PROBABILIDAD Y ESTADÍSTICA

una introducción

CÉSAR A. HERAZO HENRIQUEZ

PROBABILIDAD Y ESTADÍSTICA una introducción

Autor: **CÉSAR A. HERAZO HENRIQUEZ**

Primera edición, primer semestre 2014

ISBN: 978-958-8736-57-0

519.2 / H413
Herazo Henríquez, César A.
Probabilidad y estadística: Una introducción / César A. Herazo Henríquez; Fredy Badran Padaui, Editor -- Cartagena de Indias: Editorial Universitaria, c2013
328p.
Incluye índice alfabético.
Incluye referencias bibliográficas (p.323 -324)

ISBN: 978-958-8736-57-0
1. Probabilidades 2. Estadística Matemática I. Badran Padaui, Freddy, Ed.
CEP: Universidad de Cartagena. Centro de Información y Documentación José Fernández de Madrid.

Editor: Freddy Badrán Padauí, Jefe de Sección de Publicaciones

Universidad de Cartagena

Diseño de Portada: Jorge Barrios Alcalá

Fotografía solapa: Mario Lorduy Benedetti

Diagramación: Alicia Mora Restrepo

E- mail: mcherazoh@yahoo.com

Editorial Universitaria, Centro, Calle de la Universidad, Cra. 6, Nº 36 -100, Claustro de San Agustín, primer piso.

A la memoria de mis abuelos, Tío Toño y pa' César,
y a la de mi Papá, el mono Herazo.
A Maita.
A mi esposa y mis hijos.

Agradecimientos

Quiero agradecer a la Universidad de Cartagena por permitirme culminar este proyecto que hace muchos años inicié y a la Facultad de Ciencias Exactas y Naturales las cuales me dieron el amplio apoyo para terminar el texto. Agradezco al Doctor Humberto Pérez G. quien me sacó de los innumerables problemas que se presentaron en la transcripción, gráficas y corrección del libro. También al Magíster y estudiante de doctorado Pedro Ortega P. por sus atinados comentarios, al Magíster Elias Salazar B. por la precisión en sus correcciones y al matemático Jhon F. Cantillo por la elaboración de las tablas del apéndice D.

Doy gracias a mis hijos Ericka y César Andrés, quienes transcribieron capítulos del libro y ayudaron a corregir el manuscrito final. Por último, reconozco el apoyo de todos aquellos que han usado parte de este texto antes de editar y que hicieron los comentarios pertinentes para que el presente volumen tomara su forma final.

Índice general

Introducción

Aunque en la literatura estadística existen muchos libros de texto buenos y reconocidos, el presente se diferencia de los previos ya que hace un esfuerzo por combinar los contenidos de la probabilidad y la estadística con la enseñanza de estos temas. Creo que una recolección bibliográfica de autores tradicionales y modernos, descriptivos y formales, produciría un texto de buen nivel para quienes se inician o desean profundizar en el estudio de esta disciplina. Además, considero que esto conjugado con unas notas empíricas que condensen el programa seguido, redundará en beneficio del estudiante en aras de una mejor comprensión de los temas expuestos.

La secuencia de los temas tratados en el libro, se inicia con los elementos de la estadística descriptiva en el capítulo 1 donde se incluyen los conceptos de población y muestra, y se presenta el estudio descriptivo de los datos mediante tablas, gráficos y medidas de localización. El concepto de probabilidad se presenta en el capítulo 2 y se dan sus propiedades; se trata además la probabilidad condicional y la independencia de eventos. Una descripción general de las variables aleatorias se encuentran en el capítulo 3, las cuales son fundamentales para la teoría estadística. Luego en el capítulo 4 se presentan algunas variables aleatorias de uso común, que tienen nombres especiales y que aparecen constantemente en las diversas aplicaciones. En los capítulos 5 y 6 se trabaja la teoría de las variables aleatorias distribuidas conjuntamente y se dan dos de los resultados más importantes de la teoría estadística como son la *ley de los grandes números* y el *teorema central del límite*. El capítulo de enlace entre la probabilidad y la inferencia estadística es el 7, en el cuál se tratan las muestras, los estadísticos y las distribuciones muestrales. Finalmente, el objeto de la estadística, como es la inferencia, se trata en los capítulos 8, 9 y 10 con los títulos de *estimación puntual*, *estimación por intervalo* y *pruebas de hipótesis.*

El capítulo 1 está destinado para ser leido por estudiantes de diversos programas que requieran un mínimo de formación de estadística y sin necesidad de la orientación de un docente. El resto del texto está orientado a la enseñanza de la probabilidad y la estadística, y se recomienda la constante orientación de un profesor para afianzar los temas presentados.

Este texto está concebido para dos cursos: uno de *probabilidad* que comprende los capítulos del 2 al 6, y el otro de *inferencia estadística* que abarca los capítulos del 7 al 10. Para lograr una buena apropiación del libro se recomienda tener como requisitos cursos formales de cálculo diferencial e integral. Puede adaptarse, omitiendo las demostraciones más elaboradas, a estudiantes de todas las áreas que requieran de la probabilidad y la estadística en su proceso de formación.

Por último, quiero anotar que el presente libro es el producto de un buen tiempo dedicado a la investigación y a la docencia, y que solo fue posible condensar durante el año sabático concedido por la Universidad de Cartagena al autor en el período enero de 2012 a enero de 2013, durante el cual me fue posible organizar, mejorar y editar las notas acumuladas en mis años de labor en esta disciplina.

Capítulo 1

Estadística descriptiva

1.1. Introducción

La palabra *estadística* viene del latín *status* que significa estado y por varias décadas fue asociada solamente con la presentación de hechos y figuras relacionados con la economía, la demografía y otras situaciones políticas sobresalientes en un país. El término estadística es hoy ampliamente usado y es así, entre algunos ejemplos, como los médicos llaman a la historia de los pacientes, la estadística de estos; los comentaristas del deporte llaman al reporte de los equipos (partidos jugados, ganados, perdidos, puntos, etc), estadística de los equipos. Sin embargo, este uso del término, no es el más apropiado, pues la estadística se ocupa de situaciones en las cuales la ocurrencia de un evento no se puede predecir con certeza.

De una forma general, la **estadística es la ciencia de resolver problemas en presencia de la variabilidad**. Sin embargo, un concepto más aterrizado es el siguiente: la **estadística es un cuerpo de conceptos y métodos usados para coleccionar e interpretar datos provenientes de un área particular de la investigación y para sacar conclusiones en situaciones donde la incertidumbre y la variación están presentes**. De acuerdo a esto, la estadística contempla el arte y la ciencia de la colección, interpretación, y análisis de datos, y la habilidad para sacar generalidades lógicas que relacionen el fenómeno bajo investigación. Si se recuerdan los estados esenciales del método científico (que incluye especificación de objetivos, recolección de información, análisis de datos y exposición de lo encontrado), es claro que la estadística penetra extensamente en el dominio de la investigación científica.

Después que los datos han sido recolectados, hay una gran necesidad de usar métodos estadísticos. Uno de estos métodos está diseñado para resumir la información contenida en los datos y describir los rasgos prominentes, el otro método está encaminado a establecer generalidades o inferencias acerca del fenómeno de estudio. El primero se conoce como **estadística descriptiva** y el segundo como **estadística inferencial**.

En una investigación hay, en general, un conjunto de datos que puede ser muy grande y que por muchos factores no se pueden recolectar todos, sin embargo para hacer alguna inferencia sobre estos se debe tomar una parte de ellos. Esto conduce a las siguientes definiciones:

Definición 1.1. Una **población** (estadística) es el conjunto completo de medidas posibles o el registro de algún rasgo cualitativo correspondiente a la colección completa de unidades para la cual serán hechas las inferencias.

Definición 1.2. Una **muestra** de una población es el conjunto de medidas que son realmente recolectadas en el curso de una investigación.

Con base en una muestra, que debe ser *bien* tomada se obtienen las conclusiones acerca de una población, lo cual es el objetivo fundamental de la estadística.

Los objetivos principales de la estadística son:

- Obtener inferencias acerca de una población, a partir del análisis de información contenida en datos muestrales.
- Evaluar la incertidumbre incluida en estas inferencias.
- Diseñar el proceso y extensión del muestreo de manera que las observaciones formen una base para poder hacer inferencias válidas y seguras.

1.2. Estudio descriptivo de los datos

El proceso de la recolectar datos puede incluir actividades diversas tales como experimentos de laboratorio, pruebas de campo, estudios de opinión, el examen de registros históricos, etc. De esta manera resulta un conjunto de datos consistente de medidas numéricas que pueden ir desde unas pocas figuras a una complejidad de cientos o miles de números. Esta información debe ser organizada y analizada para inferir u obtener conclusiones.

El resumen y la exposición de los aspectos más importantes de un conjunto de datos se llama *estadística descriptiva* e incluye la condensación de los datos en forma de tablas, su representación gráfica y el cálculo de indicadores numéricos del centro y la variabilidad.

Los aspectos principales para describir un conjunto de datos son:

a. Resumen y descripción del patrón total de los datos por:

   - Presentación por tablas y gráficas.
   - Examen de la forma total de los datos graficados para observar rasgos importantes, incluyendo simetría.
   - Inspección de los datos graficados para examinar si hay observaciones inusuales (outliers) las cuales se alejan de la mayoría de los datos.

b. Cálculo de medidas numéricas para identificar:

   - Un valor típico o representativo que indique la localización central de los datos.
   - La cantidad de variación presente en los datos.

1.2.1. Descripción por gráficas y tablas

Cuando los datos son cuantitativos (medidos, contados, etc.), los dos métodos gráficos principales para representar un conjunto de datos son: el diagrama de puntos, que se emplea cuando se tienen pocas observaciones (menos de 25) y el histograma de frecuencias relativas que se usa cuando hay un gran número de datos.

Diagrama de puntos

Cuando hay pocas observaciones, ellas se pueden representar gráficamente dibujando una línea con una escala que cubra el rango de los valores de las medidas y sobre esta se dibujan los datos como puntos prominentes. La figura así obtenida se llama diagrama de puntos.

Ejemplo 1.1. Al hojear 20 páginas de un texto próximo a publicarse, el autor observa que los errores por página son: 1, 2, 2, 2, 12, 3, 3, 3, 4, 4, 10, 4, 4, 5, 5, 5, 6, 7, 4, 3. Construya el diagrama de puntos y hacer las observaciones pertinentes.

Solución. Se construirá una línea horizontal que tenga una escala de 0 a 12 y sobre ella se colocan los puntos de manera que si una observación se repite se coloca encima del dato idéntico a la observación. La figura 1.1 muestra el diagrama de puntos de esta situación.

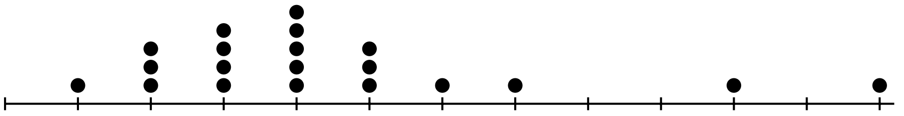


Figura 1.1. Diagrama de puntos de los errores por página

Se observa que la mayoría de los datos están distribuidos alrededor de un valor central, 4 o 5, excepto para los valores 10 y 12, los cuales se desvían sustancialmente de los otros. El autor debe examinar estas dos últimas observaciones.

Distribución de frecuencias

Cuando el conjunto de datos es grande, graficar un diagrama de puntos es tedioso, más aún, el apiñamiento de estos en el diagrama puede resultar que oscurezca detalles en las áreas donde las observaciones están fuertemente concentradas. En tales casos, es conveniente condensar los datos agrupando las observaciones cercanas y construir una tabla de frecuencias.

Para construir una distribución de frecuencias se procede de la siguiente manera:

a. Se encuentran los valores mínimo(x_{min}) y máximo(x_{max}) en el conjunto de datos.

b. Se escoge un número k, de sub-intervalos o celdas de igual longitud L, que cubra el rango R, entre el mínimo y el máximo, sin sobreposición. Estos se llaman intervalos de clase y sus puntos finales se conocen como acotamiento de clase.

c. Se cuenta el número de observaciones que caen en cada intervalo de clase. Este número se llama la frecuencia de clase, f_i. Luego se calcula la frecuencia relativa de cada clase f_{ir}, por $f_{ir} = \frac{f_i}{n}$, donde n es el número de datos. "La frecuencia relativa de una clase representa el porcentaje de información en la clase ".

Observación.
La escogencia del número y posición de los intervalos de clase es fundamentalmente

una materia de prueba y error: el número de clases usualmente va de 6 a 15 dependiendo del número de observaciones. Agrupar las observaciones en clases sacrifica la información concerniente a cómo las observaciones están distribuidas en cada celda. Si hay pocas celdas, la pérdida de información es seria, y si hay muchas y el conjunto de datos es muy pequeño, las frecuencias de una celda a la próxima tienden a subir y a bajar de una forma caótica y no producen un patrón para la distribución total de los datos. Como un paso inicial, las frecuencias se determinan con un gran número de intervalos que se pueden combinar posteriormente como se desea para obtener una escogencia de tal forma que el patrón de la distribución sea visible. Sieembargo, existen varias reglas que pretenden dar el número ideal de intervalos: Una es la regla de Sturges que establece: *tome k el entero más cercano a* $1 + 3,322 \log_{10} n$; otra establece que k debe ser el entero más cercano al número $2,5\sqrt[4]{n}$. Para n grande estos dos números son parecidos.

Un lugar decimal extra (.5, .05, .005, etc.) se usa para definir el acotamiento de clase, dependiendo si los datos son enteros, con un decimal, con dos decimales, etc. De esta manera ninguna observación puede caer exactamente sobre el acotamiento evitando así la ambigüedad en la clasificación de los datos.

Ejemplo 1.2. Dentro de un grupo de 60 trabajadores dedicados a una actividad específica se hizo una prueba sobre esa acitividad y los puntajes obtenidos se muestran abajo. Construir una distribución de frecuencias.

Puntajes de 60 trabajadores sobre una actividad X

84.1	88.9	90.2	80.4	83.7	83.6	82.3	87.5	86.4	84.4
83.3	85.0	87.7	85.8	87.2	87.7	86.0	83.9	83.9	84.2
80.1	82.8	81.3	86.7	84.7	88.2	79.1	82.9	87.0	77.4
86.3	87.8	80.7	92.0	82.1	80.6	82.6	83.1	81.8	89.1
81.9	86.4	86.1	85.5	85.7	90.8	84.5	83.9	86.7	83.4
88.8	81.5	85.5	85.2	89.8	83.6	86.9	82.6	91.2	84.8

Solución. De los datos se observa que: x_{min} =77.4 y x_{max} =92.0. Luego, $R = x_{max} - x_{min}$ =14.6.

Como $n = 60$ y 2.5 $\sqrt[4]{n} \approx 6.958$, sea $k = 7$; y como R/k =2.09, sea L =2.2. ¿Será que si se toma L=2.1, también funciona?

La distribución de frecuencias de los datos se da en la Tabla 1.1:

Intervalos	f_i	f_{ir}	$h_i = f_{ir}/L$
77.35 – 79.55	2	0.0333	0.0151
79.55 – 81.75	6	0.1000	0,0454
81.75 – 83.95	17	0.2800	0.1272
83.95 – 86.15	14	0.2333	0.1060
86.15 – 88.35	13	0.2167	0.0985
88.35 – 90.55	5	0.0833	0.0378
90.55 – 92.75	3	0.0500	0.0227
	60	0.9966	

Tabla 1.1: Distribución de frecuencias para la actividad de 60 trabajadores

En esta tabla se han usado los siguientes criterios para definir los intervalos:

- El límite inferior del primer intervalo se construye así: $V_{min} - 0.05$ y para obtener el límite superior se le suma la longitud L al límite inferior.
- El límite inferior del segundo intervalo es el límite superior del primero y su límite superior se obtiene sumándole L. En general, el límite superior del intervalo k es el límite inferior del primer intervalo más k veces L, es decir, $L_{sk} = L_{i1} + kL$.

Después de resumir un gran conjunto de datos en la forma de una distribución de frecuencias, este puede representarse gráficamente en un histograma de frecuencias relativas, el cual es una representación visual del patrón de la distribución.

Para dibujar un histograma de frecuencias relativas, los intervalos de clase se colocan en una línea horizontal y, sobre cada intervalo se levanta un rectángulo vertical con área igual a la frecuencia relativa de ese intervalo. En este punto debe tenerse en cuenta que la altura de cada rectángulo h_i se obtiene por la fórmula $h_i = \frac{f_{ir}}{L_i}$, donde f_{ir} es la frecuencia relativa del intervalo i y L_i es la longitud del mismo intervalo.

El área de cada rectángulo en un histograma de frecuencias relativas representa la proporción de las observaciones que ocurren en el intervalo de clase sobre el cual se levanta el rectángulo. De aquí que el área de la suma de todos los rectángulos en un histograma es 1. La convención de usar el área de todos los rectángulos en lugar de sus alturas para representar sus frecuencias relativas, tiene la siguiente ventaja: a ojo se pueden comparar dos partes de un histograma o dos histogramas diferentes, y cuando dos histogramas están basados en intervalos de clase de diferente amplitud, la propiedad de tener área igual a 1 hace a estos comparables.

Ejemplo 1.3. En el ejemplo 1.2 de la prueba a los 60 trabajadores, construir un histograma de frecuencias relativas.

Solución. De acuerdo a la tabla 1.1 se colocan los intervalos sobre una linea horizontal y luego se levantan los rectángulos con las alturas obtenidas en la columna 4. La Figura 1.2 es el histograma de frecuencias relativas para este ejemplo.

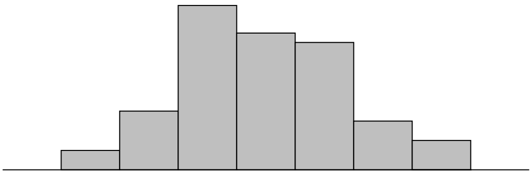


Figura 1.2. HFR de la prueba a 60 trabajadores

La regla que sugiere intervalos de clase iguales no es apropiada cuando los datos están separados sobre un amplio rango, pero que están altamente concentrados en una pequeña parte del mismo con relativamente pocos datos dispersos. Usando intervalos de pequeña longitud donde los datos están altamente concentrados e intervalos grandes cuando los datos están bien dispersos, ayuda a reducir la pérdida de información debido al agrupamiento.

Las tabulaciones de ingreso, edad y otras características en reportes oficiales se hacen a menudo con intervalos de clase desiguales. Cuando los intervalos de clase no son todos iguales, el histograma se debe graficar de acuerdo a la convención de usar el área de los rectángulos para representar las frecuencias relativas, para conducir el patrón correcto de la distribución.

Ejemplo 1.4. La distribución de frecuencias del número de personas muertas en los grandes tornados en los E.E.U.U. entre 1900 y 1973 se muestra en la tabla 1.2. Graficar el histograma de frecuencias relativas y comentar la forma de la distribución.

N de muertos	Frecuencia
≤ 24	8
$25 - 49$	16
$50 - 74$	16
$75 - 99$	11
$100 - 149$	6
$150 - 199$	2
$200 - 249$	4
≥ 250	1

Tabla 1.2: Distribución de frecuencias del número
de muertos por tornados en E.E.U.U. entre 1900 y 1973

Solución. Se construye primero una distribución de frecuencias de manera que las clases sean continuas (ver tabla 1.3). Con el objeto de construir el histograma de la figura 1.3 se han acotado el primer y último intervalo. La longitud de estos depende de cómo están relacionadas las frecuencias relativas del primero con el segundo y la del penúltimo con la del último. Por ejemplo, si la frecuencia relativa del último intervalo es menor que la del penúltimo, se toma ese intervalo con igual longitud a la del anterior.

Se observa además, que la distribución de muertes en los tornados es sesgada a la derecha.

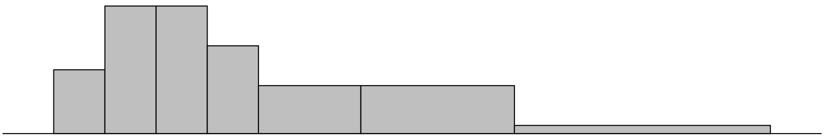


Figura 1.3. Histograma del número de muertos en tornados en USA entre 1900 y 1973.

Diagrama de líneas de frecuencias relativas

Algunas veces el conjunto de datos resulta de contar, tales como el número de niños por familia o el número de artículos defectuosos por día en un proceso de

Intervalos	f_i	f_{ir}
$-0.5 - 24.5$	8	0.125
$24.5 - 49.5$	16	0.250
$49.5 - 74.5$	16	0.250
$74.5 - 99.5$	11	0.172
$99.5 - 149.5$	6	0.094
$149.5 - 249.5$	6	0.094
≥ 249.5	1	0.016
	64	1.001

Tabla 1.3: Distribución de frecuencias corregida del número de muertes por tornados en Estados Unidos entre 1900 y1973.

producción o el número de accidentes de tránsito por día, en vez de ser medidos sobre una escala continua. Si el número de valores distintos en tal conjunto de datos no es muy grande, se construye una distribución de frecuencias usando los valores individuales como las clases en vez de usar intervalos de clase. Los datos se presentan en la forma de un *Diagrama de Líneas de Frecuencias Relativas*.

Un diagrama de líneas de frecuencias relativas se construye así: Los valores distintos se localizan sobre una línea horizontal y luego se levantan líneas verticales sobre estos valores con alturas iguales a las frecuencias relativas.

Las líneas reemplazan los rectángulos para enfatizar que las frecuencias no están realmente cubriendo intervalos.

Un histograma puede construirse también dibujando rectángulos centrados sobre los valores distintos de los datos, con tal que las frecuencias relativas sean consideradas como las alturas. Aunque los diagramas de línea y los rectángulos se usan por datos contados, los diagramas de línea nunca deben dibujarse para medidas sobre una escala continua.

Ejemplo 1.5. En un centro de cómputo, el número diario de interrupciones de un computador debido a error de máquina fueron registrados por un período de 40 días y los siguientes datos se obtuvieron. Construir una distribución del número de interrupciones por día y un diagrama de líneas.

1	2	0	0	1	5	4	3	3	1	2	5	0	0	3
1	2	0	0	1	0	0	0	1	1	0	2	0	2	4
4	0	2	2	6	1	0	0	1	3					

Solución. La distribución de frecuencia se realiza teniendo en cuenta que los datos son contados y el número de valores distintos es poco. Esta se encuentra en la tabla 1.4 y el diagrama de líneas se da en la figura 1.4.

N de Interrupciones	Frecuencia	Frecuencia relativa
0	14	0.350
1	9	0.225
2	7	0.175
3	4	0.100
4	3	0.075
5	2	0.050
6	1	0.025
Total	40	1.000

Tabla 1.4: Distribuición de frecuencias del número diario de interrupciones de un computador debido a error de máquina.

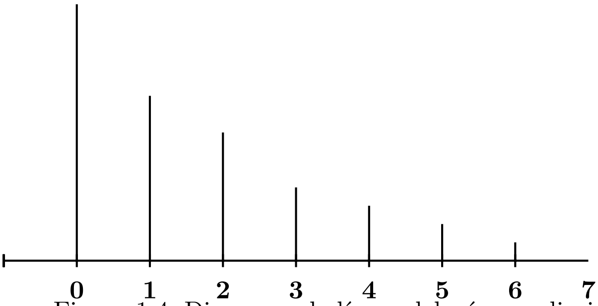


Figura 1.4. Diagrama de líneas del número diario de interrupciones de un computador debido a error de máquina

Polígono de frecuencias relativas

Esta gráfica se obtiene del histograma de frecuencias relativas uniendo mediante trazos de recta los puntos medios (altos) de los rectángulos; como punto de partida se toma el punto medio del intervalo anterior al primero y como punto de llegada el punto medio del intervalo siguiente al último . La figura 1.5 muestra el polígono de frecuencias para la prueba de los 60 trabajadores.

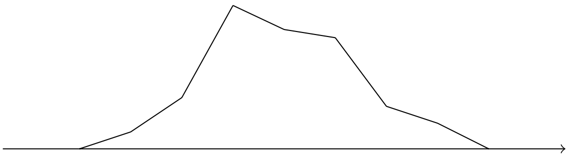


Figura 1.5. Polígono de frecuencias de la prueba a 60 trabajadores

Otros métodos gráficos

Cuando las categorías básicas no son cuantificables tales como: tendencia política, preferencias de color, fuentes de energía, entre otros, se emplean el diagrama de pastel y el diagrama de franjas.

El diagrama de pastel es un gráfico donde se dibuja un círculo (o un pastel) y se divide este en sectores (rebanadas) de acuerdo al total o porcentaje de cada categoría.

Ejemplo 1.6. En una fábrica se encontró que el personal total se distribuye de la siguiente manera: De oficina 25 %, operarios 40 %, profesionales 20 % y otros 15 %. Dibuje un diagrama de pastel.

Solución. En la figura 1.6 aparece el diagrama donde, por ejemplo, el sector correspondiente a **operarios** es el 40 % equivalente a 144 de los 360 grados del círculo.

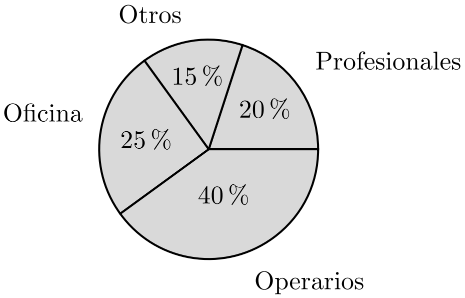


Figura 1.6 Diagrama de pastel para la distribución del personal de una fábrica.

El diagrama de franjas se construye colocando las categorías sobre una línea horizontal (o vertical) separadas por una longitud establecida y luego se dibujan rectángulos con el total o proporción de cada una de ellas.

Ejemplo 1.7. En un curso de matemáticas de primer semestre las notas definitivas fueron: 6 excelentes (E), 9 buenas (B), 16 regulares (R) y 19 malas (M). Construya un diagrama de franjas.

Solución. Las notas de este curso corresponden en porcentajes a: excelentes 12 %, buenos 18 %, regulares 32 % y malos 38 %. En la figura 1.7 está construido el diagrama de franjas vertical de esta información.

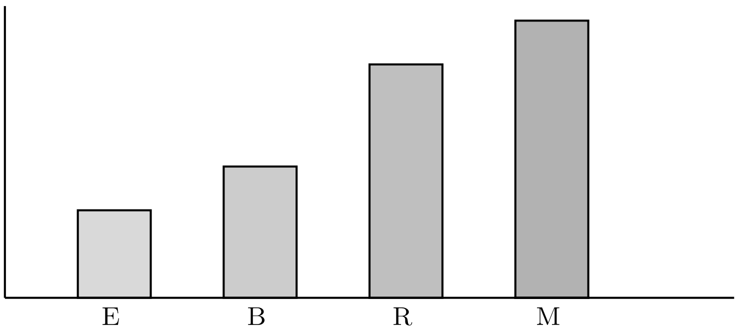


Figura 1.7. Diagrama de franjas para la clasificación de las notas del curso de matemáticas.

Debe tenerse en cuenta que tanto en el diagrama del pastel como el de franjas el orden de ubicación de las categorías no es importante. Por ejemplo, el diagrama de franjas de la figura 1.8 da la misma información que el anterior.

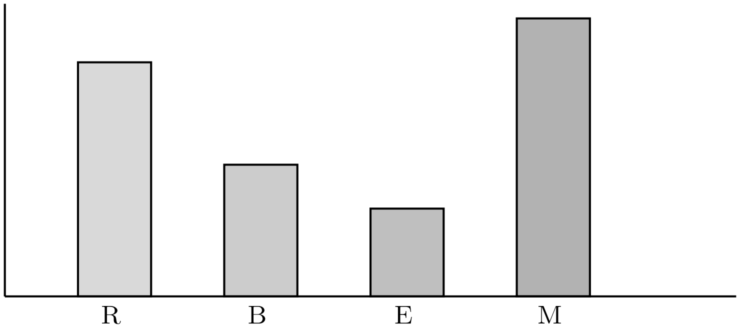


Figura 1.8. Diagrama de franjas para la clasificación de las notas del curso de matemáticas.

1.2.2. Descripción por medidas

Otra forma muy importante de describir un conjunto de datos es mediante el cálculo de medidas numéricas llamadas de localización y de variabilidad.

Medidas de localización

Estas medidas intentan localizar puntos importantes en el conjunto de datos como son el centro o cualquier "percentil". Quizás el aspecto más importante del estudio de la distribución de una muestra de medidas es la posición de un valor central, es decir, un valor representativo alrededor del cual las medidas están distribuidas. Cualquier medida numérica que intente representar el centro de un conjunto de datos, se llama "medida de tendencia central". Las dos medidas del centro más usadas son la media y la mediana.

Definición 1.3. La media muestral o promedio de un conjunto de n medidas $x_1, x_2, ..., x_n$ es la suma de las medidas dividida por n. La media se denota por $\bar{x}$ y de acuerdo con la definición es:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}. \tag{1.1}$$

A partir del concepto de promedio, la media representa el centro de un conjunto de datos. Si se dibuja el diagrama de puntos de un conjunto de datos como una barra horizontal delgada sobre la cual cubos de igual tamaño y peso se colocan en las posiciones de los puntos que representan los datos, entonces la media $\bar{x}$ representa el punto de equilibrio sobre el cual la barra está balanceada. El cálculo de $\bar{x}$ y su interpretación geométrica se ilustra en el ejemplo 8.

Ejemplo 1.8. Las medidas en centímetros (cm) de 7 bebés al nacer en un hospital durante un cierto día son: 45, 49, 52, 48, 50, 55, 53. Luego la media de estas medidas es

$$\bar{x}=\tfrac{45+49+52+48+50+55+53}{7}=\tfrac{352}{7}=50.29.$$

El diagrama de puntos aparece en la figura 1.9 donde la media muestral es el punto de balanceo o centro del gráfico.

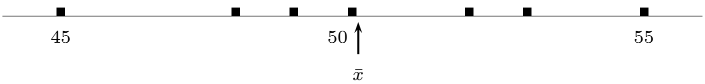


Figura 1.9. Interpretación geométrica de la media

Definición 1.4. La mediana muestral de un conjunto de medidas $x_1, x_2, ..., x_n$ es el valor intermedio cuando las medidas están ordenadas de menor a mayor. Si n es un número impar, hay un valor intermedio único y este es la mediana. Si n es par, hay dos valores intermedios y la mediana se define como el promedio de ellos. La mediana se denota por $\xi_{0.5}$ o por $\tilde{x}$, y

$$\xi_{0.5} = \tilde{x} = \begin{cases} x_{(\frac{n+1}{2})} & \text{si } n \text{ es impar,} \\ \dfrac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} & \text{si } n \text{ es par.} \end{cases} \tag{1.2}$$

En lenguaje común, la mediana es el valor que divide los datos en dos partes iguales; es decir, 50 % de los datos están a la izquierda o encima de la mediana y 50 % están a la derecha o debajo de la mediana.

Ejemplo 1.9. Encontrar la mediana de las medidas de los nacimientos de los bebés del ejemplo 1.8.

Solución. Las medidas arregladas de menor a mayor son:

$$45 \quad 48 \quad 49 \quad 50 \quad 52 \quad 53 \quad 55.$$

Se observa que el único que está en el centro es 50, por lo tanto la mediana es $\xi_{.5}$=50cm.

Ejemplo 1.10. Los ingresos mensuales (por diezmil pesos) para 10 técnicos de una empresa son: 95, 92, 105, 122, 125, 132, 237, 90, 145 y 280. Calcular la media y la mediana.

Solución. La media es

$$\overline{x}=\frac{\sum_{i=1}^{n} x_i}{10}=\frac{1420}{10}=142.$$

Luego, el ingreso promedio mensual para el grupo es $1.420.000. Este valor es cuestionable como ingreso mensual típico porque 7 de los 10 valores están por debajo de 142. Los valores 237 y 280 inflan drásticamente la media.

Para encontrar la mediana se ordenan los datos de menor a mayor,

$$90 \quad 92 \quad 95 \quad 102 \quad 122 \quad 125 \quad 132 \quad 145 \quad 237 \quad 280.$$

Los dos valores del medio son 122 y 125. Luego, el ingreso mediano es

$$\xi_{.5}=\frac{122+125}{2}=123.5.$$

De manera que el ingreso mediano es de $1.235.000, lo cual parece ser una medida central más sensata que la media en este caso.

Comentario
Este ejemplo muestra que la mediana no está afectada por unas pocas observaciones muy pequeñas o muy grandes, mientras que la presencia de tales extremos tendrá un efecto significante sobre la media. Para distribuciones extremadamente asimétricas la mediana es una medida más sensata del centro que la media. Esta es la razón del por que los reportes del gobierno sobre distribuciones de ingresos, citan el ingreso mediano como un resumen antes que el promedio. Cuando la distribución no es altamente asimétrica, se prefiere la media. Esta es más usada que la mediana, la cuál carece de algunas ventajas teóricas cuando se hacen inferencias.

Si el número de observaciones es grande (mayor que 25) es útil extender la noción de la mediana y dividir el conjunto de datos en cuartos, obteniendose los llamados *cuartiles*. Sin embargo, un concepto más general y que involucra a los cuartiles son los *percentiles*, pues como se verá enseguida, el primer, el segundo y el tercer cuartil son respectivamente los percentiles del 25, 50 y 75 % . Observese que si p, $0 < p < 1$, es una fracción entonces $100p$ representa un porcentaje. Por ejemplo, $\frac{1}{5}$ es sinónimo de 20 %.

Definición 1.5. El percentil del $100p\,\%$, denotado ξ_p, es un valor tal que, después que los datos se han ordenados de menor a mayor, al menos $100p\,\%$ de las observaciones están a la izquierda (arriba) de este valor y al menos $100(1-p)\,\%$ están a la derecha (abajo) de este valor, $0 < p < 1$.

Los cuartiles son simplemente el percentil del 25 % ($\xi_{.25}$), la mediana ($\xi_{.5}$) y el percentil del 75 %($\xi_{.75}$).

Se hace la siguiente identificación:

- Cuartíl inferior: Q_1=$\xi_{0.25}$
- Segundo cuartíl: Q_2=$\xi_{0.5}$
- Cuartíl superior: Q_3=$\xi_{0.75}$

Se tomará un valor observado para el percentil muestral excepto cuando dos valores adyacentes satisfagan la definición, en cuyo caso se toma el promedio de estos. Esto coincide con la definición de la mediana cuando el tamaño muestral es par. Cuando todos los valores en un intervalo satisfacen la definición de un percentil, la convención particular usada para localizar un punto en el intervalo no altera apreciablemente el resultado, excepto quizás, para la determinación de percentiles extremos, aquellos que están antes $\xi_{0.05}$ o después de $\xi_{0.95}$.

Se ilustra el método de encontrar percentiles mediante el siguiente ejemplo.

Ejemplo 1.11. Los datos de 40 medidas del nivel de tráfico en una intersección se presentan a continuación y están ordenados de menor a mayor. Localizar los cuartiles y los percentiles del 5, 12 y 95 %.

52.0	54.4	54.5	55.7	55.8	55.9	55.9	56.2	56.4	56.4
56.7	56.8	57.2	57.6	58.9	59.4	59.4	59.5	59.8	60.0
60.2	60.3	60.5	60.6	60.8	61.0	61.4	61.7	61.8	62.0
62.1	62.6	62.7	63.1	63.6	63.8	64.0	64.6	64.8	64.9

Solución. Para determinar el primer cuartil, se deben contar al menos $0.25 \times 40 = 10$ observaciones desde la medida más pequeña y al menos $0.75 \times 40 = 30$ observaciones desde la mayor. La observación décima es 56.4 y la trigésima a partir de la mayor es 56.7. Luego Q_1 es

$$Q_1 = \frac{56.4+56.7}{2} = 56.55.$$

Para determinar la mediana se cuenta al menos $0.5 \times 40 = 20$ obervaciones a partir de la menor y 20 observaciones a partir de la mayor hacia la izquierda. La observación 20 a partir de la menor es 60.0 y la 20 desde la mayor es 60,2, luego la mediana es

$$Q_2 = \frac{60.0+60.2}{2} = 60.1.$$

Para el tercer cuartil se cuenta al menos 30 observaciones desde la menor, que es 62,0 y 10 observaciones desde la mayor, que es 62,1. Por tanto,

$$Q_3 = \frac{62.0+62.1}{2} = 62.05.$$

Para $\xi_{.05}$ se deben contar al menos $0.05 \times 40 = 2$ observaciones desde la menor, esta es 54.4; y contar al menos $0.95 \times 40 = 38$ observaciones desde la mayor, esta es 54.5. Luego,

$$\xi_{0.05} = \frac{54.4+54.5}{2} = 54.45.$$

Para $\xi_{0.95}$ se tiene que:

$$\xi_{0.95} = \frac{64.6+64.8}{2} = 64.7.$$

Para $\xi_{0.12}$, se calcula $0.12 \times 40 = 4.8$, se toma la observación 5; se calcula $0.88 \times 40 = 35.2$ y se toma la observación 36. Como estas dos observaciones coinciden se tiene que $\xi_{0.12}$=55.8.

Otras medidas de tendencia central con usos específicos son:

- La *moda*, la cual se define como la observación que más aparece en el conjunto de datos. Puede no ser única.
- La *media geométrica* G, se define por:

$$G = \sqrt[n]{x_1 x_2 ... x_n} \ . \tag{1.3}$$

- La *media armónica* H, se define por:

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \ . \tag{1.4}$$

Es de anotar que la media geométrica es útil para promediar valores en proporciones o porcentajes, mientras que la armónica se usa para promediar ciertos tipos de relaciones o tasas: tasas de nacimiento, de mortalidad, velocidades en km/h, hectárea/h, etc.

Existen dos medidas, llamadas la *media recortada* y la *media de Winzor*, las cuales han ganado importancia en los últimos años.

Como se dijo, la media es sensible a la existencia de unas pocas observaciones muy grandes o muy pequeñas, mientras que la mediana los ignora. Las medias recortada y de Winzor pueden considerarse como un término medio entre la media y la mediana.

- La *media recortada o truncada*, $\bar{x}_r$, es la media de las observaciones que quedan cuando se remueven las observaciones inferiores a Q_1 y las superiores a Q_3.

- La *media de Winzor*, $\bar{x}_w$, es la media de las observaciones de los datos que se obtienen al remplazar las observaciones inferiores a Q_1 por Q_1 y las superiores a Q_3 por Q_3.

Estas medias no se alteran por la presencia de una fracción pequeña de observaciones inusuales o erróneas que son o extremadamente grandes o pequeñas. Los estadísticos han encontrado que estas dos medidas son casi tan buenas como la media en distribuciones simétricas cuando no hay observaciones inusuales, pero ambas medidas son usualmente mejores en la presencia de valores extraños.

Cuando una distribución tiene una larga cola, estas medidas suministran índices alternativos de un valor central diferente de aquellos suministrados por la media y la mediana.

Las medias anteriores también se definen considerando los percentiles $\xi_{0.05}$ y $\xi_{0.95}$, y, $\xi_{0.10}$ y $\xi_{0.90}$, dependiendo de las colas de la distribución de los datos.

Ejemplo 1.12. Con los datos del ejemplo 1.11, encontrar la media, la media recortada y la media de Winzor.

Solución. Para la media:

$$\bar{x}=\frac{2358}{40}=59.625.$$

Para la media recortada, se eliminan las observaciones por debajo de $Q_1 = 56.55$ y las que están por encima de $Q_3 = 62.05$, quedando las observaciones:

56.5	56.7	56.8	57.2	57.6	58.9	59.4	59.4	59.5	59.8	60.0
60.2	60.3	60.5	60.6	60.8	61.0	61.4	61.7	61.8	62.0	62.05

$$\bar{x}_r = \frac{1314.2}{22} = 59.736.$$

Para la media de Winzor se tiene que

$$\bar{x}_w = \frac{10Q_1+1195.6+10Q_3}{40} = \frac{2381.6}{40} = 59.54.$$

Medidas de variación

Además de la localización del centro de los datos, un aspecto importante de un estudio descriptivo de datos es una medida numérica del grado de variación alrededor del centro. Dos conjuntos pueden mostrar similar posición del centro pero pueden ser marcadamente diferentes con respecto a la variabilidad. En esta sección se presentan algunas medidas numéricas que facilitan una comparación objetiva del grado de variación en diferentes conjuntos de datos. Los principales son la *varianza* y la *desviación estándar*

Definición 1.6. La varianza muestral de un conjunto de medidas $x_1, x_2, ..., x_n$, se denota por s^2 y se define por

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}. \tag{1.5}$$

Cada $x_i - \bar{x}$ se llama desviación y podría pensarse que el promedio de todas las desviaciones sería una buena medida de dispersión, pero no es así pues

$$\textstyle\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

Porque en la varianza influye la suma de cuadrados, sus unidades son el cuadrado de la unidad en la cual las medidas están expresadas. Para obtener una medida de variabilidad en la misma magnitud de los datos, se toma la raíz cuadrada de la varianza, llamada desviación estándar y se denota por s. Esta, mejor que la varianza, sirve como una medida básica de variabilidad.

$$s = \sqrt{Varianza} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \tag{1.6}$$

A diferencia de la interpretación simple de $\bar{x}$ como el punto de balanceo para la distribución del conjunto de medidas, una interpretación física de la desviación estándar s no es tan transparente. En la comparación de dos conjuntos de datos, un alto valor de s refleja la presencia de una mayor variación en ese conjunto de datos que en el otro. Sin embargo, en el contexto de un conjunto de datos particular, el valor numérico de s en relación a la dispersión de los puntos no es claro.

Un resultado debido al matemático ruso Chebyshev, suministra una fracción de los datos en conexión con intervalos alrededor de la media y que involucran a s.

Regla de Chebyshev. *Para todo conjunto de datos, el intervalo* $(\bar{x} - ks, \bar{x} + ks)$ *contiene al menos* $100(1 - \frac{1}{k^2})\,\%$ *de los datos, para* $k > 1$.

La regla de Chebyshev garantiza la inclusión de una fracción mínima de los datos en un intervalo que está centrado en $\bar{x}$ y se extiende en un múltiplo de s en ambas direcciones.

Ejemplo 1.13. Examinar la regla de Chebyshev para la prueba de la actividad de los 60 trabajadores del ejemplo 1.2.

Solución. En este ejmplo $\bar{x}$=84.95, y s=3.1. Para ilustrar la regla de Chebyshev, se consideran los dos intervalos ($\bar{x}$ - 1.5s, $\bar{x}$+1.5s) y $(\bar{x} - 2s, \bar{x} + 2s)$.

El primer intervalo es equivalente a (80.3 , 89.6) y según la regla debe contener al menos el 55.56 % de los datos, es decir, 34. Realmente este intervalo contiene 52.

El segundo intervalo es equivalente a (78.75 , 91.15) y debe contener al menos el 75 % de los datos, es decir, 45. Hay 57 datos en este intervalo .

Otras medidas de Variación y que se usan algunas veces son:

1. La *desviación media*, notada D_m, está definida por la expresión

$$D_m = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}. \tag{1.7}$$

2. La *desviación mediana*, notada D_{me} y definida por

$$D_{me} = \frac{\sum_{i=1}^{n} |x_i - \xi_{.5}|}{n}. \tag{1.8}$$

3. El *rango muestral*, notado R y definido por $R = x_{max} - x_{min}$, da la longitud del intervalo que contiene todo el conjunto de datos. El rango se usa, en vez de la varianza, como una medida de dispersión cuando el conjunto de datos es menor que 8.

4. El *rango intercuartílico*, notado R_i, se define como $R_i = Q_3 - Q_1$. Esta medida da la longitud del intervalo que cubre la mitad del centro de los datos. Se usa frecuentemente en reportes gubernamentales sobre ingresos, y en distribuciones que tienen largas colas en una dirección.

5. El *coeficiente de variación*, notado CV, se define por $CV = \frac{S}{\bar{x}} \times 100$. Se utiliza cuando se comparan diferentes poblaciones o diferentes muestras, considerándose más variable aquella cuyo CV sea mayor.

Método gráfico para detectar outliers- boxplots

Los boxplots o diagrama de cajas y bigotes son técnicas gráficas que muestran la distribución de uno o más conjuntos de datos. Estos métodos alternos a los histogramas son particularmente útiles cuando se comparan dos o más conjuntos de datos. Ellos muestran la localización, el sesgo, la dispersion, la longitud de las colas y los Outliers.

Para dibujar un boxplot se debe:

- Obtener el primero, segundo y tercer cuartil; es decir, Q_1,Q_2 y Q_3.
- Calcular R_i= (Q_3 - Q_1), y luego $1.5R_i$ y $3R_i$.
- Dibujar una caja rectangular que empiece en Q_1 y termine en Q_3. Dibuje también una barra vertical en la posición de la mediana.
- Antes de Q_1 y hasta Q_1, y después de Q_3 dibuje líneas a cada lado de la caja. Estos son los **bigotes**.
- Cualquier dato que no se encuentre entre los bigotes se considera un outlier (dato atípico) y se grafica. Datos entre un bigote y la linea $Q_1{-}3R_i$ (o $Q_3{+}3R_i$) se consideran outliers suaves (representados por un círculo) y datos después de estas líneas se consideran outliers fuertes (simbolizados por un círculo obscuro).

Hay varias reglas para determinar los extremos de los bigotes. Algunas son:

- El dato más pequeño que se encuentre en $[Q_1$- $1.5R_i, Q_1)$ y el dato más alto que se encuentre en $(Q_3, Q_3{+}1.5R_i]$.
- El noveno percentil ($\xi_{.09}$) y el percentil del 91 % ($\xi_{.91}$).

La siguiente gráfica muestra cómo se construye un diagrama de caja y bigotes:

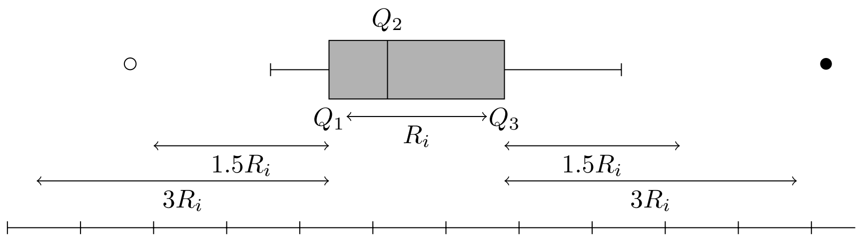


Figura 1.10. Construcción de un diagrama de caja y bigotes

En la práctica, el diagrama se presenta solo con la caja, los bigotes y los outliers como se muestra en el ejemplo.

Ejemplo 1.14. Para los 20 datos del ejemplo 1.1, $Q_1 = 3,\ Q_2 = 4$ y $Q_3 = 5$. De manera que $R_i = 2$ y así $Q_1 - 1.5R_i = 0,\ Q_1 - 3R_i = -3,\ Q_3 + 1.5R_i = 8$ y $Q_3 + 3R_i = 11$. Con esta información se construye el boxplot y es el de la figura 1.11. Se observa claramente que el valor 10 es un outlier suave y 12 un outlier fuerte.

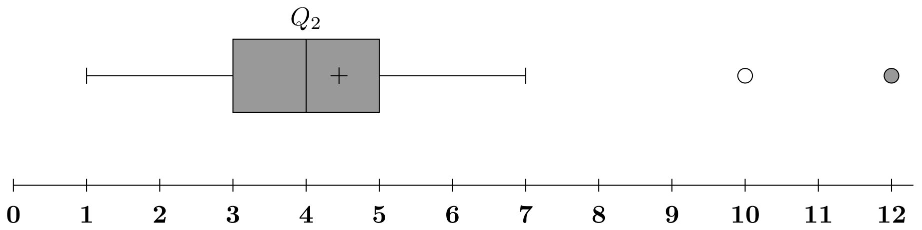


Figura 1.11. Diagrama de caja y bigotes del ejemplo 1.1

1.3. Medidas descriptivas para datos agrupados

Para el cálculo de estas medidas se amplía la distribución de frecuencias estudiada hasta ahora, ubicando algunas columnas adicionales de manera que se faciliten los cálculos. Estas columnas son:

- La de las marcas de clase m_i, las que se definen como el punto medio de cada intervalo. Es decir, $m_i = \frac{L_i+L_s}{2}$.

- La de las frecuencias acumuladas F_i, que se definen de la siguiente manera: $F_1 = f_1$, $F_2 = F_1 + f_2$, $F_3 = F_2 + f_3$,...; en general $F_k = F_{k-1} + f_k$.
- Ocasionalmente la de las frecuencias acumuladas relativas F_{ir}, las cuales se definen por $F_{ir} = \frac{F_i}{n}$.

Suponiendo que se tiene una distribución de frecuencias con k intervalos y que se desea calcular la media, la mediana, la varianza y los percentiles en general, se definen:

La media por

$$\bar{x} = \frac{\sum_{i=1}^{k} m_i f_i}{n}. \tag{1.9}$$

La Varianza por

$$s^2 = \frac{\sum_{i=1}^{k} (m_i - \bar{x})^2 f_i}{n}. \tag{1.10}$$

De esta expresión se puede obtener fácilmente que

$$s^2 = \frac{\sum_{i=1}^{k} m_i^2 f_i}{n} - \bar{x}^2. \tag{1.11}$$

Para encontrar el percentil del $100p\,\%$, ξ_p, se procede de la siguiente manera:

1. Se determina el intervalo cuya frecuencia acumulada sea igual o mayor al valor np, donde n es el número total de observaciones. Este intervalo se llama la clase percentil del $100p\,\%$, y denote al límite inferior de este intervalo como L_{ip}.
2. Calcule la fracción $(np - F_{i-1})/f_i$, donde F_{i-1} es la frecuencia acumulada del intervalo anterior al de la clase del percentil y f_i es la frecuencia de la clase del percentil.
3. El percentil de orden p es,

$$\xi_p = L_{ip} + \frac{np - F_{i-1}}{f_i} \times L, \tag{1.12}$$

donde L es la longitud del intervalo.

Ejemplo 1.15. En el ejemplo 1.2, referente a los 60 trabajadores dedicados a una actividad, encontrar la media, la desviación estándar, la mediana, los percentiles del 25 % y del 95 %.

Intervalos	m_i	f_i	F_i	f_k	F_{ir}
77.35 – 79.55	78.45	2	2	0.0333	0.033
79.55 – 81.75	80.65	6	8	0.1000	0.133
81.75 – 83.95	82.85	17	25	0.2800	0.416
83.95 – 86.15	85.05	14	39	0.2333	0.650
86.15 – 88.35	87.25	13	52	0.2167	0.866
88.35 – 90.55	89.45	5	57	0.0833	0.950
90.55 – 92.75	91.65	3	60	0.0500	1.000

Tabla 1.5: Distribución de frecuencias completa del ejemplo 2

Solución En la distribución encontrada se adicionan las columnas de las marcas de clase y de las frecuencias acumuladas, obteniéndose la tabla 1.5.

Luego,

$$\bar{x} = \frac{\sum_{i=1}^{7} m_i f_i}{60} = \frac{5096.4}{60} = 84.94$$

$$s^2 = \frac{\sum_{i=1}^{7} m_i^2 f_i}{60} - \bar{x}^2 = 7224.39 - (84.94)^2 = 9.587 \text{ y } s = 3.096.$$

Para calcular la mediana, obsérvese que p =0.5 y que el intervalo cuya frecuencia acumulada alcanza o excede justamente a $np = 30$ es (83.95, 86.15). Luego, este es el intervalo de la clase mediana y $L_{i0.5}$=83.95. En ese intervalo $f_i = f_4 = 14$ y $F_{i-1} = F_3 = 25$, por lo tanto $\xi_{0.5} = 83.95 - \frac{(30-25)}{14} \times 2.2 = 84.74$.

Para calcular $\xi_{0.25}$, obsérvese que el intervalo cuya frecuencia acumulada alcanza o excede justamente a $np = 15$ es (81.75, 83.95). Luego, este es el intervalo de la clase del percentil del 25 % y se tiene que $L_{i.25} = 81.75$. En ese intervalo $f_i = f_3 = 17$ y además $F_{i-1} = F_2 = 8$, por lo tanto $\xi_{.25} = 81.75 - \frac{15-8}{17} \times 2.2 = 82.66$.

Para encontrar $\xi_{,95}$, el intervalo cuya frecuencia acumulada excede o alcanza justamente a $np = 57$ es (88.35, 90.55), entonces $f_i = f_6 = 5$ y $F_{i-1} = 52$, por tanto $\xi_{.95} = 88.35 - \frac{57-52}{5} \times 2.2 = 90.55$.

Método gráfico para hallar percentiles

Cuando un conjunto de datos está agrupado en intervalos de clase, también se pueden obtener los percentiles utilizando un gráfico de frecuencia acumulada relativa.

Este consiste en dibujar los intervalos de clase sobre el eje horizontal del plano xy y sobre el eje vertical positivo se colocan las frecuencias relativas acumuladas, F_{ir}. Al límite superior de cada intervalo de clase se le asigna su frecuencia relativa acumulada y al límite inferior del primero se le asigna la frecuencia relativa acumulada 0 (como es lógico). Luego se dibujan trazos de recta que unan los puntos así obtenidos, resultando una curva lineal a trozos. Para hallar el percentil $100p\,\%$, se ubica p en el eje vertical, y a partir de esta se traza una recta paralela al eje X hasta que intercepte el gráfico; desde este nuevo punto se baja una recta paralela al eje Y hasta que corte la recta de los intervalos. Este punto, que se puede medir, es el percentil buscado.

Ejemplo 1.16. En el ejemplo anterior, obtener por el método gráfico la mediana y los percentiles de $25\,\%$ y $95\,\%$.

Solución. En la tabla 1.5 se construyeron dos columnas adicionales, la de la frecuencia relativa y la de la frecuencia acumulada relativa. Usando el proceso descrito, se obtiene la figura 1.12 en la cual se observa que $\xi_{.25} = 82.70$, $\xi_{.5} = 84.90$, $\xi_{.95} = 90.80$, los cuales son muy próximos a los obtenidos en el ejemplo anterior.

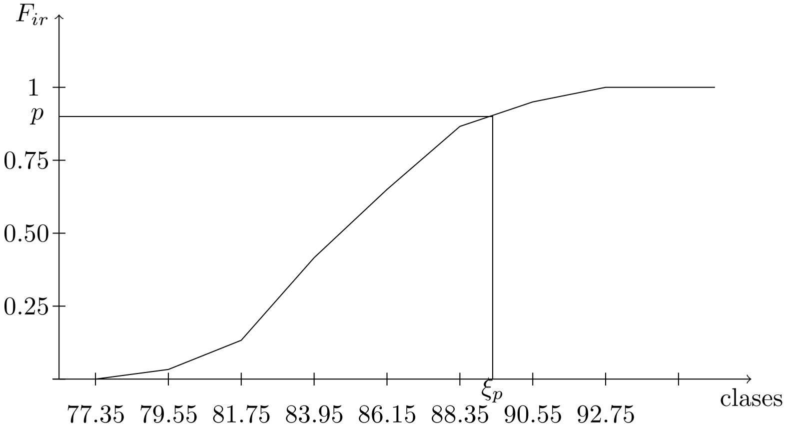


Figura 1.12. Percentiles por el método gráfico, ejemplo 1.16

1.4. Ejercicios

1. Sea los más claro posible:

a) Distinga entre Estadística Descriptiva e Inferencial.
b) ¿Cuál es el valor de una representación gráfica de un conjunto de datos?
c) Mencione 5 medidas de tendencia central y defínalas.
d) ¿Qué uso se puede hacer de la desigualdad de Chebyshev en la descripción de un conjunto de datos?

2. Las siguientes notas resultaron de una prueba hecha a un curso de Cálculo I: 3.5, 3.2, 3.8, 2.5, 3.8, 3.5, 4.0, 3.7, 3.0, 2.7, 2.0, 3.2, 2.8, 5.0, 3.5, 2.7

   a) Construir un diagrama de puntos.
   b) ¿Entre qué valores se encuentra el mayor volumen de los datos?
   c) Interpretar este conjunto de datos.

3. Las medidas en centímetros, de los huevos de 50 gallinas en experimentación en un día fueron:

7.08	7.44	6.90	8.14	7.40	6.15	7.55	8.90	9.40	7.65
6.80	8.00	7.65	6.96	8.55	7.50	7.30	8.45	7.25	6.57
7.98	6.25	9.00	7.25	8.10	7.80	8.60	7.90	6.22	8.35
8.19	7.12	8.70	8.25	6.00	8.75	7.00	9.25	6.40	7.75
6.68	7.20	8.75	7.30	8.15	6.45	7.95	8.75	6.45	7.99

   a) Clasificar la longitud de los huevos usando 10 intervalos de clase.Use también el número de intervalos dado por la regla de Sturges.
   b) En los casos anteriores, contruir los histogramas de frecuencias relativas.
   c) Obtener la media, la mediana y los percentiles del 25 %, 75 % y 95 %
   d) Por el método gráfico obtener la mediana y los percentiles del inciso anterior. Comparar.

4. Los porcentajes de cambios en la producción de petróleo crudo de 2009 a 2010 para 51 países de Norteamérica, Latinoamérica, Europa, Medio Oriente, Africa y Asia fueron (Fuente: OPEC. Annual Statistical Bulletin 2010/2011):

-0.9	2.8	-2.3	5.3	9.1	17.1	2.5	-1.0	1.1	-8.4
-0.8	38.2	1.2	-9.5	6.1	-3.1	2.0	-5.8	-1.3	-10.1
16.0	-21.8	-9.5	4.6	-7.2	-1.0	-0.4	0.9	2.2	6.4
0.1	-0.2	2.4	3.7	-6.6	-2.2	-2.7	-11.9	7.6	2.1
3.3	0.9	11.2	-4.3	-8.4	3.1	7.4	10.4	-1.8	-3.6
-5.1									

a) Construir un histograma de frecuencias relativas.

b) Obtener el porcentaje medio de cambio y la mediana de cambio.

c) Verificar que se cumple la regla de Chebyshev para $k = 2$.

d) El valor extremadamente grande 38.2 fue de Albania y se debió al aumento de su reserva, y el valor -21.8 de los Países Bajos, se debe a una reducción anual de su reserva. Elimine estos valores y responda los tres puntos anteriores. Calcule el coeficiente de variación y en general, compare.

e) Una regla empírica sugiere que la desviación estándar puede ser aproximada por "un cuarto del rango". Calcular esta aproximación de s y compararla con la obtenida anteriormente.

5. La variación diaria, en porcentaje, de las acciones de mayor rendimiento en Colombia fueron (El Tiempo 11/11/2011):

-3.33	-1.38	-0.41	-2.26	-4.62	0.59	0.41	0.50	-1.06	-0.06
0.00	-0.12	0.42	0.00	0.94	0.17	0.69	0.00	-0.43	-0.63
0.00	-0.80	0.86	-1.86	0.00	-0.33	0.00	0.00	0.76	-2.95
-1.49	0.00	0.00	-1.17	-0.42	0.00	-1.00	0.00	0.00	

a) Construya un histograma de frecuencias relativas y comente sobre la forma de esta información.

b) Obtener la variación diaria media de cambio.

c) Verificar que se cumple la regla de Chebyshev para $k = 2$.

d) ¿Es el valor $-4.62\,\%$ un outlier?.

6. La reserva de petroleo, en millones de barriles, para América Latina en el 2010 se da a continuación (Fuente: OPEC. Annual Statistical Bulletin 2010/2011). Dibuje un diagrama de pastel, utilizando porcentajes.

País	Reserva
Argentina	2.505
Brasil	12.857
Colombia	1.360
Ecuador	7.206
México	11.692
Venezuela	296.501
Otros	2.760

7. La población de votantes de un país se identifican como 30 % demócratas, 28 % republicanos, 20 % socialistas, 15 % comunistas y 7 % otros. Utilice el método gráfico que usted considere más apropiado para analizar esta información.

8. En la siguiente distribución de frecuencia, x_i es el número de estudiantes que participaron en 30 talleres especiales en un colegio.

x_i	4	5	6	7	8	9	10	11
f_i	5	4	6	9	1	3	1	1

   a) Encontrar la media, la mediana, la moda y la varianza.

   b) ¿Qué porcentaje de la distribución está entre $(\bar{x} - 2s, \bar{x} + 2s)$. Concuerda esto con la regla de Chebyshev?

9. La distribución de estaturas x_i, de 125 mujeres de una comunidad se da en la tabla siguiente

x_i	1.56	1.58	1.60	1.62	1.64	1.66	1.68	1.70	1.72	1.74	1.76
f_i	3	5	8	11	15	19	21	15	14	9	5

   a) Calcule $\bar{x}$, $\xi_{0.5}$ y s^2.

   b) Construya una buena distribución de frecuencias en una escala continua y obtenga $\bar{x}$, $\xi_{0.5}$ y s^2.

   c) compare los reultados en a y b.

10. Reconstruya la distribución de frecuencias de abajo. Para su interés $n = 54$, $L = 11$ y $m_5 = 73$.

Intervalos	m_i	f_i	f_{ir}	F_i	F_{ir}
			0.111		
			0.185		
			0.259		
			0.222		
			0.148		
			0.074		
Total			0.999		

Calcule ahora media, mediana, varianza y cualquier prcentil.

11. Los puntajes de 50 estudiantes para ingresar a una Universidad fueron:

70.8	74.4	69.0	81.4	74.0	61.5	75.5	89.0	94.0	76.5
68.8	80.0	76.5	69.6	85.5	75.5	82.5	77.5	85.0	73.0
84.5	72.5	65.7	79.8	62.5	90.0	72.5	81.0	78.0	86.2
79.0	62.7	83.5	81.9	71.2	87.3	82.5	67.3	87.4	71.9
92.5	64.8	77.6	66.8	72.6	58.9	73.1	81.4	64.5	79.3

a) Construir una *buena* distribución de frecuencias.

b) Verificar que se cumple la regla de Chebyshev.

c) Obtener la mediana y los percentiles del 5 % y del 95 %.

d) ¿Cuántos estudiantes estan en el 5 % superior?

e) Mediante el método gráfico obtener la mediana y los percentiles del 5 % y del 95 %. Comparar con los obtenidos en el inciso anterior.

12. El tiempo de vida (en años) de ciertas baterías para automóviles similares se garantiza por 3 años. En una muestra de 50 tiempos de vida de estas baterías se obtuvieron los siguientes datos:

3.3	3.4	2.5	4.7	2.2	3.1	3.5	2.6	3.7	3.4
4.1	3.8	1.6	4.3	3.2	3.9	4.2	2.0	3.3	2.8
3.6	3.4	2.9	3.5	3.1	3.7	3.3	4.2	3.0	3.7
3.8	3.2	3.4	3.3	4.1	2.8	3.7	3.2	3.3	3.6
2.9	3.5	3.2	3.1	3.0	3.0	3.3	3.0	3.8	3.2

a) Encuentre la media, la mediana y la varianza.

b) Encuentre la media recortada y la de Winzor al 10 %. Compare.

c) ¿Cree usted que la garantía está bien sustentada?

d) Determine outliers, si los hay

e) Verifique la Regla de Chebyshev

13. Una regla empírica establece que *para un conjunto de datos que tiene una distribución en forma de campana, aproximadamente, el 68 % de las observaciones están en el intervalo* $(\bar{x}-s, \bar{x}+s)$*, el 95 % de las observaciones están en* $(\bar{x}-2s, \bar{x}+2s)$ *y el 97 % de las observaciones están en* $(\bar{x}-3s, \bar{x}+3s)$. ¿Será que los datos del ejercicio anterior satisfacen esta regla?

14. 500 resultados graficados de los exÃ¡menes de admisión de una universidad muestran una distribución de frecuencia relativa en forma de campana. Si se

conoce que estos resultados tienen una media de 70 y desviación estándar de 9, aproximadamente ¿cuantos de estos resultados espera usted que esten en el intervalo $(52, 88)$?

15. El tiempo medio de vida de ciertas baterías de litio es de 1.5 años y desviación estándar de 0.3 años. ¿Que porcentaje de tiempos de vida estarán al menos entre 0.9 y 2.1 años?

16. En un estudio acerca de los efectos del tabaquismo sobre los patrones de sueño se obtuvieron los siguientes datos (tiempo en minutos):
Fumadores:

45.6	36.4	11.8	25.2	50.7	22.2	54.0	50.1	41.8	62.2
51.2	67.3								

No fumadores:

35.0	31.6	27.8	28.6	24.1	30.4	31.8	36.9	15.9	40.6
37.5	27.4	33.9	31.2	22.1	18.5	38.4	12.7	36.4	17.5

Use los diagramas de cajas y bigotes para comparar estos dos conjuntos de datos. ¿Que le dicen los coeficientes de variación?

17. La rentabilidad efectiva mensual y semestral de 38 fondos de inversión abiertas de sociedades fiduciarias en Colombia (LA REPUBLICA 13-14 de julio de 2013), se dan a continuación en porcentajes.

1 mes	6 meses	1 mes	6 meses	1 mes	6 meses	1 mes	6 meses
-1.83	3.15	-1.82	2.97	-2.91	2.43	-0.84	2.51
0.28	2.73	1.59	2.93	-0.39	4.92	-0.89	2.51
0.10	2.66	1.06	0.84	0.49	4.28	1.48	3.43
-5.86	-0.84	-3.06	2.28	-4.24	2.16	-4.41	1.34
-4.45	2.98	-4.47	2.00	-3.51	2.37	-5.68	2.05
-5.27	2.05	-2.32	2.62	-0.28	3.43	1.36	4.23
-4.48	2.31	-9.50	-3.24	-1.24	3.09	3.89	4.26
-3.34	2.61	-6.40	1.86	-2.96	2.57	0.18	3.19
-6.69	2.47	1.64	3.40	-2.80	2.55	0.12	3.34
-2.90	3.17	-21.19	-11.96				

En cada caso encuentre:

a) Media, mediana, varianza y tres percentiles cualesquiera.

b) Verifique la regla de Chebyshev.

c) Determine *outliers* si los hay.

d) Dibuje histogramas de frecuencia relativas para los dos conjuntos de datos.

e) Compare la rentabilidad efectiva mensual con la rentabilidad efctiva semestral

18. Para los datos del ejercicio anterior construya los histogramas de frecuencias relativas y responda las preguntas anteriores para datos agrupados.

19. El calor específico de ciertas sustancias a 25^0C y presión constante se da a continuación en $\frac{kj}{k_g k}$:

1.005	2.093	0.5203	0.2253	1.694	0.4781	0.6496	0.6228
0.8439	0.806	1.744	1.527	0.8237	5.193	1.649	14.30
1.663	0.248	2.226	1.039	1.030	1.040	0.9992	0.918
1.664	1.669	1.531	0.5415	0.8334	0.9291	1.865	0.1583

a) Calcule media, mediana, varianza y tres percentiles cualesquiera.

b) Verifique la regla de Chebyshev.

c) Obtenga una *buena* distribución de frecuencias(completa).

d) Calcule media, mediana, moda, varianza y los tres percentiles del punto a) para los datos agrupados y compare.

e) Obtenga los tres percentiles anteriores por el método gráfico.

f) Determine mediante un diagrama de cajas y bigotes (box-plots) los posibles tipos de **outliers** si los hay.

20. La media de 15 observaciones es 8 y la varianza 12. Se encuentra que la observación 16 es 4. Halle la media y la varianza de las 16 observaciones.

Capítulo 2

Elementos de probabilidad

2.1. Introducción

Una de las herramientas fundamentales de la estadística es la teoría de probabilidad, la cual tuvo su comienzo formal con los juegos de azar en el siglo XVII. Estos, como su nombre lo indica, implican actividades como lanzadas de una moneda, de un dado, la extracción de una carta de una baraja, etc. en los cuales el resultado de estos experimentos es incierto. Sin embargo, hay cierta regularidad en la distribución de los resultados, aunque los mismos no sean completamente predecibles.

Se conoce, por ejemplo, que en muchas tiradas de una moneda simétrica, la mitad de las pruebas resulten caras. En este largo plazo de regularidad predecible es que las casas de juegos son capaces de entrar en negocios.

Un tipo de incertidumbre y regularidad a largo plazo ocurre a menudo en ciencia experimental. Por ejemplo, en genética es incierto si el nacimiento es niño o niña, pero en una gran racha de nacimientos se conoce, aproximadamente, qué porcentaje de nacimientos serán niños y qué porcentaje serán niñas.

En la siguiente sección se dan algunos conceptos y resultados necesarios para presentar posteriormente la teoría básica de la probabilidad.

2.2. Conceptos básicos

Definición 2.1. Experimento: es un fenómeno por el cual se hace una observación. Se dice aleatorio si un resultado particular no se puede predecir con certeza. Los experimentos se representarán por ε.

Ejemplo 2.1. Son ejemplos de experimentos,

- ε_1: Lanzar un dado y observar el número que aparece en la cara superior.
- ε_2: Lanzar dos monedas y observar las caras superiores.
- ε_3: Observar el tiempo de duración(en horas) de una bombilla eléctrica.
- ε_4: Observar la estatura x y el peso y de los estudiantes del curso de Probabilidad.
- ε_5: Sacar un artículo de un proceso de producción y observar el nivel de calidad.

Se observa en cada uno de estos experimentos, que existe un conjunto de resultados posibles para cada uno de ellos. Esto es lo que se llama *espacio muestral*.

Definición 2.2. Espacio muestral. El conjunto de todos los resultados posibles de un experimento se llama *espacio muestral* o *espacio de sucesos* y se denota con la letra mayúscula S o por Ω. En este texto se emplea preferiblemente Ω.

En los ejemplos anteriores los espacios muestrales son,

- $\Omega = \{1, 2, 3, 4, 5, 6\}$ para ε_1.
- $\Omega = \{cc, cs, sc, ss\}$ para ε_2, dode c denota cara y s sello.
- $\Omega = \{t|0 \leq t < \infty\}$ para ε_3, donde $t < \infty$ significa que en algún tiempo finito la bombilla se funde.
- $\Omega = \{(x, y)|x > 1.5,\ y > 45\}$, x está dado en metros y y en kilos, para ε_4.
- $\Omega = \{\text{b,m}\}$, si solo se aceptan las condiciones de bueno(b) o malo(m), en ε_5.

Un espacio muestral Ω se dice **discreto** si es finito o enumerable. Un conjunto que es finito o contablemente infinito se dice **contable**. Cuando el espacio incluye resultados que pueden asumir cualquier valor en un intervalo se dice que es un **espacio muestral continuo**. Ω de $\varepsilon_1, \varepsilon_2$ y ε_5 son discretos, mientras que los de

ε_3 y ε_4 son continuos.

En la realización de un experimento se está interesado en la ocurrencia de algunos subconjuntos del espacio muestral Ω, estos se llaman **eventos o sucesos** y se denotan con las primeras letras mayusculas del alfabeto, A, B, C, D, E, etc. Sin embargo cuando se está interesado en n eventos se empleará una sola de estas letras con subindice, como $E_1, E_2, ..., E_n$.

Definición 2.3. Evento. Dado un experimento, ε, cualquier conjunto caracterizado por un rasgo común se llama *evento* o *suceso*.

En los experimentos anteriores se consideran los siguientes eventos:

- A es el evento en que aparezca un número par en ε_1, entonces $A = \{2, 4, 6\}$.
- B es el evento en que aparezca una cara en ε_2, entonces $B = \{cs, sc\}$.
- C es el evento en que la bombilla dure menos de 200 horas en ε_3, entonces $C = \{t|0 < t < 200\}$.
- D es el evento en que la estatura sea menor que $1.70m$ en ε_4, entonces $D = \{(x, y)|1.5 < x < 1.7,\ y > 45\}$.
- E es el evento en que el artículo sea bueno en ε_5, entonces $E = \{\text{b}\}$.

Es de anotar que un evento ocurre si uno cualquiera de sus resultados sucede. Además, como los eventos son conjuntos, las operaciones de unión, intersección, complemento y diferencia son válidas.

Si E y F son eventos entonces:

a) $E \cup F$ es el evento que ocurre si y solo si E o F ocurren o ambos ocurren.

b) $E \cap F$ es el evento que ocurre si y solo si E ocurre y F ocurre.

c) E^c es el evento que ocurre si y solo si E no ocurre.

d) $F - E$ es el evento que ocurre si y solo si F ocurre y E no.

Definición 2.4. Evento elemental. Un evento se dice *elemental* si este contiene exactamente un resultado del experimento. $E = \{3\}$ en ε_1 es un evento elemental.

Definición 2.5. Eventos mutuamente excluyentes. Dos eventos E y F se dicen *mutuamente excluyentes* si $E \cap F = \phi$. Los eventos $E_1, E_2, \cdots$ se dice que son *mutuamente excluyentes* si son mutuamente excluyentes a pares; es decir, si $E_i \cap E_j = \phi$ para $i \neq j$.

La probabilidad surge antes de la creación de la teoría de conjuntos, sin embargo en 1933 A. N. Kolmogorov propone un sistema de axiomas(teoría) basado en conjuntos que involucra todos los conceptos conocidos de probabilidad. Una definición fundamental es la de una σ- *álgebra*.

Definición 2.6. **σ- álgebra.** Sea $\mathcal{F}$ un conjunto no vacío de subconjuntos de Ω, se dice que $\mathcal{F}$ es una σ - álgebra sobre Ω si:

1. si $E \in \mathcal{F}$, entonces $E' \in \mathcal{F}$.

2. si $E_1, E_2, \cdots$ es una colección finita o infinita de conjuntos de $\mathcal{F}$, entonces $\bigcup_n E_n \in \mathcal{F}$.

De esta definición es claro que $\Omega \in \mathcal{F}$, pues $E \cup E' = \Omega$ para cualquier conjunto E, y adémas ϕ también pertenece a $\mathcal{F}$.

Observación.
Los elementos de $\mathcal{F}$ se llamarán eventos, Ω se llamará evento seguro , y ϕ evento imposible.

Decir que un evento E "ocurre "significa que el resultado obtenido al realizar el experimento es un elemento de E.

2.3. Axiomas de probabilidad

Una de los propósitos de la ciencia es el de predecir y describir eventos en el mundo en el cual vivimos. Una manera mediante la cual se da esto es en la construcción de modelos matemáticos, los cuales describen adecuadamente el mundo real. Por ejemplo, la ecuación $v = \frac{e}{t}$ expresa una relación entre las cantidades v, e y t, y es un modelo matemático. La razón para mencionar esta ecuación es que la teoría de la probabilidad hace algo similar: se construye un modelo de probabilidad que se usa para describir eventos en el mundo real. Por ejemplo, se puede desear encontrar una ecuación para predecir el sexo de cada nacimiento en una cierta localidad. Tal ecuación sería muy compleja y no ha sido encontrada. Sin embargo, un modelo de probabilidad se puede construir, el cual aunque no ayude mucho con el tratamiento de un nacimiento individual, es ampliamente útil en los trabajos con grupos de nacimientos. De aquí que se puede postular un número p que representa la probabilidad que un nacimiento sea niño. De esta probabilidad fundamental se pueden resolver preguntas como: ¿Cuál es la probabilidad que en 20 nacimientos al menos 8 sean niños? ¿Cuál es la probabilidad que haya 5 nacimientos niños consecutivos en los próximos 12?.

Para responder estas preguntas y muchas similares, se debe desarrollar un *modelo de probabilidad idealizado*.

Definición 2.7. Función de probabilidad. Sea Ω un espacio muestral y $\boldsymbol{\mathcal{F}}$ una σ-álgebra sobre Ω. Una función de probabilidad $P(\cdot)$ es una función de conjunto con dominio en $\boldsymbol{\mathcal{F}}$ y recorrido en el intervalo $[0, 1]$, la cual satisface los siguientes axiomas:

a) $P(E) \geq 0$ para cualquier evento $E \in \boldsymbol{\mathcal{F}}$.

b) $P(\Omega) = 1$

c) Si $E_1, E_2, ..., E_n, ...$, es una sucesión de eventos mutuamente excluyentes en $\boldsymbol{\mathcal{F}}$, y si

$$E_1 \cup E_2 \cup ... \cup E_n \cup ... = \bigcup_{i=1}^{\infty} E_i$$

entonces,

$$P\big(\bigcup_{i=1}^{\infty} E_i\big) = \sum_{i=1}^{\infty} P(A_i)$$

$P(E)$ o $P[E]$ se lee *la probabilidad del evento E* o *la probabilidad de que el evento E ocurra*.

La terna $(\Omega, \boldsymbol{\mathcal{F}}, P)$ se llama *espacio de probabilidad*, y es un término único que da un camino expedito para asumir le existencia de las tres componentes en su notación, las cuales están relacionadas: $\boldsymbol{\mathcal{F}}$ es una colección de subconjuntos de Ω y $P(.)$ es una función que tiene a $\boldsymbol{\mathcal{F}}$ como su dominio.

Ejemplo 2.2. Sea $\Omega = \{1, 2, 3, 4\}$ y $\boldsymbol{\mathcal{F}} = \{\Omega, \phi, \{1, 3\}, \{2, 4\}\}$ la función P definida sobre $\boldsymbol{\mathcal{F}}$ por

$$P(E) = \begin{cases} 0, & \text{si } E = \phi, \\ \frac{1}{3}, & \text{si } E = \{1,3\}, \\ \frac{2}{3}, & \text{si } E = \{2,4\}, \\ 1, & \text{si } E = \Omega \end{cases}$$

es una *medida* de probabilidad.

Definición 2.8. Evento nulo. Sea $(\Omega, \boldsymbol{\mathcal{F}}, P)$ un espacio de probabilidad, cualquier evento E con probabilidad cero se llama evento nulo.

Ejemplo 2.3. Sea $\Omega = \{1,2,3\}$, $\boldsymbol{\mathcal{F}} = \{\Omega, \phi, \{1\}, \{2,3\}\}$ y sea P la medida de probabilidad siguiente;

$$P(E) = \begin{cases} 1, & \text{si } 2 \in E, \\ 0, & \text{si } 2 \notin E \end{cases}$$

Entonces son eventos nulos $\{1\}$ y ϕ.

2.4. Propiedades de $P(\cdot)$

En cada uno de los teoremas sigiuentes, sea $(\Omega, \boldsymbol{\mathcal{F}}, P)$ un espacio de probabilidad

Teorema 2.1. $P(\phi) = 0$.

Demostración. Sea $E_1 = \phi, E_2 = \phi, E_3 = \phi, ...$, entonces

$$\phi = \bigcup_{i=1}^{\infty} E_i.$$

Luego

$$P(\phi) = P\big(\bigcup_{i=1}^{\infty} E_i\big) = \sum_{i=1}^{\infty} P(\phi)$$

y en consecuencia

$$0 = \sum_{i=2}^{\infty} P(\phi), \text{ lo que implica que } P(\phi) = 0.$$

Teorema 2.2. Si $E_1, E_2, ..., E_n$, son eventos mutuamente excluyentes en $\boldsymbol{\mathcal{F}}$, entonces,

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{i=1}^{n} P(E_i).$$

Demostración. Sea $E_{n+1} = E_{n+2} = \cdots = \phi$, entonces

$$\bigcup_{i=1}^{n} E_i = \bigcup_{i=1}^{\infty} E_i,$$

y por tanto,

$$P\left(\bigcup_{i=1}^{n} E_i\right) = P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{n} P(E_i) + \sum_{i=n+1}^{\infty} P(E_i) = \sum_{i=1}^{n} P(E_i).$$

Teorema 2.3. Si E es un evento en $\boldsymbol{\mathcal{F}}$, entonces,

$$P(E') = 1 - P(E)$$

Demostración. Como E y E' son mutuamente excluyentes y $E \cup E' = \Omega$, se tiene que:

$$P(\Omega) = P(E \cup E') = P(E) + P(E'), \text{pero } P(\Omega) = 1.$$

Por tanto,

$$P(E') = 1 - P(E)$$

Teorema 2.4. Si E y F están en $\boldsymbol{\mathcal{F}}$, entonces

$$P(E) = P(E \cap F) + P(E \cap F')$$

y

$$P(E - F) = P(E \cap F') = P(E) - P(E \cap F)$$

Demostración.

$$E = E \cap (F \cup F') = (E \cap F) \cup (E \cap F').$$

Entonces,

$$P(E) = P(E \cap F) + P(E \cap F').$$

Además

$$P(E \cap F') = P(E - F) = P(E) - P(E \cap F).$$

Teorema 2.5. Para dos eventos E y F cualesquiera en $\boldsymbol{\mathcal{F}}$, se tiene que:

$$P(E \cup F) = P(E) + P(F) - P(E \cap F).$$

Para tres eventos E, F y G cualesquiera en $\boldsymbol{\mathcal{F}}$, se tiene que

$$\begin{aligned}&P(E \cup F \cup G) = \\ &\ P(E) + P(F) + P(G) - P(E \cap F) - P(E \cap G) - P(F \cap G) + P(E \cap F \cap G).\end{aligned}$$

En general, para eventos $E_1, E_2, \cdots, E_n$ en $\boldsymbol{\mathcal{F}}$, se tiene que

$$\begin{aligned}P\left(\bigcup_{i=1}^{n} E_i\right) &= \sum_{i=1}^{n} P(E_i) - \sum_{i\ <}\sum_{j} P(E_i \cap E_j) + \\ &\sum_{i\ <}\sum_{j\ <}\sum_{k} P(E_i \cap E_j \cap E_k) - \cdots + (-1)^{n+1} P(E_1 \cap E_2 \cdots \cap E_n)\end{aligned}$$

Demostración. Para esta prueba se descompone $E \cup F$ en dos conjuntos mutuamente excluyentes:

$$E \cup F = (E \cup F) \cap \Omega = (E \cup F) \cap (F \cup F') = F \cup (E \cap F').$$

Obsérvese que F y $E \cap F'$ son mutuamente excluyentes. Luego,

$$P(E \cup F) = P(F \cup (E \cap F')) = P(F) + P(E \cap F') = P(F) + P(E) - P(E \cap F).$$

La demostración en el caso general se hace por inducción matemática.

Teorema 2.6. Si E y F son dos eventos en $\boldsymbol{\mathcal{F}}$ y $E \subset F$, entonces

$$P(E) \leq P(F).$$

Demostración. El evento F se puede descomponer como la unión de dos eventos mutuamente excluyentes. En efecto $F = F \cap \Omega = F \cap (E \cup E') = (F \cap E) \cup (F \cap E')$. Luego,

$$P(F) = P(F \cap E) + P(E' \cap F) = P(E) + P(E' \cap F).$$

Como $P(E' \cap F) \geq 0$, entonces

$$P(F) = P(E) + P(E' \cap F)) \geq P(E)$$

Teorema 2.7. Desigualdad de Boole. Si $E_1, E_2, \cdots, E_n$, están en $\boldsymbol{\mathcal{F}}$, entonces

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) \leq P(E_1) + \cdots + P(E_n).$$

Demostración. (Usar Inducción Matemática).

El siguiente ejemplo ilustra algunas de estas propiedades y es además una introducción para la siguiente sección.

Ejemplo 2.4.

i) Una moneda es insesgada o legal si siempre que se lanza, la probabilidad de obtener cara es la misma de obtener sello. Si se lanza una moneda legal y se mira la parte superior, ¿cuál es la probabilidad de obtener cara?

   Solución. Aquí $\Omega = \{c, s\}$ y además $\Omega = \{c\} \cup \{s\}$ y $\{c\} \cap \{s\} = \phi$, luego $P(\Omega) = P(\{c\} \cup \{s\}) = P(\{s\}) + P(\{c\})$. Por la definición de probabilidad se sabe que $P(\Omega) = 1$, luego $P(\{s\}) + P(\{c\}) = 1$. Como la moneda es legal $P(\{c\}) = P(\{s\})$, así que

$$P(\{c\}) + P(\{c\}) = 2P(\{c\}) = 1$$

   por lo tanto $P(\{c\}) = \frac{1}{2}$.

ii) Supóngase ahora que la moneda lanzada no es legal, de manera que es tres veces más probable salir sello que cara, ¿cuál es la probabilidad de obtener cara?

   Solución

   $P(\{s\}) + P(\{c\}) = 1$, ahora $P(\{s\}) = 3P(\{c\})$ por lo tanto $P(\{c\}) + 3P(\{c\}) = 1$, así $P(\{c\}) = \frac{1}{4}$ y además $P(\{s\}) = \frac{3}{4}$.

2.5. Espacios muestrales finitos

Para un gran número de experimentos, existe un número finito de resultados, por ejemplo N y pueden ser igualmente posibles o no. Si los resultados son igualmente posibles, es relativamente fácil calcular la probabildad de un evento. En la siguiente subsección se discute este caso.

2.5.1. Espacios muestrales finitos con resultados igualmente posibles

Si el espacio muestral Ω consta de N resultados igualmente posibles, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ entonces

$$\Omega = \bigcup_{i=1}^{N} \{\omega_i\}$$

de donde

$$1 = P(\Omega) = P\left(\bigcup_{i=1}^{N} \{\omega_i\}\right) = \sum_{i=1}^{N} P(\{\omega_i\})$$

pero $P(\{\omega_i\}) = p$ para $i = 1, 2, \ldots, N$ entonces

$$1 = \sum_{i=1}^{N} p = Np$$

por lo tanto

$$p = P(\{\omega_i\}) = \frac{1}{N}$$

Teorema 2.8. Concepto a priori o clásico de probabilidad. Sea Ω el espacio muestral de un experimento que tiene N resultados igualmente posibles, entonces para cualquier evento E que contiene n_E elementos de Ω,

$$P(E) = \frac{n_E}{N}.$$

Demostración.
Sea $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ con estos resultados igualmente posibles, y sea $E = \left\{\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_{n(E)}}\right\}$ donde ω_{i_j} es algún ω_k de Ω, luego

$$P(E) = P\left(\bigcup_{j=1}^{n_E} \{\omega_{i_j}\}\right) = \sum_{j=1}^{n_E} P\left(\omega_{i_j}\right) = \sum_{j=1}^{n_E} \frac{1}{N} = \frac{n_E}{N}.$$

Ejemplo 2.5.

I) Considérese el experimento de lanzar un dado legal y observar la cara superior. Cualquiera de las seis caras puede aparecer arriba y estos resultados son mutuamente excluyentes, puesto que dos caras no pueden aparecer arriba simultáneamente. Además, el dado es legal pues es un cubo simétrico, así que los seis resultados son igualmente posibles.

Suponiendo que se quiere encontrar la probabilidad del evento E obtener un número impar, luego tres de los seis resultados posibles tienen el atributo de ser impar, es decir, $E = \{1, 3, 5\}$ y puesto que $n = 6$, entonces $P(E) = \frac{3}{6} = 1/2$.

Similarmente, si E es el evento de obtener un número primo impar, entonces $E = \{3, 5\}$ y por consiguiente $P(E) = \frac{2}{6} = 1/3$.

II) En el experimento de sacar una carta de una baraja inglesa. Sea F el evento de obtener una pica. Como de las 52 cartas 13 son picas, entonces $n_F = 13$ y de esta manera $P(F) = \frac{13}{52} = 1/4$.

Sea ahora G el evento de obtener una letra. Puesto que hay 16 letras en toda la baraja, entonces $P(G) = \frac{16}{52} = 4/13$.

La aplicación de la definición de estos ejemplos sencillos es obvia, pero se debe tener cuidado en que los resultados deben ser mutuamente excluyentes, e igualmente posibles. Obsérvese los siguientes ejemplos:

a) En el experimento de lanzar una moneda legal dos veces (o dos monedas simultáneamente) se quiere obtener la probabilidad del evento A de obtener dos sellos.

Una persona desprevenida puede razonar de la siguiente manera: Hay tres resultados posibles para las dos tiradas $\Omega = \{cc, cs, ss\}$. Uno de estos resultados tiene el atributo deseado, es decir, el conjunto $E = \{ss\}$, y por tanto $P(E) = \frac{1}{3}$.

Este razonamiento no es correcto, porque los tres resultados no son igualmente posibles, ya que el resultado de una cara y un sello puede ocurrir de dos maneras, puesto que puede aparecer la cara en la primera tirada y el sello en la segunda y viceversa. Luego hay cuatro resultados igualmente posibles así que $\Omega = \{cc, cs, sc, ss\}$. Si $E = \{ss\}$, entonces $P(E) = \frac{1}{4}$, que es la probabilidad correcta.

b) Considérese de nuevo el experimento de sacar una carta de una baraja y sea F el evento de que la carta sacada sea un as o una pica.

En la enumeración de los resultados favorables se puede contar que hay 4 ases y 13 picas, y por tanto hay 17 resultados con el atributo deseado, o sea que,

$P(F) = \frac{17}{52}$. Este resultado es claramente incorrecto, porque los 17 resultados no son mutuamente excluyentes, ya que el as de pica es as y a la vez pica. Luego, hay 16 resultados con el atributo deseado y la probabilidad correcta es $P(F) = \frac{16}{52} = 4/13$.

Ejemplo 2.6. Se lanza una moneda legal tres veces (o lanzar tres monedas de una vez). Halle la probabilidad de obtener a lo sumo una cara.
Solución.

$$\Omega = \{ccc, ccs, csc, scc, css, scs, ssc, sss\}$$

Sea E el evento de obtener a lo sumo una cara, entonces $E = \{css, scs, ssc, sss\}$, los 8 resultados de Ω son igualmente posibles y además $n(E) = 4$, luego la probabilidad de E es

$$P(E) = \frac{4}{8} = \frac{1}{2}$$

Ejemplo 2.7. Un elevador con dos pasajeros se detiene en el 2^{do}, 3^{er} y 4^{to} piso. Si es igualmente posible que un pasajero baje en cualquier piso, ¿cual es la probabilidad que los pasajeros bajen en pisos diferentes?

Solución. Supongase que a y b denotan los pasajeros, un resultado de Ω es por ejemplo, a_3 y b_2 que significa que a se baja en el tercer piso y b se baja en el segundo piso, luego $\Omega = \{a_2b_2, a_2b_3, a_2b_4, a_3b_2, a_3b_3, a_3b_4, a_4b_2, a_4b_3, a_4b_4\}$ y son igualmente posibles. Sea E el evento de que los pasajeros se bajen en el mismo piso entonces, $E = \{a_2b_2, a_3b_3, a_4b_4\}$ entonces $P(E) = \frac{3}{9} = \frac{1}{3}$, así que $P(E^c) = 1 - P(E) = \frac{2}{3}$ es la probabilidad de que los pasajeros se bajen en pisos diferentes.

Ejemplo 2.8. Un número se selecciona al azar de $N = \{1, 2, 3\ldots, 1000\}$. ¿Cuál es la probabilidad que el número sea divisible por 3 o 5?

Solución. El número de resultados igualmente posibles es 1000. Sea E el evento de obtener un número divisible por 3 y F el evento de obtener un número divisible por 5. Se debe hallar $P(E \cup F)$.

$$\begin{aligned}
E &= \left\{x | x = 3m,\ 1 \leq m \leq \left[\frac{1000}{3}\right] = 333\right\}, \quad n(E) = 333 \\
F &= \left\{x | x = 5m,\ 1 \leq m \leq \left[\frac{1000}{5}\right] = 200\right\}, \quad n(F) = 200 \\
E \cap F &= \left\{x | x = 15m,\ 1 \leq m \leq \left[\frac{1000}{15}\right] = 66\right\}, n(E \cap F) = 66
\end{aligned}$$

así que

$$\begin{aligned} P(E \cup F) &= P(E) + P(F) - P(E \cap F) \\ &= \frac{333}{1000} + \frac{200}{1000} - \frac{66}{1000} \\ &= 0,467 \end{aligned}$$

En este ejemplo, $[a]$ es la parte entera de a, la cual se define como el mayor entero menor o igual que a(ver apendice B).

Ejemplo 2.9. Una urna contiene 6 bolas blancas y 9 bolas rojas. Se eligen al *azar* 5 bolas. Calcule la probabilidad que entre las bolas extraidas hayan 2 bolas blancas y 3 rojas.

Solución. Cuando se dice que la extracción es al azar significa que no existe una tendencia a tomar un color determinado y que cada bola tiene la misma posibilidad de salir. Si E es el evento de obtener 2 bolas blancas y 3 rojas, entonces $n_E = \binom{6}{2}\binom{9}{3}$(ver apendice A).

El espacio muestral, Ω tiene entonces, $N = \binom{15}{5}$ resultados posibles, luego

$$P(E) = \frac{n_E}{N} = \frac{\binom{6}{2}\binom{9}{3}}{\binom{15}{5}} = 0.4196$$

2.5.2. Espacios muestrales finitos sin resultados igualmente posibles

Si los resultados no son igualmente posibles, las cosas no son tan simples como en el caso anterior, pero los valores de $P(E)$ se pueden encontrar para cada uno de los $2^{N(\Omega)}$ eventos. Sea $\Omega = \{\omega_1, \omega_2, \cdots, \omega_N\}$ y se supone que $p_i = P(\{\omega_i\})$, para $i = 1, 2, \cdots, N$. Puesto que $P(\Omega) = 1$ entonces,

$$1 = P(\bigcup_{j=1}^{N} \omega_j) = \sum_{j=1}^{N} P(\omega_j) = \sum_{j=1}^{N} p_j.$$

Para cualquier evento E se define como $P(E) = \sum_j p_j$, donde la suma es sobre aquellos ω_j que están en E.

Ejemplo 2.10. Un experimento tiene N resultados posibles $w_1, w_2, \cdots, w_N$, donde se conoce que el resultado s_{j+1} es dos veces más posible que el resultado

$w_j, j = 1, 2, \cdots, N$. Hallar $P(A_k)$, donde $A_k = \{w_1, w_2, \cdots, w_k\}$.

Solución. Por hipótesis $p_{j+1} = 2p_j$, de manera que $p_2 = 2p_1, p_3 = 2p_2 = 2^2p_1, p_4 = 2p_3 = 2^3p_1, \cdots, p_N = 2^{N-1}p_1$, entonces,

$$1 = \sum_{j=1}^{N} p_j = \sum_{j=1}^{N} 2^{j-1}p_1 = p_1 \sum_{j=1}^{N} 2^{j-1} = p_1(2^N - 1)$$

de ahí que,

$$p_1 = \frac{1}{2^N - 1}$$

Ahora,

$$p(A_k) = \sum_{j=1}^{k} p_j = \frac{1}{2^N - 1} \sum_{j=1}^{k} 2^{j-1} = \frac{2^k - 1}{2^N - 1}.$$

En este ejemplo el resultado fué fácil de obtener porque el evento seleccionado toma los primeros k resultados del espacio muestral y hay además una ley de formación para las probabilidades de los eventos elementales, sin embargo cuando los resultados no son igualmente posibles o no son mutuamente excluyentes el concepto clásico no es aplicable y, en este caso, el experimento se repite *varias veces* y un *gran número de veces*, para luego tomar la estabilización de la frecuencia relativa del evento (cuando se da) como la probabilidad del evento(concepto a posteriori o frecuencial).

Definición 2.9. Concepto a posteriori o frecuencial de probabilidad. Si $f_n(E)$ es la estabilización de la frecuencia relativa del evento E en la n-ésima repetición de un experimento, entonces

$$P(E) = \lim_{n\to\infty} f_n(E).$$

Ejemplo 2.11. Si el 25 % de la población de una ciudad lee el periódico A, 20 % lee el periódico B, 15 % lee el C, 10 % lee A y B, 7 % lee A y C, 5 % lee B y C, y 3 % lee los tres periódicos. ¿Cuál es la probabilidad que una persona seleccionada de esta población no lea ninguno de los tres periódicos?

Solución. Sean E, F y G eventos, donde cada uno de ellos indican que el individuo seleccionado lea el periódico A, el individuo lea el periódico B y que el individuo lea el periódico C, respectivamente.

Los porcentajes de lectura equivalen a la frecuencias relativas de los eventos, así que $P(E) = 0,25$; $P(F) = 0,2$; $P(G) = 0,15$; $P(E \cap F) = 0,10$; $P(E \cap G) = 0,07$; $P(F \cap G) = 0,05$ y $P(E \cap F \cap G) = 0,03$.

La probabilidad que el individuo no lea ni A ni B ni C, es

$$P(E^c \cap F^c \cap G^c)$$

pero

$$P(E^c \cap F^c \cap G^c) = P([E \cup F \cup G]^c) = 1 - P(E \cup F \cup G)$$

Como,

$$\begin{aligned} P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(E \cap F) - P(E \cap G) \\ - P(F \cap G) + P(E \cap F \cap G) \\ = 0,25 + 0,2 + 0,15 - 0,1 - 0,07 - 0,05 + 0,03 = 0,41 \end{aligned}$$

Luego, la probabilidad que un individuo seleccionado al azar no lea ninguno de los tres periódicos es

$$P(E^c \cap F^c \cap G^c) = 1 - P(E \cup F \cup G) = 1 - 0,41 = 0,59.$$

Se ha hablado de "aleatoriedad"pretendiendo con esto que cada resultado de un experimento tiene la misma oportunidad de selección. Los dos caminos básicos de tomar una muestra son: con reemplazamiento y sin reemplazamiento. Por ejemplo, si un experimento consiste en sacar bolas de una urna y si se toma una muestra, se dice que esta es con reemplazamiento si cada vez que se saca una bola se retorna a la urna antes de sacar la próxima; y es sin reemplazamiento si la bola no es retornada.

Ejemplo 2.12.

a) Supóngase que una urna contiene M bolas numeradas de 1 a M, donde las primeras K bolas son defectuosas y el resto $M - K$ no lo son. El experimento consiste en sacar n bolas de la urna. Sea A_k el evento de que en la muestra de n bolas haya exactamente k defectuosas. Hallar $P(A_k)$.

   Solución. $\Omega = \{(w_1, w_2, \cdots, w_n) : w_j =$ número de la bola tomada en la *j-ésima* sacada$\}$.

   Si el muestreo es con reemplazamiento, el número de resultados de Ω es $N(\Omega) = M^n$ y A_k es aquel subconjunto de Ω para el cual exactamente k de las w_j son bolas numeradas de 1 a K, inclusive. Estas K bolas numeradas deben caer en algún subconjunto de k posiciones del número total de las n posiciones posibles. Hay $\binom{n}{k}$ caminos de seleccionar las k posiciones para las bolas numeradas de 1 a K. Para cada una de las $\binom{n}{k}$ posiciones diferentes hay $K^k(M-K)^{n-k}$ *n-tuplas* diferentes, por tanto:

$$P(A_k) = \frac{\binom{n}{k} K^k(M-K)^{n-k}}{M^n} \tag{2.1}$$

Si el muestreo es sin reemplazamiento, para cada una de las $\binom{n}{k}$ posiciones diferentes hay $(K)_k(M-K)_{n-k}$ *n-tuplas* diferentes, donde

$$(N)_k = N(N-1)\cdots(N-(k-1)).$$

Luego,

$$P(A_k) = \frac{\binom{n}{k} (K)_k(M-K)_{n-k}}{(M)_n} \tag{2.2}$$

Puede demostrarse que este último resultado es:

$$P(A_k) = \frac{\binom{K}{k}\binom{M-K}{n-k}}{\binom{M}{n}} \tag{2.3}$$

b) Obtener la probabilidad de que en una mano de 10 cartas, 5 sean picas.

Solución. Una baraja consta de 52 cartas, de las cuales $K = 13$ son picas. Se quiere encontrar $P(A_5)$, luego $k = 5$ y $n = 10$. A_5 denota el evento de sacar 5 picas, entonces

$$P(A_5) = \frac{\binom{13}{5}\binom{52-13}{10-5}}{\binom{52}{10}} = 0.0468$$

2.6. Continuidad de la función P

Una función $f:\mathbb{R}\longrightarrow\mathbb{R}$ es continua en a si $\lim\limits_{x\to a} f(x) = f(a)$, f es continua en $\mathbb{R}$ si es continua en todo $a\in\mathbb{R}$. Esta definición es equivalente al criterio secuencial, $f:\mathbb{R}\longrightarrow\mathbb{R}$ es continua en $\mathbb{R}$ si y solo si para una sucesión $\{x_n\}_{n\geq 1}$ convergente en $\mathbb{R}$

$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right).$$

Definición 2.10. Sea $\{E_n\}_{n\geq 1}$ una sucesión de eventos.

a) $\{E_n\}$ se dice creciente si $E_1\subset E_2\subset\cdots\subset E_n\subset E_{n+1}\subset\ldots$

b) $\{E_n\}$ se dice decreciente si $E_1\supset E_2\supset\cdots\supset E_n\supset E_{n+1}\supset\ldots$

c) la sucesión $\{E_n\}$ se dice monótona si $E_1\subseteq E_2\subseteq\cdots\subseteq E_n\subseteq\ldots E_{n+1}\subseteq\ldots$, ó, $E_1\supseteq E_2\supseteq\cdots\supseteq E_n\supseteq E_{n+1}\supseteq\ldots$

Definición 2.11. Sea $\{E_n\}$ es una sucesión monótona, se define el límite de E_n por

$$\lim_{n\to\infty} E_n = \begin{cases} \bigcup_{n=1}^{\infty} E_n, & \text{si } E_n \text{ es creciente.}\\ \bigcap_{n=1}^{\infty} E_n, & \text{si } E_n \text{ es decreciente.}\end{cases}$$

Teorema 2.9. Continuidad de P. Sea $\{E_n\}$ una sucesión de eventos tal que cuando $n\to\infty$, E_n es creciente o E_n es decreciente. Entonces

$$P(\lim_{n\to\infty} E_n) = \lim_{n\to\infty} P(E_n).$$

Demostración. Se supone que E_n es creciente, entonces

$$\lim_{n\to\infty} E_n = \bigcup_{n=1}^{\infty} E_n$$

Pero el miembro derecho de esta igualdad se puede escribir como

$$\bigcup_{n=1}^{\infty} E_n = E_1\cup(E_1^c\cap E_2)\cup(E_1^c\cap E_2^c\cap E_3)\cup\cdots.$$

Luego,

$$\begin{aligned}
P(\lim_{n\to\infty} E_n) &= P(E_1 \cup (E_1^c \cap E_2) \cup (E_1^c \cap E_2^c \cap E_3) \cup \cdots) \\
&= P(E_1 \cup (E_2 - E_1) \cup (E_3 - E_2) \cup \cdots \cup (E_n - E_{n-1}) \cup \cdots) \\
&= P(E_1) + P(E_2 - E_1) + P(E_3 - E_2) + \cdots + P(E_n - E_{n-1}) + \cdots \\
&= P(E_1) + P(E_2) - P(E_1 \cap E_2) + \cdots + P(E_n) - P(E_n \cap E_{n-1}) + \cdots \\
&= P(E_1) + P(E_2) - P(E_1) + \cdots + P(E_n) - P(E_{n-1}) + \cdots \\
&= \lim_{n\to\infty} P(E_n)
\end{aligned}$$

La siguiente sección es una aplicación de la continuidad.

2.6.1. Probabilidades 0 y 1

Supóngase que un experimento consiste en seleccionar al azar un punto del intervalo $(0,1)$. Puesto que cada punto en $(0,1)$ tiene una representación decimal como $0,30434783\ldots$. El experimento es equivalente a tomar un decimal infinito en $(0,1)$ aleatoriamente. Si en tal experimento se quiere hallar la probabilidad de seleccionar el punto $\frac{1}{3} = 0,3333\ldots$, sea E_n el evento de que el decimal seleccionado tenga el 3 en los primeros n dgitos, así

$$\begin{array}{rcl}
E_1 & = & \{x | x = 0,3\alpha_1\alpha_2,\ldots,\ \alpha_i = 0,1,2,\ldots,9\} \\
E_2 & = & \{x | x = 0,33\alpha_1\alpha_2,\ldots,\ \alpha_i = 0,1,2,\ldots,9\} \\
E_3 & = & \{x | x = 0,333\alpha_1\alpha_2,\ldots,\ \alpha_i = 0,1,2,\ldots,9\} \\
\vdots & \vdots & \vdots \\
E_n & = & \left\{x | x = \underset{n-veces 3}{0,33\ldots3}\beta_1\beta_2\ldots,\ \beta_i = 0,1,2,\ldots,9\right\},
\end{array}$$

luego $E_1 \supset E_2 \supset \cdots \supset E_n \supset \ldots$. Además $P(E_1) = \frac{1}{10}$ porque se debe escoger el 3 de los números $0,1,2,\ldots,9$, $P(E_2) = \frac{1}{10^2}$ ya que se debe escoger el número 33 de los números $00,\ldots,10,\ldots,99$, y en general $P(E_n) = \frac{1}{10^n}$. Puesto que $\bigcap_{i=1}^{\infty} E_n = \{\frac{1}{3}\}$, la probabilidad de tener el punto $\frac{1}{3}$ es

$$\begin{array}{rcl}
P\left(\bigcap_{n=1}^{\infty} E_n\right) & = & P\left(\lim_{n\to\infty} E_n\right) \\
& = & \lim_{n\to\infty} P(E_n) \\
& = & \lim_{n\to\infty} \frac{1}{10^n} \\
& = & 0
\end{array}$$

De esta manera se tiene un evento, $\left\{\frac{1}{3}\right\} \neq \phi$ y la probabilidad es 0. El punto $\frac{1}{3}$ no es trascendental, se puede escoger cualquier $t \in (0,1)$ y $P(\{t\}) = 0$.
El evento $E_t = (0,1) - \{t\} \neq \Omega$, para cualquier $t \in (0,1)$

$$\begin{aligned} P(E_t) &= P((0,1)) - P((0,1) \cap \{t\}) \\ &= 1 - P(\{t\}) \\ &= 1 - 0 \\ &= 1 \end{aligned}$$

Es decir, existen eventos diferentes de Ω con probabilidad 1.

2.6.2. Selección aleatoria de puntos de intervalos

El hecho que la probabilidad de seleccionar cualquier punto de un intervalo (a,b) sea 0 implica que si $[\alpha,\beta] \subset (a,b)$ entonces el evento de que el punto caiga en $[\alpha,\beta]$, (α,β), $(\alpha,\beta]$, $[\alpha,\beta)$ son todas equivalentes. Ahora, considerando los intervalos $\left(a, \frac{a+b}{2}\right]$ y $\left(\frac{a+b}{2}, b\right)$ como $\frac{a+b}{2}$ es el punto medio del intervalo es razonable asumir que

$$p_1 = P\left(t \in \left(a, \frac{a+b}{2}\right]\right) = P\left(t \in \left(\frac{a+b}{2}, b\right)\right) = p_2.$$

Como $\left(a, \frac{a+b}{2}\right] \bigcap \left(\frac{a+b}{2}, b\right) = \phi$ y $(a,b) = \left(a, \frac{a+b}{2}\right] \bigcup \left(\frac{a+b}{2}, b\right)$ entonces,

$$\begin{aligned} P((a,b)) &= P\left(t \in \left(a, \frac{a+b}{2}\right] \bigcup \left(\frac{a+b}{2}, b\right)\right) \\ &= P\left(t \in \left(a, \frac{a+b}{2}\right]\right) + P\left(t \in \left(\frac{a+b}{2}, b\right)\right) \\ &= p_1 + p_2 = 2p_1, \end{aligned}$$

luego $2p_1 = 1$ y por tanto, $p_1 = \frac{1}{2} = p_2$, es decir, la probabilidad de que un punto seleccionado aleatoriamente de (a,b) caiga en el intervalo $\left(a, \frac{a+b}{2}\right]$ es $\frac{1}{2}$. Observe que la longitud de cada uno de estos intervalos es $\frac{1}{2}L(a,b)$.

Si (a,b) se divide en tres intervalos de igual longitud,$\left(a, \frac{2a+b}{3}\right]$, $\left(\frac{2a+b}{3}, \frac{a+2b}{3}\right]$, $\left(\frac{a+2b}{3}, b\right)$, es razonable asumir que $p_1 = p_2 = p_3$ donde $p_1 = P\left(t \in \left(a, \frac{2a+b}{3}\right]\right)$, $p_2 = P\left(t \in \left(\frac{2a+b}{3}, \frac{a+2b}{3}\right]\right)$ y $p_3 = P\left(\left(\frac{a+2b}{3}, b\right)\right)$ entonces $p_1 + p_2 + p_3 = 1$ así que $3p_1 = 1$ por lo tanto $p_1 = \frac{1}{3}$ de donde se concluye que $p_1 = p_2 = p_3 = \frac{1}{3}$. También se observa que la longitud de estos intervalos es $\frac{1}{3}L(a,b)$.

Definición 2.12. Un punto se dice que se selecciona aleatoriamente de (a, b) si cualquier dos subintervalos de (a, b) que tengan la misma longitud son igualmente probables de incluir el punto. La probabilidad asociada con el punto de que el subintervalo (α, β) de (a, b) lo contenga es $\frac{\beta-\alpha}{b-a}$, es decir

$$P\left(t \in (\alpha, \beta)\right) = \frac{\beta - \alpha}{b - a}.$$

2.7. Probabilidad condicional

Para ambientar el concepto, considérese la siguiente situación: se sabe que de los estudiantes de primer año del 2012 de un programa de matemáticas, 70 % ganaron geometría, 50 % ganaron cálculo diferencial y 40 % ganaron ambas asignaturas. Si se seleccionó al azar un estudiante de primer año de matemáticas y se encontró que ganó geometría, ¿cual es la probabilidad que haya ganado cálculo?

Sean E y F los eventos que el estudiante seleccionado haya ganado geometría y cálculo respectivamente. No se pide obtener $P(F)$ (porque se conoce), lo que se quiere calcular es P(F dado que E ocurrió).

Sea n el número de estudiantes de primer año de matemáticas del 2012; entonces el número de estudiantes que ganaron geometría es $0.70n$, y el número de estudiantes que ganaron geometría y cálculo es $0.40n$, luego

$$P\,(\text{ganar cálculo dado que ganó geometría}) = P(F|E) = \frac{0.40n}{0.70n} = \frac{4}{7}.$$

Definición 2.13. Sea $(\Omega, \boldsymbol{\mathcal{F}}, P(.))$ un espacio de probabilidad y sean E yF dos eventos en $\boldsymbol{\mathcal{F}}$. La probabilidad condicional del evento E dado el evento F, denotado por $P(E|F)$ se define por

$$P(E|F) = \frac{P(E \cap F)}{P(F)}, \text{ si } P(F) > 0. \tag{2.4}$$

Ejemplo 2.13. Del conjunto de familia con dos hijos se selecciona una familia al azar y se encontró que tiene una niña. ¿Cuál es la probabilidad que el otro hijo sea niña?

Solución Sean E y F los eventos tales que la familia seleccionada tiene una niña y dos niñas respectivamente.

Aquí $\Omega = \{nn, nN, Nn, NN\}$ donde n denota niño y N denota niña, entonces $E = \{nN, Nn, NN\}$ y $F = \{NN\}$. Luego

$$P(F|E) = \tfrac{P(E\cap F)}{P(E)} = \tfrac{P(F)}{P(E)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \tfrac{1}{3}.$$

Es evidente de la definición 2.13 que

$$P(E \cap F) = P(E|F)P(F), \text{ si } P(F) > 0$$

y

$$P(E \cap F) = P(F|E)P(E), \text{ si } P(E) > 0.$$

Estas fórmulas se conocen como la *regla de multiplicación*

La definición de probabilidad condicional es compatible con la aproximación frecuencial de probabilidad, pues si se observa un gran número, a saber N, de ocurrencias de un experimento ε, para el cual los eventos E y F están definidos, entonces $P(E|F)$ representa la proporción de ocurrencia en la cual F ocurrió, así como también E; es decir, $P(E|F) = \frac{N_{E\cap F}}{N_F}$, donde N_F denota el número de ocurrencias del evento F en las N ocurrencias del experimento, y $N_{E\cap F}$ denota el número de ocurrencias del evento $E \cap F$ en las N ocurrencias del experimento.

Ahora, $P(E \cap F) = \frac{N_{E\cap F}}{N}$ y $P(F) = \frac{N_F}{N}$, así que:

$$\frac{P(E \cap F)}{P(F)} = \frac{\frac{N_{E\cap F}}{N}}{\frac{N_F}{N}} = \frac{N_{E\cap F}}{N_F} = P(E|F)$$

Ejemplo 2.14. Se lanzan dos monedas. Si se supone que los cuatro resultados en el espacio muestral $\Omega = \{cc, cs, sc, ss\}$ son igualmente posibles, ¿cuál es la probabilidad:

a) De que ambas resulten caras, dado que una cara apareció en la primera moneda?

b) De que ambas resulten caras, dado al menos una cara?

Solución. Si $A = \{cc\}$ es el evento que denota que ambas monedas caen caras y si $B = \{cc, cs\}$ es el evento que denota obtener cara en la primera moneda entonces,

a) La probabilidad de obtener dos caras dado que en la primera salió cara es

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{1/4}{1/2} = \frac{1}{2}$$

b) Sea $C = \{cc, cs, sc\}$ el evento de obtener al menos una cara, entonces la probabilidad de obtener dos caras dada al menos una cara es

$$P(A|B) = \frac{P(A \cap C)}{P(C)} = \frac{P(A)}{P(C)} = \frac{1/4}{3/4} = \frac{1}{3}$$

Ejemplo 2.15. Una urna contiene 10 bolas blancas, 7 amarillas y 8 negras. Se selecciona una bola aleatoriamente de la urna y se nota que no es amarilla. ¿Cuál es la probabilidad que sea negra?

Solución. Sea A el evento que denota una bola negra y B el evento que denota que la bola seleccionada no es amarilla. Entonces B es el evento de obtener bola blanca o negra, por lo tanto, $P(A \cap B) = P(A)$.

Se desea encontrar $P(A|B) = P(A \cap B)/P(B) = P(A)/P(B)$.

Es decir, $P(A|B) = \frac{8/25}{18/25} = \frac{8}{18} = \frac{4}{9}$.

Cuando se habla de probabilidad condicional se está condicionando sobre algún evento dado F. Este evento es como un nuevo espacio muestral. La pregunta que surge inmediatamente ¿es $P(\cdot|F)$ una función de probabilidad?

En efecto, $P(\cdot|F)$ es una función de probabilidad puesto que,

1. $P(E|F) = P(E \cap F)/P(F) \geq 0$, para todo $E \in \boldsymbol{\mathcal{F}}$.

2. $P(\Omega|F) = P(\Omega \cap F)/P(F) = P(F)/P(F) = 1$.

3. Si $E_1, E_2, \cdots$, es una sucesión de eventos tales que $E_i \cap E_j = \phi$ para $i \neq j$ y si $\bigcup\limits_{i=1}^{\infty} E_i \in \boldsymbol{\mathcal{F}}$, entonces

$$P\left(\bigcup_{i=1}^{\infty} E_i|F\right) = \frac{P\left(\bigcup\limits_{i=1}^{\infty} E_i \cap F\right)}{P(F)} = \frac{\sum\limits_{i=1}^{\infty} P(E_i \cap F)}{P(F)} = \sum_{i=1}^{\infty} P(E_i|F).$$

Dado un espacio de probabilidad $(\Omega, \boldsymbol{\mathcal{F}}, P(\cdot))$ y $F \in \boldsymbol{\mathcal{F}}$ tal que $P(F) > 0$, se puede demostrar que las propiedades de $P(\cdot)$ también se satisfacen para $P(\cdot|F)$.

Teorema 2.10. Teorema de las probabilidades totales. Dado un espacio de probalidad $(\Omega, \mathcal{F}, P(\cdot))$, si $F_1, F_2, \cdots, F_n$ es una colección de eventos mutuamente excluyentes en $\mathcal{F}$ tales que

$$\bigcup_{i=1}^{n} F_i = \Omega \text{ y } P(F_i) > 0, \text{ para } i = 1, 2, \cdots, \text{ entonces para todo } E \in \mathcal{F},$$

$$P(E) = \sum_{i=1}^{n} P(E|F_i)P(F_i) \tag{2.5}$$

Demostración. Observe que $E = E \cap \Omega = E \cap (\bigcup_{i=1}^{n} F_i) = \bigcup_{i=1}^{n} (E \cap F_i)$.

Como para $i \neq j$, $E \cap F_i$ y $E \cap F_j$ son mutuamente excluyentes, entonces

$$P(E) = P(\bigcup_{i=1}^{n} (E \cap F_i)) = \sum_{i=1}^{n} P(E \cap F_i) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$$

Corolario 2.1. Dado un espacio de probabilidad $(\Omega, \mathcal{F}, P(.))$ y sea $F \in \mathcal{F}$ tal que $0 < P(F) < 1$. Entonces para cualquier $E \in \mathcal{F}$, se tiene que

$$P(E) = P(E|F)P(F) + P(E|F')P(F')$$

.

Teorema 2.11. Fórmula de Bayes. Dado un espacio de probabilidad $(\Omega, \mathcal{F}, P(.))$, si $F_1, F_2, \cdots, F_n$ es una colección de eventos mutuamente excluyentes en $\mathcal{F}$ que satisface $\bigcup_{i=1}^{n} F_i = \Omega$ y $P(F_i) > 0$, para $i = 1, 2, \cdots, n$, entonces para todo $E \in \mathcal{F}$

$$P(F_k|E) = \frac{P(E|F_k)P(F_k)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}, \; P(E) > 0. \tag{2.6}$$

Demostración. Puesto que $\bigcup_{i=1}^{n} F_i = \Omega$ y $F_i \cap F_j = \Phi$ para $i \neq j$, entonces para $E \in \mathcal{F}$, $P(E) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$, por 2.5.

Como $P(F_k|E) = \frac{P(E \cap F_k)}{P(E)} = \frac{P(E|F_k)P(F_k)}{P(E)}$ entonces, $P(F_k|E) = \frac{P(E|F_k)P(F_k)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}$.

Ejemplo 2.16. Cinco urnas enumeradas de 1 a 5, tienen 10 bolas cada una. La urna i tiene i bolas defectuosas y $10 - i$ no defectuosas, para $i = 1, 2, 3, 4, 5$. Si se selecciona al azar una urna y luego se toma una bola de la urna escogida, ¿cuál es la probabilidad que la bola venga de la urna 4 dado que la bola es defectuosa?.

Solución. Sea E el evento que la bola seleccionada sea defectuosa y sea F_i el evento de que la urna i sea seleccionada. Se desea encontrar $P(F_4|E)$.

Como $P(F_i) = 1/5$ y $P(E|F_i) = i/10$ para $i = 1, 2, 3, 4, 5$ entonces $P(E) = \sum\limits_{i=1}^{5} P(E|F_i)P(F_i) = \frac{1}{5}\sum\limits_{i=1}^{5} \frac{i}{10} = \frac{1}{50}(1+2+3+4+5) = \frac{3}{10}$.

Luego $P(F_4|E) = \frac{P(E|F_4)P(F_4)}{P(E)} = \frac{(\frac{4}{10})(\frac{1}{5})}{\frac{3}{10}} = \frac{4}{15}$.

Teorema 2.12. Regla de multiplicación. Sea $(\Omega, \boldsymbol{\mathcal{F}}, P(.))$ un espacio de probabilidad y sean $E_1, E_2, \cdots, E_n$ eventos en $\boldsymbol{\mathcal{F}}$ para los que $P(E_1 \cap E_2 \cap \cdots \cap E_{n-1}) > 0$ entonces,

$$P\Big(\bigcap_{i=1}^{n} E_i\Big) = P(E_1)P(E_2|E_1)P(E_3|(E_1 \cap E_2)) \cdots P(E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1}). \tag{2.7}$$

Demostración. Usar inducción matemática.

Ejemplo 2.17. Una urna contiene 12 bolas de las cuales 4 son blancas y 8 son rojas. Se proyecta el siguiente juego: se toma una bola aleatoriamente, se observa su color y esta se reemplaza con dos bolas del mismo color. ¿Cuál es la probabilidad de sacar una bola blanca en cada una de las tres primeras extracciones?

Solución. Suponga que E_i denota el evento de seleccionar una bola blanca en la *i-ésima* extracción, se debe encontrar $P(E_1 \cap E_2 \cap E_3)$.

Por la regla de multiplicación se tiene que

$P(E_1 \cap E_2 \cap E_3) = P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2)$.

De los datos del problema, $P(E_1) = 4/12 = 1/3$. Ahora, dado que la primera bola fue blanca, $P(E_2|E_1) = 6/14 = 3/7$; y además dado que la primera y la segunda bola fueron blancas, $P(E_3|(E_1 \cap E_2)) = 8/16 = 1/2$.

Luego, $P(E_1 \cap E_2 \cap E_3) = (1/3)(3/7)(1/2) = 1/14$.

2.8. Independencia entre eventos

Si $P(E|F)$ no depende del evento F, es decir, $P(E|F) = P(E)$, parece natural decir que E es independiente de F.

Definición 2.14. Dado un espacio de probabilidad $(\Omega, \boldsymbol{\mathcal{F}}, P(\cdot))$. Sean E y F dos eventos en $\boldsymbol{\mathcal{F}}$. Se dice que E yF son *independientes* si y solo si se satisface una cualquiera de las siguientes condiciones,

a. $P(E \cap F) = P(E)P(F)$.

b. $P(E|F) = P(E)$, si $P(F) > 0$.

c. $P(F|E) = P(F)$, si $P(E) > 0$.

Para argumentar que estas 3 condiciones son equivalentes, basta mostrar que a implica b, b implica c y c implica a.

Ejemplo 2.18. Sean E y F eventos independientes tales que $P(E \cup F) = 0.6$ y $P(E) = 0.4$, encuentre $P(F)$.

Solución. Como $P(E \cup F) = P(E) + P(F) - P(E \cap F)$ y $P(E \cap F) = P(E)P(F)$ por ser E y F independientes, entonces

$P(E \cup F) = P(E) + P(F) - P(E)P(F)$. Luego, $P(F) = \frac{P(E \cup F) - P(E)}{1 - P(E)}$.

Por tanto, $P(F) = \frac{0.6 - 0.4}{1 - 0.4} = \frac{0.2}{0.6} = \frac{1}{3}$.

Teorema 2.13. Si E y F son eventos independientes definidos en $(\Omega, \boldsymbol{\mathcal{F}}, P(.))$, entonces E y F', E' y F, y E' y F' son independientes.

Demostración. Se probará que $P(E' \cap F') = P(E')P(F')$, los otros dos casos son similares.

$$P(E' \cap F') = P([E \cup F]') = 1 - P(E \cup F) = 1 - P(E) - P(F) + P(E \cap F).$$

Por ser E y F independientes se tiene que:

$$\begin{aligned} P(E' \cap F') &= 1 - P(E) - P(F) + P(E)P(F) \\ &= 1 - P(E) - P(F)[1 - P(E)] \\ &= [1 - P(E)][1 - P(F)] = P(E')P(F'). \end{aligned}$$

Definición 2.15. Sea $(\Omega, \boldsymbol{\mathcal{F}}, P(\cdot))$ un espacio de probabilidad y sean $E_1, E_2, \cdots, E_n$ eventos en $\boldsymbol{\mathcal{F}}$. Estos eventos se dice que son independientes si y solo si

$$\begin{aligned} P(E_i \cap E_j) =& P(E_i)P(E_j), \text{ para } i \neq j \\ P(E_i \cap E_j \cap E_k) =& P(E_i)P(E_j)P(E_k), \text{ para } i \neq j \ j \neq k \ k \neq i \\ &\vdots \\ P\left(\bigcap_{i=1}^{n} E_i\right) =& \prod_{i=1}^{n} P(E_i) \end{aligned} \tag{2.8}$$

Observación.
La independencia a pares no implica la independencia de los eventos. En efecto, considere el experimento de lanzar dos dados. Si E_1 es el evento de obtener un número impar en el primer dado, E_2 es el evento de obtener un número impar en el segundo dado, y E_3 es el evento de obtener una suma impar. Entonces,

$P(E_1)P(E_2) = (1/2)(1/2) = P(E_1 \cap E_2)$.
$P(E_1)P(E_3) = (1/2)(1/2) = P(E_1 \cap E_3)$.
$P(E_2)P(E_3) = (1/2)(1/2) = P(E_2 \cap E_3)$.

Luego, E_1, E_2, E_3 son independiente a pares.

Sin embargo, $P(E_1 \cap E_2 \cap E_3) = 0 \neq P(E_1)P(E_2)P(E_3) = (1/1)(1/2)(1/2) = 1/8$.

Luego, E_1, E_2, E_3 no son independientes.

2.9. Ejercicios

1. Dado $P(A) = 0.5$ y $P(A \cup B) = 0.6$. Hallar $P(B)$, si:
   a) A y B son mutuamente excluyentes.
   b) A y B son independientes.
   c) $P(A|B) = 0.4$
2. Si A y B son independientes, $P(A) = 1/3$ y $P(B') = 1/4$. Hallar $P(A \cup B)$.
3. Suponga que A_1, A_2, A_3 son eventos mutuamente excluyentes y que $P(A_i) = 1/3$ y $P(B|A_i) = i/6$, para $i = 1, 2, 3$. Hallar $P(B)$.
4. Sean los eventos A_j, con $j = 1, 2, 3$. Si $A_1 \subset A_2 \subset A_3$ y $P(A_1) = 1/12$, $P(A_2) = 5/12$ y $P(A_3) = 7/12$. Calcular la probabilidad de los siguientes eventos:

a) $A_1' \cap A_2'$.

b) $A_1' \cap A_3'$.

c) $A_1' \cap A_2' \cap A_3'$.

5. Si $P(A) = 1/3$ y $P(B') = 1/4$, ¿pueden ser A y B disjuntos? Explique.

6. Si se lanzan dos dados una vez, ¿cuál es la probabilidad que el número total de puntos:

   a) Sea igual a 5?

   b) Sea divisible por 3?

7. Dos monedas, una con $P(cara) = \alpha$ y la otra con $P(cara) = \beta$ se lanzan simultanea e independientemente. Defina $p_0 = P(0\ caras)$, $p_1 = P(1\ cara)$ y $p_2 = P(2\ caras)$. ¿Se pueden escoger α y β tales que $p_0 = p_1 = p_2$? Probar la respuesta.

8. 20 bolas numeradas de 1 a 20 se colocan en una urna y se mezclan. Se toma una bola y después otra sin reemplazamiento. Si x_1 y x_2 son los número escritos en la primera y segunda bola, respectivamente, ¿cuál es la probabilidad de que:

   a) $x_1 + x_2 = 8$

   b) $x_1 + x_2 \leq 5$

9. Si las letras A, C, I, I, F, N y O se escriben en 7 fichas y se colocan en una urna, ¿Cuál es la probabilidad de formar la palabra OFICINA?

10. Si $P(A') = a$ y $P(B') = b$, probar que $P(A \cap B) \geq 1 - a - b$.

11. Si $P(A') = c$ y $P(B') = d$, probar que $P(A|B) \geq \frac{c+d-1}{d}$.

12. Si A y B son eventos tales que $P(A) = P(B) = 1$, probar que $P(A \cap B) = 1$.

13. Si A y B son eventos probar que

$$P(A \triangle B) = P(A) + P(B) - 2P(A \cap B),$$

   donde $A \triangle B = (A - B) \cup (B - A)$ es la diferencia simétrica de A y B.

14. Entre los números $1,\ 2, \cdots,\ 50$ se escoge un número al azar. ¿Cuál es la probabilidad que el número seleccionado sea divisible por 6 o por 8?

15. De 6 números positivos y 8 negativos, se eligen 4 al azar (sin sustitución) y se multiplican. ¿Cuál es la probabilidad que el producto sea un número positivo?

16. Una urna contiene 3 bolas rojas, 2 blancas y 1 amarilla. Una segunda urna contiene 1 bola roja, 2 blancas y 3 amarillas.

    a) Una bola se selecciona aleatoriamente de cada una. Describa el espacio muestral para este experimento. Encontrar la probabilidad que ambas sean del mismo color. ¿Es la probabilidad que ambas sean rojas, mayor que la probabilidad que ambas sean blancas?

    b) Las bolas de las dos urnas se combinan en una sola urna y se toma una muestra de tres. Encontrar la probabilidad que aparezcan los tres colores cuando:

        i) El muestreo es con reemplazamiento.

        ii) El muestreo es sin reemplazamiento.

17. Los 24 volúmenes de la Enciclopedia Británica se arreglan en un estante. ¿Cuál es la probabilidad que:

    a) Todos los 24 volúmenes aparezcan en orden ascendente?

    b) Todos los 24 volúmenes aparezcan en orden ascendente dado que los volúmenes 14 y 15 aparcen en orden ascendente y que los volúmenes 1 a 13 preceden al volúmen 14?

18. Si el espacio muestral de un experimento es $\Omega = \{w : 0 < w < \infty\}$ y $A = \{w : 7 < w < \infty\}$ es un evento con $P(A) = \int_A e^{-x} dx$, evaluar $P(A), P(A')$ y $P(A \cup A')$.

19. Suponga que la probabilidad que un par de gemelos sean niños es 0.30 y que la probabilidad que ambos sean niñas es 0.26. Dado que la probabilidad que un hijo sea niño es 0.52, ¿cuál es la probabilidad que:

    a) El segundo gemelo sea un niño, dado que el primero es un niño?

    b) El segundo gemelo sea una niña, dado que la primera es una niña?

20. Suponga que el 3 % de los hombres y el 0.3 % de las mujeres son daltónicas. Una persona se escoge al azar y resultó ser daltónica, ¿Cuál es la probabilidad que esa persona sea hombre?

21. La probabilidad que un niño sea asmático es 0.15 si el padre y la madre son fumadores, 0.13 si la madre es fumadora pero el padre no, 0.05 si el padre es fumador y la madre no y 0.04 si ninguno de los padres son fumadores. Además, la probabilidad que el padre sea fumador es de 0,5; si el padre es fumador, la probabilidad que la madre lo sea es 0,6, mientras que si el padre no es fumador, la probabilidad que la madre lo sea es 0,2.

    a.) ¿Cuál es la probabilidad que el padre sea fumador y que la madre no lo sea?

    b.) Hallar probabilidad no condicional que el niño tenga asma.

    c.) Si el niño tiene asma, ¿cuál es la probabilidad que el padre sea fumador?

    d.) ¿Cuál es la probabilidad que la madre sea fumadora si el niño tiene asma?

22. Tres máquinas I, II y III fabrican el 20 %, 30 % y 50 %, respectivamente, de la producción total de ciertos artículos. De estos 4 %, 3 % y 2 % respectivamente, son defectuosos. Un artículo se toma aleatoriamente, se prueba y se encuentra que es defectuoso. ¿Cuál es la probabilidad que el artículo fuese manufacturado por la máquina III?.

23. Una urna I contiene dos bolas blancas y dos negras; una urna II contiene tres bolas blancas y dos negras. Una bola se transfiere de I a II, luego una bola se selecciona de la urna II y resultó ser blanca. ¿Cuál es la probabilidad que la bola transferida haya sido blanca?

24. Demostrar el Teorema 2.5.

25. Demostrar la desigualdad de Boole.

26. Demuestre la Regla de la Multiplicación (Teorema 2.12).

Capítulo 3

Variables aleatorias y distribuciones de probabilidad

3.1. Variables aleatorias

Es frecuente el caso que en el desarrollo de un experimento se esté interesado principalmente en alguna función de los resultados en vez de ellos mismos. Por ejemplo, en la lanzada de dos dados se está interesado a menudo en la suma de las caras superiores y no realmente en los valores individuales de los dados. También , en las tiradas de una moneda, se puede estar interesado en el total de caras que ocurren y no en la sucesión de caras y sellos que resulten. Estas cantidades de interés o estas funciones de valor real definidas en el espacio muestral, se conocen como *variables aleatorias*.

Definición 3.1. Variable aleatoria. Sea $(\Omega,\ \boldsymbol{\mathcal{F}},\ P(.))$ un espacio de probabilidad dado, una variable aleatoria es una función X definida de Ω a los reales R, tal que el conjunto A, definido por

$$A = \{w \in \Omega : X(w) \leq x\} \in \boldsymbol{\mathcal{F}},\ x \in R. \tag{3.1}$$

La ecuación 3.1 establece que la función X es *medible* y se puede escribir como $\{X \leq x\} \in \boldsymbol{\mathcal{F}}$, o $X^{-1}((-\infty, x]) \in \boldsymbol{\mathcal{F}}$, para todo $x \in R$.

En general, para denotar variables aleatorias se usarán las últimas letras mayúsculas X, Y, U, V, W y Z del alfabeto inglés, y para denotar un valor de la variable aleatoria, se usarán letras minúsculas x, y, u, v, w y z. Cuando sobre un mismo espacio de probabilidad se definen varias variables aleatorias se usa una sola de estas letras acompañada de un subíndice, como $X_1, X_2, \cdots$, etc.

Ejemplo 3.1. Considérese el experimento de lanzar una moneda legal. Sea X la variable aleatoria que denota el número de caras. Entonces $\Omega = \{c, s\}$ y $X(w) = 0$, si $w = s$, y $X(w) = 1$ si $w = c$. Así, la variable aleatoria X asocia un número real con cada resultado del experimento. Puede demostrarse que esta función es realmente aleatoria. El lector puede comprobarlo fácilmente.

Observación.
Dado que los valores de una variable aleatoria están determinados por el resultado de un experimento, se pueden asignar probabilidades a los valores posibles de la variable aleatoria. En el ejemplo anterior,

$$P(X = 0) = P(X^{-1}\{0\}) = P(\{s\}) = 1/2$$

y

$$P(X = 1) = P(X^{-1}\{1\}) = P(\{c\}) = 1/2.$$

Ejemplo 3.2. Suponga que el experimento consiste en lanzar dos monedas legales. Si X denota el número de caras obtenidas, entonces X es la variable aleatoria que toma uno de los valores 0, 1, 2 con probabilidades respectivas:

$P(X = 0) = P(\{ss\}) = \frac{1}{4}$,
$P(X = 1) = P(\{cs, sc\}) = \frac{1}{2}$,
$P(X = 2) = P(\{cc\}) = \frac{1}{4}$.

Puesto que X toma solo uno de los valores 0, 1, 2, se debe tener que

$1 = P(\bigcup_{i=0}^{2} \{X = i\}) = \sum_{i=0}^{2} P(X = i)$, lo cual es cierto.

Ejemplo 3.3. Considere el experimento de lanzar dos dados legales, defínanse las dos variables aleatorias: X como la suma de las caras superiores, y Y como la diferencia absoluta de las caras superiores.

La variable aleatoria X puede tomar los valores 2, 3, $\cdots$, 12 y puesto que el espacio muestral consta de 36 resultados posibles, se puede dar una “distribución de probabilidades”de esta, como se muestra a continuación.

X	2	3	4	5	6	7	8	9	10	11	12
$P(X=i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

donde, por ejemplo $P(X=4) = P(\{(i,j) : i+j=4\}) = P(\{(1,3),(2,2),(2,1)\})$.

Esta tabla se puede expresar algunas veces por una fórmula y también por un gráfico. Una fórmula que expresa la distribución de probabilidades anterior es,

$$P(X=i) = \begin{cases} \frac{i-1}{36}, & i = 2,3,\cdots,7 \\ \frac{13-i}{36}, & i = 8,9,\cdots,12 \end{cases}$$

La variable aleatoria Y puede tomar uno de los valores 0, 1, $\cdots$, 5 y su distribución de probabilidades se da a continuación:

Y	0	1	2	3	4	5
$P(Y=j)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

donde, por ejemplo $P(Y=5) = P(\{(i,j) : |i-j| = 5\}) = P(\{(1,6),(6,1)\})$.

En el ejemplo anterior se habló de una “distribución de probabilidades ”y ésta se presentó como una tabla, una fórmula o un gráfico y ésto es precisamente lo que se entiende por la distribución de probabilidades de una variable aleatoria.

Definición 3.2. Distribución de probabilidad. Una tabla, una fórmula o un gráfico que involucra los valores de una variable aleatoria y sus respectivas probabilidades, se llama la distribución de la variable aleatoria.

Ejemplo 3.4. Se seleccionan tres bolas aleatoriamente de una urna que contiene 3 bolas blancas, 4 rojas y 5 negras. Si se gana $1.oo por cada bola blanca seleccionada y se pierde \$1.oo por cada bola roja seleccionada, y X denota la variable aleatoria *ganancia total*, hallar su distribución de probabilidades.

Solución. La variable aleatoria puede tomar uno de los valores $0, \pm 1, \pm 2$, o ± 3, y para obtener las probabilidades respectivas debe notarse que los eventos:

a) $\{X = 0\}$ ocurre si todas las bolas seleccionadas son negras o hay una de cada color, es decir,

$$P\{X = 0\} = \frac{\binom{5}{3} + \binom{3}{1}\binom{4}{1}\binom{5}{1}}{\binom{12}{3}} = \frac{10 + 60}{220} = \frac{70}{220}$$

b) $\{X = 1\}$ ocurre si una bola blanca y dos negras o dos bolas blancas y una roja son seleccionadas, es decir,

$$P\{X = 1\} = \frac{\binom{3}{1}\binom{5}{2} + \binom{3}{2}\binom{4}{1}}{\binom{12}{3}} = \frac{42}{220}$$

c) $\{X = 2\}$ ocurre si dos bolas blancas y una negra son selecionadas, es decir,

$$P\{X = 2\} = \frac{\binom{3}{2}\binom{5}{1}}{\binom{12}{3}} = \frac{15}{220}$$

d) $\{X = 3\}$ ocurre si tres bolas blancas son seleccionadas, es decir,

$$P\{X = 3\} = \frac{\binom{3}{3}}{\binom{12}{3}} = \frac{1}{220}$$

e) $\{X = -1\}$ ocurre si una bola blanca y dos rojas o una roja y dos negras son seleccionadas, es decir,

$$P\{X=-1\} = \frac{\binom{3}{1}\binom{4}{2} + \binom{4}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{58}{220}$$

f) $\{X=-2\}$ ocurre si dos bolas rojas y una bola negra es seleccionada, es decir,

$$P\{X=-2\} = \frac{\binom{4}{2}\binom{5}{1}}{\binom{12}{3}} = \frac{30}{220}$$

g) $\{X=-3\}$ ocurre si 3 bolas rojas son seleccionadas, es decir,

$$P\{X=-3\} = \frac{\binom{4}{3}}{\binom{12}{3}} = \frac{4}{220}$$

Luego, la distribución de probabilidades de la variable aleatoria "ganancia total"es

X	-3	-2	-1	0	1	2	3	Total
$P(X=i)$	$\frac{4}{220}$	$\frac{30}{220}$	$\frac{58}{220}$	$\frac{70}{220}$	$\frac{42}{220}$	$\frac{15}{220}$	$\frac{1}{220}$	1

3.2. Función de distribución

Definición 3.3. Función de distribución. La función de distribución acumulativa de una variable aleatoria X (o simplemente la función de distribución de X), denotada por F o por $F_X(.)$, se define para todo real b como

$$F_X(b) = P(X \leq b) = P(\{w : X(w) \leq b\}). \tag{3.2}$$

En palabras, $F(b)$ denota la probabilidad que la variable aleatoria X tome un valor menor o igual que b.

3.2.1. Propiedades de la función de distribución F

Sea F la función de distribución de una variable aleatoria X entonces,

a) F es una función no decreciente, es decir, si $a < b$, entonces $F_X(a) \leq F_X(b)$.

En efecto, si $a < b$, entonces el evento $\{X \leq a\}$ está contenido en el evento $\{X \leq b\}$ y por tanto $P(X \leq a) \leq P(X \leq b)$, o sea que $F_X(a) \leq F_X(b)$.

b) F es normalizada; es decir,

$$\lim_{x \to -\infty} F_X(x) = 0 \text{ y } \lim_{x \to +\infty} F_X(x) = 1.$$

En efecto, si $x \to -\infty$, el evento $\{X \leq x\}$ converge al evento $\{X < -\infty\}$, es decir,

$$\lim_{x \to -\infty} \{X \leq x\} = \{X < -\infty\}$$

y por tanto,

$$\lim_{x \to -\infty} F_X(x) = P(\lim_{x \to -\infty} \{X \leq x\}) = P(X < -\infty) = P(\phi) = 0.$$

Además , si $x \to +\infty$, el evento $\{X \leq x\}$ converge al evento $\{X < \infty\}$, es decir,

$$\lim_{x \to +\infty} \{X \leq x\} = \{X < +\infty\}$$

y por tanto,

$$\lim_{x \to +\infty} F_X(x) = P(\lim_{x \to +\infty} \{X \leq x\}) = P(X < +\infty) = P(\Omega) = 1.$$

c) $F_X(.)$ es continua por la derecha, es decir,

$$F(x) = F(x^+), \text{ donde } F(x^+) = \lim_{y \to x} F(y) \text{ y } y > x.$$

La definición de función de distribución y sus propiedades se recogen de una forma más general en el siguiente teorema.

Teorema 3.1. Una función F con recorrido en $[0, 1]$ es una función de distribución para alguna probabilidad P si y solo si satisface las condiciones

a.) F es no decreciente.

b.) F es normalizada.

c.) F es continua por la derecha.

Ejemplo 3.5. En el ejemplo 3.1 la función de distribución es

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}, & 0 \leq x < 1 \\ 1, & x \geq 1, \end{cases}$$

o también

$$F_X(x) = \frac{1}{2} I_{(0,1)}(x) + I_{[1,\infty)}(x).$$

Ejemplo 3.6. En el ejemplo 3.3 la función de distribución de la variable aleatoria diferencia absoluta es

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{6}{36}, & 0 \leq x < 1 \\ \frac{16}{36}, & 1 \leq x < 2 \\ \frac{24}{36}, & 2 \leq x < 3 \\ \frac{30}{36}, & 3 \leq x < 4 \\ \frac{34}{36}, & 4 \leq x < 5 \\ 1, & x \geq 5 \end{cases}$$

Para calcular la probabilidad de que X sea estrictamente menor que x, se aplica la propiedad de continuidad para obtener

$$\begin{aligned} P(X < b) &= P(\lim_{n\to\infty} \{X \leq b - 1/n\}) = \lim_{n\to\infty} P(X \leq b - 1/n) \\ &= \lim_{n\to\infty} F_X(b - 1/n) = F(b^-). \end{aligned}$$

Además, como $\{X = a\} \cup \{X < a\} = \{X \leq a\}$ se tiene también que

$$P(X = a) = F_X(a) - P(X < a) = F(a) - F(a^-), \text{ o } F(a) = P(X = a) + F(a^-).$$

Esto conduce al lema siguiente,

Lema. Sea F la función de distribución de X entonces,

1. $P(X = x) = F(x) - F(x^-)$.
2. $P(x < X \leq y) = F(y) - F(x)$.

3. $P(X > x) = 1 - F(x)$.

4. Si X es una variable aleatoria continua(ver sección 3.3.2) entonces,

$$\begin{aligned} P(x < X < y) &= P(x \le X < y) = P(x < X \le y) = P(x \le X \le y) \\ &= F(y) - F(x). \end{aligned}$$

Ejemplo 3.7. Supóngase que la función de distribución de una variable aleatoria X está dada por

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x/4, & 0 \le x1 \\ (1/2) + (x-1)/4, & 1 \le x < 2 \\ 11/12, & 2 \le x < 3 \\ 1, & x \ge 3 \end{cases}$$

Hallar $P(X = 2),\ P(1/2 < X \le 3/2)$ y$P(X < 2)$.

Solución. Se encontrará primero $P(X < 2)$.

$$\begin{aligned} P(X < 2) &= \lim_{n\to\infty} F(2 - 1/n) = \lim_{n\to\infty} \left[\frac{1}{2} + \frac{(2 - \frac{1}{n} - 1)}{4}\right] \\ &= \frac{1}{2} + \lim_{n\to\infty} \left(\frac{1}{4} - \frac{1}{4n}\right) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}. \end{aligned}$$

Luego,

$$\begin{aligned} P(X = 2) = F(2) - P(X < 2) &= \frac{11}{12} - \frac{3}{4} = \frac{1}{6},\ \text{y} \\ P(1/2 < X \le 3/2) = F(3/2) - F(1/2) &= \left(\frac{1}{2} + \frac{\frac{3}{2} - 1}{4}\right) - \frac{1/2}{4} = \frac{1}{2}. \end{aligned}$$

3.3. Clases de variables aleatorias

En este texto se distinguen dos clases de variables aleatorias: *discretas* y *continuas*.

3.3.1. Variables aleatorias discretas

Definición 3.4. Una variable aleatoria X se dice que es discreta si puede tomar un número finito o infinito enumerable de valores. Para una variable aleatoria discreta

X se define la *función de masa de probabilidad* por

$$p(x) = p_X(x) = P(X = x). \tag{3.3}$$

La función de masa de probabilidad p es positiva para a lo más un número contable de valores de x. Es decir, si X es discreta con valores posibles $x_1,\ x_2,\ \cdots$, entonces,

$$p(x_i) \geq 0$$

y además se cumple también que

$$\sum_{i=1}^{\infty} p(x_i) = 1.$$

Teorema 3.2. Si X es una variable aleatoria discreta, $F_X(\cdot)$ se obtiene de $p_X(\cdot)$ y viceversa.

Demostración. Sean $x_1,\ x_2,\ \cdots$, los valores posibles de la variable aleatoria X y suponga que $p(\cdot) = p_X(\cdot)$ es dado, entonces

$$F_X(x) = P(X \leq x) = \sum_{\{j:x_j \leq x\}} p_X(x_j).$$

Recíprocamente, si $F_X(\cdot)$ es dado, entonces $p_X(x) = P(X = x)$.

En efecto,

$$\begin{aligned} p_X(x) =& P(X \leq x) - P(X < x) = F_X(x) - \lim_{n\to\infty} P\left(X \leq x - \frac{1}{n}\right) \\ =& F_X(x) - \lim_{n\to\infty} F\left(x - \frac{1}{n}\right). \end{aligned}$$

Para ilustrar el teorema anterior se considera el siguiente ejemplo.

Ejemplo 3.8. Considere el experimento de lanzar un dado legal y observe la cara superior. Si X denota el número de puntos en la cara superior, obtener la función de masa de probabilidad y la función de distribución de X.

Solución. Puesto que X toma los valores posibles 1, 2, 3, 4, 5, 6, con probabilidades 1/6, entonces

$$p_X(x) = \frac{1}{6} I_{\{1,\ 2,\ \cdots,\ 6\}}(x)$$

Otra manera de expresar esta función es

$$p(x) = \begin{cases} \frac{1}{6} & \text{para } x = 1,2,3,4,5,6 \\ 0 & \text{en otro caso.} \end{cases}$$

La función de distribución de esta variable aleatoria es

$$F(x) = P(X \leq x) = \begin{cases} 0, & x < 1 \\ \frac{1}{6}, & 1 \leq x < 2 \\ \frac{2}{6}, & 2 \leq x < 3 \\ \frac{3}{6}, & 3 \leq x < 4 \\ \frac{4}{6}, & 4 \leq x < 5 \\ \frac{5}{6}, & 5 \leq x < 6 \\ 1, & x \geq 6 \end{cases}$$

En términos de la función indicadora

$$F_X(x) = \sum_{i=1}^{5} \frac{i}{6} I_{[i,i+1)}(x) + I_{[6,\infty)}(x)$$

Si $x = 2.8$ entonces,

$$F_X(2.8) = \sum_{\{j: x_j \leq 2.8\}} p_X(x_j) = p_X(1) + p_X(2) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

Si X es una variable discreta cuyos valores posibles son $x_1,\ x_2,\ \cdots$, donde $x_1 < x_2 < \cdots$, entonces la función de distribución F es una *función escalonada.* Es decir, el valor de F es constante en los intervalos $[x_{i-1}, x_i)$ y luego toma un paso o un salto de tamaño $p(x_i)$ en x_i. En el ejemplo anterior, la función de distribución se representa en la figura 3.1.

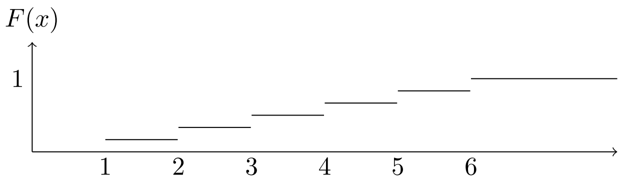


Figura 3.1.Función de distribución del ejemplo 3.8

3.3.2. Variables aleatorias continuas

Definición 3.5. Una variable aleatoria X se dice que es continua, si existe una función $f_X(.)$, no negativa tal que

$$F(x) = \int_{-\infty}^{x} f_X(t)dt, \tag{3.4}$$

para cada número real x. La función f se llama la *función de densidad de probabilidad de la variable aleatoria* X y es tal que

$$\int_{-\infty}^{\infty} f_X(x)dx = 1.$$

Todas la probabilidades acerca de X se obtienen por medio de f. En efecto

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx,$$

Así,

$$P(a < X \leq b) = \int_{a}^{b} f(x)dx.$$

Si $a = b$ en la ecuación anterior entonces

$$P(X = a) = \int_{a}^{a} f(x)dx = 0.$$

El resultado anterior nos lleva a que

$$P(X < a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx = F_X(a).$$

Teorema 3.3. Si X es una variable aleatoria continua, entonces $F_X(.)$ se obtiene de $f_X(.)$ y viceversa.

Demostración. Si $f_X(.)$ es dado entonces,

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(u)du.$$

Si $F_X(x)$ es dado entonces,

$$\frac{d}{dx}F_X(x) = \frac{d}{dx}\int_{-\infty}^{x} f_X(u) = f_X(x).$$

Ejemplo 3.9. El tiempo, en horas X, en que un computador funciona antes de hacerle mantenimiento es una variable aleatoria continua con función de densidad $f_X(x) = \lambda e^{-x/120}$, $x \geq 0$. ¿Cuál es la probabilidad que funcione al menos 80 horas?

Solución. La función dada tiene una constante que se puede conocer usando el hecho que

$$1 = F_X(\infty) = \int_{-\infty}^{\infty} f(x)dx = \lambda \int_{0}^{\infty} e^{-x/120}dx$$

$$1 = 120\lambda \int_{0}^{\infty} \frac{1}{120}e^{-\frac{1}{120}x}dx = 120\lambda \int_{0}^{\infty} e^{-u}du = 120\lambda \lim_{b\to\infty}\int_{0}^{b} e^{-u}du = 120\lambda,$$

luego $\lambda = \frac{1}{120}$ y de esta manera,

$$P(X < 80) = \frac{1}{120}\int_{0}^{80} e^{-x/120}dx = -e^{-x/120}|_{0}^{80} = 1 - e^{-80/120} = 0.487,$$

Por tanto,

$$P(X \geq 80) = 0.513.$$

3.4. Esperanzas y momentos de variables aleatorias

Uno de los conceptos más importantes en la teoría de probabilidad es el de valor esperado de una variable aleatoria. De igual forma es interesante conocer los momentos y la función generadora de estas.

Definición 3.6. Sea X una variable aleatoria, la media o la esperanza o el valor promedio de X denotado por $\mu,\ \mu_X$ o $E(X)$ se define por

$$E(X) = \mu_X = \begin{cases} \sum\limits_{i=1}^{\infty} x_i p(x_i), & \text{si } X \text{ es una variable aleatoria discreta que toma} \\ & \text{valores } x_1,\ x_2,\ \cdots, \text{ con función de probabilidad } p. \\ \int\limits_{-\infty}^{\infty} x f(x) dx, & \text{si } X \text{ es una variable aleatoria continua, con} \\ & \text{función de densidad } f. \end{cases}$$

El valor esperado existirá siempre que la suma o la integral que la definen converja.

Una definición más general que no incluye el conocimiento de la función de masa de probabilidad o de la función de densidad de probabilidad es la siguiente

Definición 3.7. Sea X cualquier variable aleatoria cuya función de distribución es $F_X(.)$, se define el valor esperado de X por

$$E(X) = \int_0^{\infty} [1 - F(x)]dx - \int_{-\infty}^{0} F(x)dx \tag{3.5}$$

Ejemplo 3.10. Considérese el experimento de lanzar dos dados legales y sean X y Y como se definieron en el Ejemplo 3.3. Obtener $E(X)$ y $E(Y)$.

Solución.

$$\begin{aligned} E(X) = \sum_{x=2}^{12} xp(x) =& 2\left(\frac{1}{36}\right) + 3\left(\frac{2}{36}\right) + \cdots + 11\left(\frac{2}{36}\right) + 12\left(\frac{1}{34}\right) \\ =& \frac{252}{36} = 7 \end{aligned}$$

Esto no significa que en una lanzada particular de los dos dados, necesariamente se espere obtener la suma 7, sino que $E(X) = 7$ es el número que se espera obtener en un gran número de desarrollos del experimento.

$$E(Y) = \sum_{y=0}^{5} yp(y) = 0\left(\frac{6}{36}\right) + 1\left(\frac{10}{36}\right) + \cdots + 5\left(\frac{2}{36}\right) = \frac{70}{36} = 1.94$$

La acotación anterior vale también para este caso.

Ejemplo 3.11. De 6 candidatos que buscan tres posiciones en un centro de consultoría, 2 son graduados en ciencias sociales y 4 no. Si los tres candidatos son seleccionados aleatoriamente, hallar el número esperado de graduados en ciencias sociales.

Solución. Sea X el número de personas seleccionadas en este experimento y que son graduados en ciencias sociales. Entonces X toma los valores 0, 1 y 2, con probabilidades

$$p(0) = P(X=0) = \frac{\binom{4}{3}}{\binom{6}{3}} = \frac{4}{20}$$
$$p(1) = P(X=1) = \frac{\binom{2}{1}\binom{4}{2}}{\binom{6}{3}} = \frac{12}{20}$$
$$p(2) = P(X=2) = \frac{\binom{2}{2}\binom{4}{1}}{\binom{6}{3}} = \frac{4}{20}$$

Luego, la distribución de probabilidad de X es

$X = x$	0	1	2
$p(x)$	$\frac{4}{20}$	$\frac{12}{20}$	$\frac{4}{20}$

Por tanto,

$$E(X) = 0\left(\frac{4}{20}\right) + 1\left(\frac{12}{20}\right) + 2\left(\frac{4}{20}\right) = 1.$$

Ejemplo 3.12. Sea X una variable aleatoria continua con función de densidad dada por $f(x) = |x| I_{(-1,1)}(x)$. Hallar $E(X)$.

Solución

$$E(X) = \int_{-\infty}^{\infty} x f(x)dx = \int_{-1}^{1} x\ |x|dx = \int_{-1}^{0} x(-x)dx + \int_{0}^{1} xxdx.$$

Luego,

$$E(X) = \int_{-1}^{0} -x^2 dx + \int_{0}^{1} x^2 dx = \left.\frac{-x^3}{3}\right|_{-1}^{0} + \left.\frac{x^3}{3}\right|_{0}^{1} = \frac{-1}{3} + \frac{1}{3} = 0.$$

Ejemplo 3.13. Sea X una variable aleatoria continua con función de distribución dada por $F(X) = (1 - pe^{-\lambda x}),\ p > 0, \lambda > 0, x > 0$. Hallar $E(X)$.

Solución Usando la ecuación 3.5, se tiene que $E(X)$ es

$$E(X) = \int_0^{\infty} pe^{-\lambda x} dx \text{ ya que para } x \leq 0,\ F(x) = 0.$$

Luego,

$$E(X) = \lim_{b \to \infty} p \int_0^{b} e^{-\lambda x} dx = \frac{p}{\lambda} \lim_{b \to \infty} \left(-e^{-\lambda x}\Big|_0^b \right),$$

Entonces,

$$E(X) = \frac{p}{\lambda} \lim_{b \to \infty} (-e^{-\lambda b} + 1) = \frac{p}{\lambda}.$$

Otro de los conceptos de mucha importancia es el de *varianza de una variable aleatoria*, la cual se define como,

Definición 3.8. Sea X una variable aleatoria con media $E(X) = \mu_x$, la varianza de X, denotada por σ^2, σ_X^2 o $V(X)$, se define por

$$V(X) = \sigma^2 = \begin{cases} \sum_{i=1}^{\infty} (x_i - \mu_X)^2 p(x_i), & \text{si } X \text{ es discreta y toma valores posibles } x_1, x_2, \cdots. \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x), & \text{si } X \text{ es continua con función de densidad } f. \\ \int_0^{\infty} 2x[1 - F(x) + F(-x)]dx - \mu_X^2, & \text{para cualquier variable aleatoria arbitraria X.} \end{cases} \tag{3.6}$$

Observación.
$E(X)$ representa el centro de gravedad o centroide de la unidad de masa que determina la función de densidad de X y $V(X)$ representa el momento de inercia.

Definición 3.9. La desviación estándar de la variable aleatoria X, denotada por σ, se define por

$$\sigma = \sqrt{V(X)}.$$

Ejemplo 3.14. Obtener la varianza y la desviación estándar de las variables aleatorias de los ejemplos 3.11 al 3.13.

Solución. En el ejemplo 3.11 se encontró que $\mu_X = 1$ entonces,

$$V(X) = \sum_{x=0}^{2} (x-1)^2 p(x) \text{ luego,}$$

$$V(X) = (-1)^2 \frac{4}{20} + 0 + (1)^2 \frac{4}{20} = \frac{8}{20} = \frac{2}{5}; \text{ y } \sigma = \sqrt{\frac{2}{5}}.$$

En el ejemplo 3.12, para la variable aletoria X continua, se encontró que $E(X) = 0$, por tanto

$$\begin{aligned} V(X) &= \int_{-1}^{1} x^2 \ |x| dx = \int_{-1}^{0} -x^3 dx + \int_{0}^{1} x^3 dx \\ &= \left. \frac{-x^4}{4} \right|_{-1}^{0} + \left. \frac{x^4}{4} \right|_{0}^{1} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}. \end{aligned}$$

Como $\sigma^2 = 1/2$, entonces $\sigma = \sqrt{1/2}$.

En el ejemplo 3.13, también de variable aleatoria continua, se encontró que $E(X) = p/\lambda$, por tanto

$$V(X) = \int_{0}^{\infty} 2x(pe^{-\lambda x}) dx - \left(\frac{p}{\lambda}\right)^2, \text{ ya que } F(-x) = 0$$

Luego,

$$\begin{aligned} V(X) &= 2p \int_{0}^{\infty} x e^{-\lambda x} dx - \left(\frac{p}{\lambda}\right)^2 = \frac{2p}{\lambda^2} - \frac{p^2}{\lambda^2} \\ &= \frac{p(2-p)}{\lambda^2}. \end{aligned}$$

Es decir,

$$\sigma^2 = \frac{p(2-p)}{\lambda^2}, \text{ y } \sigma = \frac{\sqrt{p(2-p)}}{\lambda}$$

En muchas ocasiones interesa encontrar el valor de una función de la variable aleatoria, esto condujo a la siguiente definición la cual se conoce como la *Ley del estadístico inconsciente.*

Definición 3.10. Sea X una variable aleatoria y $g(\cdot)$ una función real. La esperanza o el valor esperado de la función $g(\cdot)$ de la variable aleatoria X, denotada por $E[g(X)]$, se define por

$$E[g(X)] = \begin{cases} \sum\limits_{i} g(x)p(x), & \text{si } X \text{ es una variable aleatoria} \\ & \text{discreta con función de} \\ & \text{probabilidad } p \\ \int\limits_{-\infty}^{\infty} g(x)f(x)dx, & \text{si } X \text{ es una variable aleatoria} \\ & \text{continua con función de} \\ & \text{densidad } f \end{cases} \tag{3.7}$$

Es de anotar que $E(X)$, $V(X)$ y $E[g(X)]$ existen si y solo si las sumatorias o las integrales que las definen existen.

Ejemplo 3.15. Sea X el número de caras obtenidas cuando se lanzan dos monedas. Calcular $E(X^2)$.

Solución. Como la distribución de probabilidad de X es

X	0	1	2
$p(X)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Sea $Y = g(X) = X^2$, entonces

$$E(Y) = E(X^2) = \sum_{x=0}^{2} x^2 p(x) = 0 + 1\left(\frac{1}{2}\right) + 2^2\left(\frac{1}{4}\right) = \frac{3}{2}$$

Ejemplo 3.16. Sea X una variable aleatoria con función de densidad $f(x) = 1$, para $0 < x < 1$. Hallar $E(e^X)$.

Solución. Sea $Y = e^X$, entonces

$$E(e^X) = E(Y) = \int\limits_{-\infty}^{\infty} y\ dx = \int\limits_{0}^{1} e^x\ dx = e^x\big|_0^1 = e - 1$$

Observación.
De la definición 3.10 es claro que si $g(x) = x$, entonces $E[g(X)] = E(X)$, y si

$g(x) = (x-\mu)^2$, entonces $E[(X-\mu)^2] = V(X)$.

Teorema 3.4. Sean X una variable aleatoria, a y b constantes g y h funciones reales. Entonces

a) $E(a) = a$.

b) $E[ag(X)] = aE[g(X)]$.

c) $E[ag(X) + bh(X)] = aE[g(X)] + bE[h(X)]$.

d) Si $g(x) \leq h(x)$ para todo x, entonces $E[g(X)] \leq E[h(X)]$.

Demostración. Se supondrá que X es una variable aleatoria discreta y se demostrán *b)* y *d)*, las otras dos quedan como ejercicio lo mismo que todas en el caso en que X sea continua.

- b) Sea $ag(X) = L(X)$, entonces $E[ag(X)] = E[L(X)]$.
  O sea que
  $$\begin{aligned} E[ag(X)] =& \sum_x L(x)p(x) = \sum_x ag(x)p(x) \\ =& a\sum_x g(x)p(x). \end{aligned}$$
  Luego,
  $$E[ag(X)] = aE[g(X)].$$
- d) Si $g(x) \leq h(x)$ para todo x, entonces $h(x) - g(x) \geq 0$. Sea $L(x) = h(x) - g(x)$, entonces $E[L(X)] \geq 0$. Pero,
  $$E[L(X)] = E[h(X) - g(X)] = E[h(X)] - E[g(X)] \geq 0.$$
  Luego
  $$E[h(X)] \geq E[g(X)].$$

Teorema 3.5. Sea X una variable aleatoria, entonces

$$V(X) = E(X^2) - [E(X)]^2. \tag{3.8}$$

Demostración. Por la observación previa al teorema anterior se tiene que

$$V(X) = E\{[X - E(X)]^2\} = E\{X^2 - 2XE(X) + [E(X)]^2\}.$$

Es decir,

$$V(X) = E(X^2) - 2E(X)E(X) + [E(X)]^2 = E(X^2) - [E(X)]^2.$$

A continuación se dan tres desigualdades las cuales juegan un rol importante en probabilidad.

Teorema 3.6.(Desigualdad de Markov) Sea X una variable aleatoria y g una función real no negativa, entonces

$$P[g(X) \geq k] \leq \frac{E[g(X)]}{k}, \ k > 0. \tag{3.9}$$

Demostración. Sea $k > 0$ y supóngase que X toma cualquier valor en $(-\infty, \infty)$. Sea $A = \{x : g(x) < k\}$ y $B = \{x : g(x) \geq k\}$, es claro que

$$(-\infty, \infty) = \{x : g(x) < k\} \cup \{x : g(x) \geq k\} = A \cup B.$$

Luego,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx = \int_A g(x)f(x)dx + \int_B g(x)f(x)dx,$$

Ahora, puesto que $g(x) \geq 0$ y $f(x) > 0$, entonces $\int_A g(x)f(x)dx \geq 0$.

De esta manera,

$$E[g(X)] \geq \int_B g(x)f(x)dx \geq \int_B kf(x)dx = kP[g(X) \geq k]$$

Es decir,

$$\frac{E[g(X)]}{k} \geq P[g(X) \geq k], \ k > 0.$$

Corolario. Desigualdad de Chebyshev. Si X es una variable aleatoria con varianza finita σ^2 y media μ, entonces para cualquier $k > 0$

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}. \tag{3.10}$$

Demostración. Sea $g(x) = (x - \mu)^2 \geq 0$, entonces

$$P(|X - \mu| \geq k) = P[(X - \mu)^2 \geq k^2] \leq \frac{E[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2}.$$

Observación.
La desigualdad anterior es equivalente a $P(|X-\mu| \leq k) \geq 1 - \frac{\sigma^2}{k^2}$, y se obtiene usando la propiedad del complemento.

Ejemplo 3.17. Si X es una variable aleatoria con media 7 y varianza 4. ¿Que puede decirse acerca de la $P(2 < X < 12)$?

Solución. Como no se conoce la distribución exacta de X, se puede usar la desigualdad de Chebyshev, pues

$$\begin{aligned} P(2 < X < 12) =& P(2-7 < X-7 < 12-7) = P(-5 < X-7 < 5) \\ =& P(|X-7| < 5) \geq 1 - \frac{4}{25} = 0.840 \end{aligned}$$

Teorema 3.7. Desigualdad de Jensen. Sea X una variable aleatoria con media μ. Si g es una función convexa, entonces

$$E(g(X)) \geq g(E(X)). \tag{3.11}$$

Demostración. Sea g una función convexa(ver apéndice B).
Si se hace una expansión en serie de g alrededor de $\mu = E(X)$ se tiene que,

$$g(x) = g(\mu) + g'(\mu)(x-\mu) + \frac{g''(s)(x-s)^2}{2}$$

donde s es algún valor entre x y μ.

Dado que g es convexa, $g''(s) \geq 0$ y por tanto, $g(x) \geq g(\mu) + g'(\mu)(x-\mu)$.

Luego,

$$E[g(X)] \geq E[g(\mu)] + g'(\mu)E(X-\mu) = g(\mu) = g[E(X)].$$

Ejemplo 3.18. Una variable aleatoria X tiene media 5, ¿qué se puede decir acerca de $E(e^{-X})$?

Solución. Sea $g(x) = e^{-x}$, luego g es convexa porque $g''(x) = e^{-x} \geq 0$, para todo x. De manera que

$$E(e^{-X}) = E[g(X)] \geq g[E(X)] = g(5) = e^{-5}.$$

Observación.
El valor esperado y la varianza son casos especiales de los momentos, los cuales

tienen un papel importante en la teoría estadística.

Definición 3.11. Momento. Sea X una variable aleatoria, el *k-ésimo* momento o el momento de orden k de X, denotado por μ'_k se define por

$$\mu'_k = E(X^k) \tag{3.12}$$

siempre que ese valor esperado exista.

Estos momentos se llaman por algunos autores, momentos alrededor del origen y puede notarse que $\mu'_1 = E(X) = \mu$, la media de X.

Definición 3.12. Momento central. Sea X una variable aleatoria, el *k-ésimo* momento central o *k-ésimo* momento alrededor de la medida μ, denotado μ_k se define por

$$\mu_k = E[(X-\mu)^k] \tag{3.13}$$

Puede verse que $\mu_1 = 0$ y $\mu_2 = V(X)$; además los momentos centrales se expresan también en términos de los momentos $E(X),\ E(X^2),\ \cdots$. Por ejemplo,

$$V(X) = E[(X-\mu)^2] = E(X^2) - [E(X)]^2 = \mu'_2 - (\mu'_1)^2.$$

Como en el caso descriptivo, aquí se puede hablar de otras medidas como la mediana y la moda, las cuales como el primero de estos, son casos particulares de nociones más generales, llamadas percentiles.

Definición 3.13. Percentil. Sea $0 < p < 1$, el percentil del $(100p)\,\%$ o el cuantil de orden p de una variale aleatoria X, denotado por ξ_p, se define como el número más pequeño que satisface

$$F(\xi) \geq p.$$

A ξ_p se le llama también el *p-ésimo* percentil de X.

Definición 3.14. La mediana de una variable aleatoria X, denotada por $\xi_{.5}$, es el percentil del 50 %. Alternativamente se define como aquel número ξ que satisface

$$P(X \leq \xi) \geq \frac{1}{2},\ y\ P(X \geq \xi) \leq \frac{1}{2}$$

Una manera alterna de obtener los percentiles es mediante la siguiente función.

Definición 3.15. Función cuantil. Sea X una variable aleatoria con función de distribución F. La inversa de F se llama función cuantil y se define por

$$F^{-1}(\xi) = \text{ínf}\{x|F(x) > \xi\},\ \text{para } \xi \in [0,1]$$

Si F es extrictamente creciente y continua entonces $F^{-1}(\xi)$ es el único real x tal que $F(x) = \xi$. De esta manera, $\xi_{.25} = F^{-1}(\frac{1}{4})$ se llama el primer cuartil, $\xi_{.5} = F^{-1}(\frac{1}{2})$ es la mediana o segundo cuartil y $\xi_{.75} = F^{-1}(\frac{3}{4})$ es el tercer cuartil.

Observaciónes.

1. Si X es una variable aleatoria continua, el percentil del $100p\,\%$ de X está dado por el número más pequeño ξ que satisface $F(\xi) = p$.

2. Si X es una variable aleatoria continua la mediana de X, $\xi_{.5}$, satisface

$$\int_{-\infty}^{\xi_{.5}} f(x)dx = \int_{\xi_{.5}}^{\infty} f(x)dx = \frac{1}{2}$$

Comentarios

1. $E(X)$ y $\xi_{.5}$, son considerados como medidas que localizan el centro de la distribución. Otra medida considerada también como de localización central es la *moda* la cual se define como aquel valor de X en el cual $f(x)$ alcanza su valor máximo.

2. La varianza se usa como una medida de separación o de dispersión de una distribución. Otras medidas de dispersión están definidas en términos de los percentiles. Una de estas es el *rango intercuartílico*, el cuál se define por $\xi_{.75} - \xi_{,25}$. En general, $\xi_p - \xi_{1-p}$ es una medida de dispersión para $p,\ 1/2 < p < 1$.

3. μ_3 se emplea como una medida de asimetría o sesgo. Se demuestra que distribuciones simétricas como la de la figura 3.2, tienen $\mu_3 = 0$.

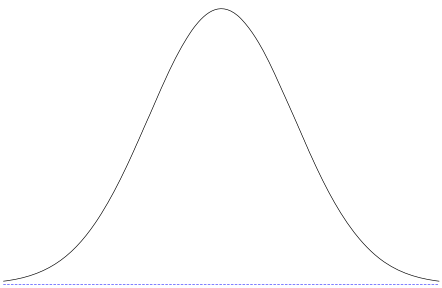


Figura 3.2. Densidad simétrica

Una curva como la de la función de densidad de la figura 3.3, se dice que es sesgada a la derecha y puede demostrarse que $\mu_3 > 0$

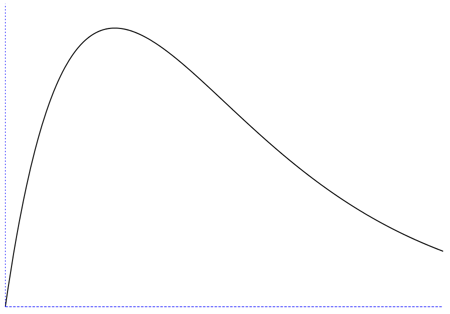


Figura 3.3. Densidad sesgada a la derecha

y una curva como la función de densidad de la figura 3.4, se dice que es sesgada a la izquierda y se demuestra que $\mu_3 < 0$.

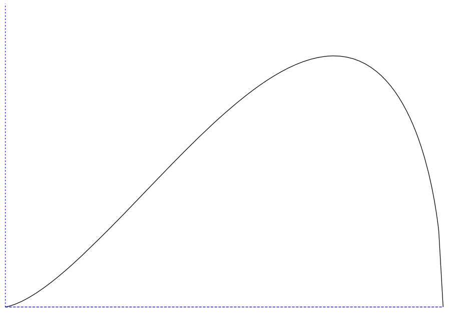


Figura 3.4. Densidad sesgada a la izquierda

4. El conocimiento de μ_3 no da siempre una pista sobre la forma de una distribución, pues a veces $\mu_3 = 0$ y sin embargo la curva de la densidad puede no

ser simétrica. Sin embargo, existe una magnitud llamada coficiente de sesgo $\alpha_3 = \mu_3/\sigma^3$ el cual es: cero si la distribución es simétrica y toma un gran valor positivo (o negativo) para una distribución que tiene una gran cola a la derecha (o a la izquierda).

La cantidad s =(media - mediana)/(desviación estándar) se usa como medida alternativa de sesgo.

5. μ_4 se usa a menudo como una medida de "exceso" o "curtosis". El coeficiente de exceso o curtosis se define por $\alpha_4 = \dfrac{\mu_4}{\sigma^4} - 3$ y da el grado de achatamiento de una densidad de probabilidad acerca de su centro. Si $\alpha_4 > 0$ la distribución se llama *leptocúrtica* y será más picuda que la densidad de la distribución normal estándar, y si $\alpha_4 < 0$, la distribución será más achatada que la densidad de la distribución normal estándar y se llama *mesocúrtica*.

Para finalizar este capítulo se hablará de una clase especial de esperanza matemática mediante la cual se pueden encontrar todos los momentos. Esta es la noción de la *función generatriz de momentos*. En la sección 6.3 se muestra el poder de esta función.

Definición 3.16. Sea X una variable aleatoria con función de densidad $f(x)$ o de masa $p(x)$. Se define la *función generatriz de momentos* de la variable aleatoria X por el valor esperado de e^{tX}, si este existe para algún valor de t, tal que $-h < t < h$ y $h > 0$.

La función generatriz de momentos de una variable aleatoria X se denota por $M_X(t)$ o $M(t)$, luego

$$M(t) = E(e^{tX}) = \begin{cases} \sum_x e^{tx} p(x), & \text{si X es discreta} \\ \int_{-\infty}^{\infty} e^{tx} f(x)dx, & \text{si X es continua} \end{cases} \tag{3.14}$$

Si la función generatriz de momentos existe, entonces $M(t)$ es continuamente diferenciable en alguna vecindad de $t = 0$.

Derivando $M(t)$ k veces con respecto a t, se tiene que

$$M^{(k)}(t) = \frac{d^k}{dt^k} M(t) = \begin{cases} \sum_x x^k e^{tx} p(x), & \text{caso discreto} \\ \int_{-\infty}^{\infty} x^k e^{tx} f(x)dx, & \text{caso continuo} \end{cases} \tag{3.15}$$

Luego $M^{(k)}(0) = E(X^k) = \mu'_k$.

Ejemplo 3.19. Una variable aleatoria X tiene función de masa de probabilidad $p(x) = p^x(1-p)^{1-x}I_{\{0,1\}}(x)$, con $0 < p < 1$. Hallar la función generatriz de X y con base en esta su media y su varianza.

Solución.

$$M_X(t) = \sum_x e^{tx}p(x) = \sum_0^1 e^{tx}p^x(1-p)^{1-x} = 1 - p + pe^t,$$

luego

$$M'_X(t) = M''_X(t) = pe^t.$$

En consecuencia,

$$E(X) = M'_X(0) = p$$

y

$$V(X) = E(X^2) - [E(X)^2] = M''_X(0) - M'_X(0)^2 = p - p^2 = p(1-p).$$

3.5. Ejercicios

1. Si la variable aleatoria X toma valores $1,\ 2,\ \cdots$, y si $P(X = i) = \frac{1}{2^i}$, para $i = 1,\ 2,\ \cdots$, calcular

   a) $P(X$ sea par$)$.

   b) $P(X > 5)$.

   c) $P(X$ sea divisible por $3)$.

2. Demostrar que la función $p(x)$ definida en el ejercicio 1 es una función de masa de probabilidad. Encontrar la moda de esta distribución.

3. Si X es una variable aleatoria continua. Demuestre que

$$f(x) = \begin{cases} -\ln x, \text{ para } 0 < x < 1 \\ 0, \text{ en otro caso} \end{cases}$$

   es una función de densidad de probabilidad. Luego calcule $P(X > \frac{2}{3})$.

4. La variable aleatoria continua X tiene función de densidad $f(x) = 3x^2$, para $-1 \leq x \leq 0$. Si b es un número tal que satisface $-1 < b < 0$. Calcular la $P[X > b | X < \frac{b}{2}]$.
   Sugerencia: use la noción de probabilidad condicional de eventos.

5. Si la variable aleatoria X toma valores $0, 1, \ldots,$ con función de masa de probabilidad $p_X(x) = \frac{C}{3^x}$, a) determinar la constante C. b) Calcular $P(X > 10)$.

6. En cada caso hallar la función de distribución de la variable aleatoria X, en donde X es discreta o continua, según el caso.

   a) $p(x) = 1/3$, $x = -1,\ 0,\ 1$.
   b) $f(x) = 3(1-x)^2$, $0 < x < 1$.
   c) $p(x) = x/15$, $x = 1,\ 2,\ 3,\ 4,\ 5$.
   d) $f(x) = 1/3$, $0 < x < 1$ o $2 < x < 4$.
   e) $f(x) = \frac{1}{\pi(1+x^2)}$, $-\infty < x < \infty$.

7. En el ejercicio anterior encontrar la media, la mediana, la moda y la varianza de la variable X.

8. Si la función de distribución de la variable aleatoria X es

$$F(x) = \begin{cases} 0,\ x < -1 \\ \frac{x+2}{4},\ -1 \leq x < 1 \\ 1,\ x \geq 1. \end{cases}$$

   Grafique F y Calcule: a) $P(-1/4 < X \leq 3/4)$, b) $P(X = 0)$ y c) $P(3/2 < X < 4)$.

9. Sea X una variable aleatoria continua con función de densidad $f(x)$ y de distribución $F(x)$. Para x_0 fijo, sea la función

$$g(x) = \begin{cases} \frac{f(x)}{1-F(x_0)},\ x \geq x_0 \\ 0,\ x < x_0. \end{cases}$$

   Probar que g es una función de densidad.

10. Sea X una variable aleatoria y F_X su función de distribución. Si F_X es estrictamente creciente, entonces F_X^{-1} esta definida por

$$F_X^{-1}(y) = x \text{ si y solo si } F_X(x) = y.$$

   Hallar F_X^{-1} si,

a) $F_X(x) = 1 - e^{-x},\ x \geq 0.$

b) $F_X(x) = \begin{cases} e^x/2, & x < 0 \\ 1/2, & 0 \leq x < 1 \\ 1 - \frac{e^{1-x}}{2}, & x \geq 1 \end{cases}$

11. En el ejercicio 4, encontrar $E(X^{\frac{1}{2}})$ y $E(X^2)$, si existen.

12. Si X es una variable aleatoria con función de densidad dada por $f(x) = |1-x|$, $0 \leq x \leq 2$, encontrar su media y su varianza.

13. Hallar la $P(\mu_X - 2\sigma < X < \mu_X + 2\sigma)$, donde X tiene función de probabilidad dada por

    a) $p(x) = (1/2)^x$, $x = 1,\ 2,\ 3,\ \cdots$

    b) $f(x) = 6x(1-x)$, $0 < x < 1$.

14. Si la varianza de la variable aleatoria X existe, demostrar que $E(X^2) \geq [E(X)]^2$.

15. Si X es una variable aleatoria que toma valores $1, 2, 3, ...,$

    a) Demuestre que $p(n) = P(X = n) = \frac{1}{n(n+1)}, n \geq 1$ es una función de masa de probabilidad.

    b) Halle $E(X)$ y $V(X)$

16. Si X es una variable aleatoria continua con función de densidad dada por

    a) $f(x) = 1/2$, $-1 < x < 1$.

    b) $f(x) = (1-x)/2$, $-1 < x < 1$.

    c) $f(x) = (x+1)/2$, $-1 < x < 1$.

    I. Demostrar que la distribución de X es simétrica en el caso *a)*; sesgada a la derecha en el caso *b)* y sesgada a la izquierda en el caso *c)*.

    II. Encontrar el coeficiente de curtosis en cada caso y hacer los comentarios pertinentes.

17. Sea X una variable aleatoria con función de densidad

$$f(x) = \begin{cases} \frac{|x|}{4}, & -2 < x < 2 \\ 0, & en \text{ otro caso.} \end{cases}$$

    Hallar el percentil del 75 %.

18. Si X es una variable aleatoria tal que $E(X) = 3$ y $E(X^2) = 13$, determinar el acotamiento inferior para $P(-2 < X < 8)$.

19. Si X es una variable aleatoria con $E(X) = \mu$ y que satisface $P(X \leq 0) = 0$, demostrar que $P(X > 2\mu) \leq 1/2$.

20. Sea X una variable aleatoria con $V(X) > 0$. ¿Cuál afirmación es verdadera? Justifique.

$$E(e^{-X}) < e^{-E(X)} \quad \text{o} \quad E(e^{-X}) > e^{-E(X)}.$$

21. Si X es una variable aleatoria cuya función de masa de probabilidad está dada por $p(x) = (1-\delta)\delta^x$, $x = 0,\ 1,\ 2,\ \cdots$ y $0 < \delta < 1$. Hallar la función generatriz de momentos y luego calcular $E(X)$ y $V(X)$.

22. Si la variable aleatoria X tiene media μ, varainza σ^2 y función generatriz de momentos $M_X(t), -h < t < h, h > 0$, probar que para $Z = \frac{X-\mu}{\sigma}$,

$$M_Z(t) = e^{-\frac{t\mu}{\sigma}} M_X\left(\frac{t}{\sigma}\right), -h\sigma < t < h\sigma.$$

23. Si $M(t)$ es la función generadora de momentos de la variable aleatoria X y si $\Psi(t) = \ln M(t)$. Probar que $\Psi'(0) = \mu_X$ y $\Psi''(0) = V(X)$.

24. Usando $M(t)$ de la variable aleatoria X del ejercicio 17 obtenga $\Psi(t)$ y encuentre $E(X)$ y $V(X)$.

25. Sea X una variable aleatoria con función generatriz de momentos $M(t)$, para $-h < t < h$ y $h > 0$. Probar que:

$$P(X \geq a) \leq e^{-at} M(t), \text{ para } 0 < t < h$$

y

$$P(X \leq a) \leq e^{-at} M(t), \text{ para } -h < t < 0.$$

Capítulo 4

Distribuciones especiales

El propósito de este capítulo está en presentar algunas variables aleatorias que se caracterizan por su función de masa o su función de densidad de probabilidad y que se conocen con nombres especiales. Entre las discretas se estudiarán, las distribuciones uniforme discreta, Bernoulli, binomial, Poisson, geométrica, binomial negativa y la hipergeométrica, y entre las continuas, la uniforme, normal, exponencial, gamma, chi-cuadrado, Weibull, beta, lognormal, t de Student y F de Snedecor.

4.1. Distribuciones discretas

4.1.1. Distribución uniforme discreta

Definición 4.1. Una variable aleatoria X se dice que tiene una distribución uniforme discreta si su función de masa de probabilidad está dada por

$$p_X(x) = \begin{cases} \frac{1}{N}, & x = 1,\ \ 2,\ \ldots\ N, \quad N \in \mathbb{Z}^+ \\ 0, & \text{en otro caso} \end{cases} \tag{4.1}$$

En términos de la función indicadora,

$$p_X(x) = \frac{1}{N} I_{\{1,\ 2,\ \ldots,\ N\}}(x).$$

Teorema 4.1. Si la variable aleatoria X tiene una distrtibución uniforme con función de masa de probabilidad dada por la definición 4.1, entonces

$$E(X) = \frac{N+1}{2} \text{ y } V(X) = \frac{N^2-1}{12}.$$

Demostración. Si X tiene una distribución uniforme discreta, entonces

$$E(X) = \sum_{x=1}^{N} x\left(\frac{1}{N}\right) = \frac{1}{N}\sum_{x=1}^{N} x = \frac{1}{N}\left[\frac{N(N+1)}{2}\right] = \frac{N+1}{2}, \text{ y}$$

$$\begin{aligned} V(X) = E(X^2) - [E(X)]^2 &= \sum_{x=1}^{N} x^2 \cdot \frac{1}{N} - \left(\frac{N+1}{2}\right)^2 \\ &= \frac{1}{N}\sum_{x=1}^{N} x^2 - \left(\frac{N+1}{2}\right)^2 \\ &= \frac{1}{N}\cdot\left[\frac{N(N+1)(2N+1)}{6}\right] - \left(\frac{N+1}{2}\right)^2 \\ &= \frac{N+1}{12}[2(2N+1) - 3(N+1)] = \frac{N^2-1}{12}. \end{aligned}$$

Este tipo de función de masa de probabilidad se emplea para describir el modelo de probabilidad cuando cada uno de los N puntos del espacio muestral, tienen la misma probabilidad $1/N$.

4.1.2. Distribuciones de Bernoulli y binomial

Un ensayo de Bernoulli es un experimento que tiene dos resultados posibles, etiquetados como “éxito”con probabilidad p y “fracaso”con probabilidad $1-p$, y $0 < p < 1$. Suponga el desarrollo de un ensayo de Bernoulli y sea X la variable aleatoria definida como el número de éxitos, entonces $X = 0$ o $X = 1$ y además, $P(X = 0) = 1 - p$ y $p(X = 1) = p$.

Esto se puede expresar por la fórmula,

$$p_X(x) = \begin{cases} p^x(1-p)^{1-x}, & \text{para } x = 0, 1, \\ 0, & \text{en otro caso} \end{cases} \tag{4.2}$$

o equivalentemente por

$$p_X(x) = p^x(1-p)^{1-x} I_{\{0,1\}}(x).$$

Definición 4.2. Una variable aleatoria X se dice que tiene una distribución de Bernoulli, con parámetro p, si su función de masa de probabilidad está dada por la ecuación 4.2. Se abrevia $X \sim B(p)$.

Teorema 4.2. Si X tiene una distribución de Bernoulli con parámetro p, entonces

$$E(X) = p,\ V(X) = p(1-p) \text{ y } M_X(t) = pe^t + 1 - p.$$

Demostración. La prueba es un caso particular del teorema 4.3, sin embargo se probó en el ejemplo 3.19.

Supóngase ahora que se desarrollan n ensayos independientes de Bernoulli con parámetro p. Si X denota el número de "éxitos" que ocurren en las n pruebas, entonces X toma valores posibles $0, 1, ..., n$ y su función de masa de probabilidad es

$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0,\ 1, \ldots,\ n, \\ 0, & \text{en otro caso.} \end{cases} \tag{4.3}$$

o

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} I_{\{0,1,\cdots,n\}}(x).$$

Donde $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ (ver apéndice A).

Definición 4.3. Una variable aleatoria X cuya función de masa de probabilidad está dada por la ecuación 4.3 se dice que tiene una distribución binomial con parámetros n y p. Para expresar esto se abrevia $X \sim B(n, p)$.

Teorema 4.3. Si X tiene una distribución binomial con parámetros n y p, entonces

$$E(X) = np,\ V(X) = np(1-p) \ \text{ y } \ M_X(t) = (pe^t - p + 1)^n.$$

Demostración. Primero se encuentra la expresión para la función generatriz y luego por derivación se obtienen la media y la varianza.

Puesto que $X \sim B(n, p)$ entonces,

$$\begin{aligned} M_X(t) = E[e^{tX}] &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x}. \end{aligned}$$

Pero $(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}$ y por tanto, $M_X(t) = (pe^t - p + 1)^n$.

Ahora, $M'_X(t) = n(pe^t - p + 1)^{n-1}pe^t$, y $M'_X(0) = E(X) = np$,

$M''_X(t) = np\{e^t(n-1)(pe^t - p + 1)^{n-2}pe^t + e^t(pe^t - p + 1)^{n-1}\}$, y

$M''_X(0) = E(X^2) = np[(n-1)p + 1]$. Por tanto,

$$V(X) = np[(n-1)p + 1] - n^2p^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p).$$

Observación.
La distribución binomial con parámetros n y p se reduce a la de Bernoulli cuando $n = 1$.

Ejemplo 4.1. Se sabe que para una moneda particular es dos veces más probable salir cara que sello. Si se lanza esta cuatro veces, hallar la función de masa de probabilidad del número de caras obtenidas.

Solución. Considere el experimento de lanzar la moneda mencionada una vez, entonces la probabilidad de obtener cara ("éxito") es $p = 2/3$. Si se lanza 4 veces y X denota el número de caras obtenidas, entonces, $X \sim B(4, 2/3)$. Por tanto,

$$p(x) = p_X(x) = \begin{cases} \binom{4}{x}\left(\frac{2}{3}\right)^x\left(\frac{1}{3}\right)^{4-x}, & x = 0,\ 1,\ 2,\ 3,\ 4. \\ 0, & \text{en otro caso.} \end{cases}$$

En particular, la probabilidad de obtener dos caras es $p(2) = \frac{24}{81}$.

El siguiente teorema da un criterio acerca del comportamiento de $p_X(\cdot)$.

Teorema 4.4. Si $X \sim B(n, p)$ entonces para $(n+1)p$ entero y $x = 1,\ 2,\ \ldots,\ n$

$$p(x-1) \begin{cases} < p(x), & \text{para } x < (n+1)p \\ > p(x), & \text{para } x > (n+1)p \\ = p(x), & \text{para } x = (n+1)p \end{cases}$$

Demostración. Se analizará el cociente $\dfrac{p_X(x)}{p_X(x-1)}$, donde $p_X(i)$ está dada por (4.3).

$$\frac{p_X(x)}{p_X(x-1)} = \frac{\binom{n}{x}p^x(1-p)^{n-x}}{\binom{n}{x-1}p^{x-1}(1-p)^{n-(x-1)}} = \frac{(n-x+1)p}{x(1-p)}. \tag{4.4}$$

De esta manera,

$$\frac{p_X(x)}{p_X(x-1)} > 1 \text{ si } \frac{(n+1)p - px}{x(1-p)} > 1,$$

es decir,

$$p_X(x) > p_X(x-1) \text{ si } (n+1)p > x.$$

También,

$$\frac{p_X(x)}{p_X(x-1)} = 1 \text{ si } \frac{(n+1)p - px}{x(1-p)} = 1,$$

lo cuál equivale a

$$p_X(x) = p_X(x-1) \text{ si } (n+1)p = x.$$

Y además

$$\frac{p_X(x)}{p_X(x-1)} < 1 \text{ si } \frac{(n+1)p - px}{x(1-p)} < 1,$$

por lo tanto

$$p_X(x) < p_X(x-1) \text{ si } (n+1)p < x.$$

De la ecuación 4.4 se observa, además que

$$p_X(x) = \frac{(n-x+1)p}{x(1-p)} p_X(x-1)$$

la cual es una fórmula de recursión para calcular $p(x)$ bajo la condición inicial de que $p(0) = (1-p)^n$.

Ejemplo 4.2. Si $X \sim B(5, \frac{1}{3})$ observe que $(n+1)p = 2$ y $p(0) = \frac{32}{243}$. Usando la fórmula de recursión se encuentra que,

$p(1) = \frac{80}{243} > p(0)$, $p(2) = \frac{80}{243} = p(1)$, $p(3) = \frac{40}{243} < p(2)$ $p(4) = \frac{10}{243} < p(3)$ y $p(5) = \frac{1}{243} < p(4)$.

Es de anotar que la distribución binomial es de las más antiguas y aunque fue publicada por James Bernoulli en 1713 en su tratado *Ars Conjectandi* [14] para $p = \dfrac{r}{r+s}$, donde r y s son enteros positivos, había sido considerada antes en el siglo XVII por Blaise Pascal para el caso en que $p = \frac{1}{2}$.

4.1.3. Distribución de Poisson

Un resultado importante en la búsqueda de una aproximación para $p(x)$ cuando $X \sim B(n,p)$ la obtuvo el matemático francés Simeon-Denis Poisson en 1837 [8], utilizando el siguiente argumento: Si el número de pruebas es grande ($n \to \infty$), la probabilidad de éxito es pequeña ($p \to 0$) y el número promedio de éxitos tiende a una cantidad fija de valor moderado ($np = \lambda$, para alguna constante λ), entonces

$$p_X(x) = \frac{n!p^x(1-p)^{n-x}}{x!(n-x)!} = \frac{n!(\lambda/n)^x(1-\lambda/n)^{n-x}}{x!(n-x)!}$$

$$= \frac{\lambda^x}{x!}\left[\frac{n!}{(n-x)!n^x}\right]\left[\frac{(1-\lambda/n)^n}{(1-\lambda/n)^x}\right]$$

$$= \frac{\lambda^x}{x!}\left[\frac{n(n-1)\cdots[n-(x-1)]}{n^x}\right]\left[\frac{(1-\lambda/n)^n}{(1-\lambda/n)^x}\right]$$

y por consiguiente

$$\lim_{n\to\infty} p_X(x) = \frac{\lambda^x}{x!}e^{-\lambda}.$$

La importancia de esta aproximación fue desconocida hasta 1889 cuando el matemático ruso-germano Bortkiewiez demostró su significado para la teoría de probabilidad y sus aplicaciones. Entre otras cosas argumentó que, puesto que

$$\sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} = 1$$

entonces la aproximación de Poisson es por sí misma una función de masa de probabilidad.

Definición 4.4. Una variable aleatoria X, que toma uno de los valores posibles $0, 1, 2, \ldots$ se dice que tiene una distribución de Poisson con parámetro λ, $\lambda > 0$, si su función de masa de probabilidad está dada por

$$p_X(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & x = 0,\ 1,\ 2,\ \ldots, \\ 0, & \text{en otro caso} \end{cases} \tag{4.5}$$

La aproximación de Poisson a la distribución binomial es buena cuando $p < 0.1$ y $np \leq 10$.

Ejemplo 4.3. Se sabe que los tornillos producidos por una compañía son defectuosos con probabilidad 0.01 independientemente uno de otro. Si esta vende los tornillos en paquetes de 100, ¿cuál es la probabilidad que en un paquete haya más de un tornillo defectuoso?

Solución. Sea X el número de tornillos defectuosos en un paquete de 100, entonces $X \sim B(100, 0.01)$. Por los valores de $n = 100$ y $p = 0.01$ se puede aproximar la probabilidad requerida usando la distribución de Poisson teniendo en cuenta que $\lambda = np = 1$, o sea que

$$\begin{aligned} P(X > 1) &= 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) \\ &= 1 - e^{-1} - \frac{(e^{-1})1}{1!} = 1 - \frac{2}{e} = 0.264241118 \end{aligned}$$

La probabilidad exacta es 0.264238021.

Notación.
Si X tiene una distribución de Poisson con parámetro λ, se abrevia $X \sim P(\lambda)$.

Teorema 4.5. Si X tiene una distribución de Poisson con parámetro λ, entonces

$$E(X) = \lambda, \ V(X) = \lambda \ \text{ y } \ M_X(t) = e^{\lambda(e^t-1)}$$

Demostración. Si $X \sim P(\lambda)$, entonces

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t-1)}.$$

De manera que,

$$M'_X(t) = \lambda e^t e^{\lambda(e^t-1)} \ \text{ y } \ M''_X(t) = \lambda^2 e^{2t} e^{\lambda(e^t-1)} + \lambda e^t e^{\lambda(e^t-1)}.$$

Así que,

$$E(X) = M'(0) = \lambda \ \text{ y } \ E(X^2) = M''_X(0) = \lambda^2 + \lambda.$$

Luego,

$$V(X) = E(X^2) - [E(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Observación.
Si $X \sim P(\lambda)$ la probabilidad del entero i puede obtenerse de la del entero $i-1$ pues,

$$P(X = x) = \frac{\lambda}{x} P(X = x-1), x = 1, 2, \cdots \text{ y donde } p(0) = e^{-\lambda}.$$

El siguiente teorema da el comportamiento de la función de masa de probabilidad de la distribución de Poisson, y su demostración es semejante al teorema 4.4. La observación anterior es de ayuda para hacer la prueba.

Teorema 4.6. Si $X \sim P(\lambda)$, entonces $p(x-1) < p(x)$, para $x < \lambda$, $p(x-1) > p(x)$, para $x > \lambda$ y $p(x-1) = p(x)$, si λ es entero y $x = \lambda$.

Ejemplo 4.4. Si $X \sim P(3)$ entonces $p(0) = e^{-3} < p(1) = 3e^{-3} < p(2) = \frac{9}{2}e^{-3} = p(3)$ y $p(3) > p(4) = \frac{27}{8}e^{-3} > p(5) = \frac{81}{40}e^{-3} > \cdots$

Algunos ejemplos que se pueden modelar como una distribución de Poisson son los siguientes:

- El número de errores de impresión en una página o grupo de páginas de un libro.
- El número de clientes que entran a una oficina postal en un día dado.
- El número de partículas α emitidas en un período fijo de tiempo por un elemento radiactivo.
- El número de temblores de tierra que ocurren durante un tiempo fijo.
- El número de guerras por año.
- El número de muertes en un período dado, de los clientes de una compañía de seguros.

Para finalizar esta distribución se enuncia el próximo teorema que, entre otras cosas, permite encontrar momentos de alto orden.

Teorema 4.7. Teorema de Hwang. Si $X \sim P(\lambda)$ y g es una función con $-\infty < E[g(X)] < \infty$, entonces

$$E[\lambda g(X)] = E[Xg(X-1)]$$

Ejemplo 4.5. Si $X \sim P(\lambda)$, entonces el segundo momento alrededor del origen, μ_2', se puede obtener empleando este teorema.

En efecto,

$$E(\lambda X) = E[X(X-1)] = E(X^2) - E(X)$$

Luego

$$E(X^2) = \mu_2' = E(\lambda X) + E(X) = \lambda \cdot \lambda + \lambda = \lambda^2 + \lambda.$$

4.1.4. Distribuciones geométrica y binomial negativa

Suponga que se desarrollan pruebas independientes, cada una con probabilidad de éxito p, hasta que un éxito ocurre. Si X denota el número de pruebas requeridas entonces,

$$p_X(n) = P(X = n) = \begin{cases} (1-p)^{n-1}p, & n = 1,\ 2,\ \dots \\ 0, & \text{en otro caso} \end{cases} \tag{4.6}$$

La ecuación 4.6 se da, ya que para $x = n$ es necesario y suficiente que las primeras $n-1$ pruebas sean fracasos y la n-ésima sea un éxito.

Definición 4.5. Una variable aleatoria X cuya función de masa de probabilidad está dada por la ecuación 4.6, se dice que tiene una distribución geométrica o de Pascal o de Furry, con parámetro p. Se abrevia por $X \sim G(p)$.

Teorema 4.8. Si X tiene una distribución geométrica con parámetro p entonces,

$$P(X \geq k) = (1-p)^{k-1}.$$

Demostración. Si $X \sim G(p)$ entonces,

$$\begin{aligned} P(X \geq k) &= \sum_{n=k}^{\infty}(1-p)^{n-1}p = p\sum_{n=k}^{\infty} q^{n-1},\ \text{donde } q = 1-p. \\ P(X \geq k) &= p(q^{k-1} + q^k + \cdots) = pq^{k-1}(1 + q + q^2 + \cdots) \\ &= pq^{k-1}\sum_{x=0}^{\infty} q^x = pq^{k-1}\left(\frac{1}{1-q}\right) = q^{k-1} = (1-p)^{k-1}. \end{aligned}$$

Esta fórmula permite encontrar cualquier percentil de esta variable aleatoria.

Ejemplo 4.6. Una urna contiene 15 bolas de las cuales 6 son blancas y 9 negras. Se seleccionan bolas una a una hasta obtener una bola blanca. Si se supone que la bola seleccionada se reemplaza antes de la próxima sacada, ¿cuál es la probabilidad de que,

a) sean necesarias exactamente 7 sacadas, y

b) al menos 8 sacadas sean necesarias?

Solución. Sea X el número de extracciones necesarias para obtener una bola blanca, entonces X tiene una distribución de Pascal con parámetro $p = 2/5$, ya que el proceso se realiza con reemplazo. Luego,

a) $P(X=7) = \left(1-\frac{2}{5}\right)^6\left(\frac{2}{5}\right) = \frac{3^6(2)}{5^7} = 0.019$

b) Por el teorema 4.8, $P(X \geq 8) = \left(1-\frac{2}{5}\right)^7 = \left(\frac{3}{5}\right)^7 = 0.028.$

Teorema 4.9. Si X tiene una distribución geométrica con parámetro p, entonces

$$M_X(t) = \frac{pe^t}{1-(1-p)e^t},\ E(X) = \frac{1}{p} \text{ y } V(X) = \frac{1-p}{p^2}.$$

Demostración. Si $X \sim G(p)$, entonces,

$$M_X(t) = \sum_{n=1}^{\infty} e^{tn}(1-p)^{n-1}p = \frac{p}{1-p}\sum_{n=1}^{\infty}[(1-p)e^t]^n = \frac{p}{1-p}\sum_{n=1}^{\infty} q^n,$$

donde $q = (1-p)e^t$. Luego

$$M_X(t) = \frac{pq}{1-p}\sum_{x=0}^{\infty} q^x = pe^t\frac{1}{1-q} = \frac{pe^t}{[1-(1-p)e^t]}$$

Ahora bien,

$$M'_X(t) = \frac{p\{e^t[1-(1-p)e^t] + e^{2t}(1-p)\}}{[1-(1-p)e^t]^2} = \frac{pe^t}{[1-(1-p)e^t]^2}$$

O sea que, $E(X) = M'_X(0) = \frac{p}{p^2} = \frac{1}{p}$

Análogamente se tiene que

$$M''_X(0) = \frac{2-p}{p^2},$$

de donde,

$$V(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Teorema 4.10. Si X tiene una distribución geométrica con parámetro p, entonces

$$P\left(X \geq i+j | X \geq i\right) = P(X > j).$$

Demostración.
(Ejercicio. Considere los eventos $A = \{X : X \geq i+j\}$ y $B = \{X : X \geq i\}$, entonces

el problema se reduce a trabajar con $P(A|B)$).

Una variable aleatoria X con distribución geométrica se reconoce, a menudo, como una variable aleatoria discreta de tiempos de espera. Esta representa cuánto se debe esperar para obtener un éxito.

Supóngase ahora que pruebas independientes, cada una con probabilidad de éxito p, se desarrollan hasta que un total de r éxitos son acumulados. Si X denota el número de pruebas requeridas, entonces

$$p_X(n) = \begin{cases} \binom{n-1}{r-1} p^r (1-p)^{n-r}, & n = r,\ r+1,\ \ldots \\ 0, & \text{en otro caso} \end{cases} \tag{4.7}$$

Definición 4.6. Una variable aleatoria X cuya función de masa de probabilidad esta dada por la ecuación 4.7, se dice que tiene una distribución binomial negativa con parámetros r y p [abreviando $X \sim Bn(r,p)$].

Teorema 4.11. Si X es una variable aleatoria con distribución binomial negativa entonces,

$$E(X) = \frac{r}{p},\ V(X) = \frac{r(1-p)}{p^2},\ \text{y } M_X(t) = \left[\frac{pe^t}{1-(1-p)e^t}\right]^r.$$

Demostración.
(Ejercicio)

Una variable aleatoria X que tiene una distribución binomial negativa se considera también, a menudo, como una variable aleatoria discreta de tiempos de espera. Esta representa cuánto se debe esperar para que r éxitos ocurran. Obsérvese que cuando $r = 1$, se tiene una distribución geométrica de parámetro p.

Ejemplo 4.7. En el ejemplo anterior, si las bolas son seleccionadas hasta que se obtienen 3 bolas blancas, entonces X el número de bolas blancas sacadas tiene distribución binomial negativa con parámetros 3 y $\frac{2}{5}$. Luego,

$$P(X = 7) = \binom{6}{2}\left(\frac{2}{5}\right)^3\left(\frac{3}{5}\right)^4 = 0.1244$$

Es de reconocer que formas especiales de la distribución binomial negativa fueron discutidas por Pascal en 1679. Una derivación como el número de lanzadas necesarias de una moneda para obtener un número fijo de caras fue obtenido en 1714 por Montmort [14] en la solución del problema de puntos, uno de los problemas que motivó el comienzo de la teoría de probabilidad moderna en el siglo XVII y que llevó a Pascal al primer razonamiento explícito acerca de lo que hoy se conoce como el valor esperado de una variable aleatoria.

4.1.5. Distribución hipergeométrica

Suponga que una muestra de tamaño n se selecciona aleatoriamente sin reemplazamiento de una urna que contiene M bolas de las cuales K son blancas y $M-K$ son negras. Si X denota el número de bolas blancas seleccionadas entonces,

$$p_X(x) = \begin{cases} \frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}}, & x = 1,\ 2,\ \ldots,\ \min(n, K) \\ 0, & \text{en otro caso} \end{cases} \tag{4.8}$$

Definición 4.7. Una variable aleatoria X cuya función de masa de probabilidad está dada por la ecuación 4.8 para algunos valores de M, K y n , se dice que tiene una distribución hipergeométrica [abreviada $X \sim hip(M, K, n)$].

Ejemplo 4.8. En una caja hay 30 fusibles de los cuales 5 son defectuosos. Se toman 10 uno a uno y sin reemplazamiento, ¿cuál es la probabilidad que se encuentre un fusible defectuoso en la muestra?

Solución. Sea X el número de fusibles defectuosos en la muestra de 10. Entonces X tiene distribución hipergeométrica y por tanto

$$P(X = 1) = \frac{\binom{5}{1}\binom{25}{9}}{\binom{30}{10}} = 0.340.$$

Teorema 4.12. Si X tiene una distribución hipergeométrica con función de masa de probabilidad dada por la ecuación 4.8 entonces,

$$E(X) = \frac{nK}{M} \text{ y } V(X) = \frac{nK}{M} \cdot \frac{M-K}{M} \cdot \frac{M-n}{M-1}.$$

Demostración. Si X tiene distribución hipergeométrica con parámetros M, K y n entonces,

$$E(X) = \sum_{x=0}^{n} x \frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} = \sum_{x=1}^{n} \left\{ \left[\frac{K!}{(x-1)!(K-x)!} \right] \frac{\binom{M-K}{n-x}}{\binom{M}{n}} \right\},$$

si $n = \min(n, K)$.

$$E(X) = K \sum_{x=1}^{n} \frac{\binom{K-1}{x-1}\binom{M-K}{n-x}}{\binom{M}{n}} = \left(\frac{nK}{M}\right) \sum_{y=0}^{n-1} \frac{\binom{K-1}{y}\binom{M-K}{n-1-y}}{\binom{M-1}{n-1}} = \frac{nK}{M}$$

ya que

$$\sum_{i=0}^{m} \binom{a}{i}\binom{b}{m-i} = \binom{a+b}{m}.$$

Para hallar $V(X)$ se calcula $E[X(X-1)]$ que es $n(n-1)[K(K-1)/M(M-1)]$ y como $V(X) = E[X(X-1)] + E(X) - [E(X)]^2$, se llega a lo que se quiere probar. El lector puede verificarlo.

Observación.
Si $p = K/M$, entonces la media de la distribución hipergeométrica coincide con la de la distribución binomial y la varianza es $(M-n)/(M-1)$ veces la varianza de la distribución binomial.

Otras variables aleatorias discretas con nombres conocidos son la *zeta* (o *zipf*), la beta-binomial y la logarítmica.

4.2. Distribuciones continuas

Hay varias clases importantes de variables aleatorias continuas que aparecen frecuentemente en teoría de probabilidad, entre otras se tienen las distribuciones uniforme, normal, gamma, chi cuadrado, Weibull, beta, lognormal, t de Student y F de Snedecor, las cuales se dicutirán a continuación.

4.2.1. Distribución uniforme

Definición 4.8. Una variable aleatoria X se dice que tiene una distribución uniforme en el intervalo (a, b) si su función de densidad de probabilidad esta dada por

$$f_X(x) = \frac{1}{b-a} I_{(a,b)}(x). \tag{4.9}$$

Es de anotar en la ecuación anterior que $f(x) = 0$ en cualquier otro caso, y si X tiene distribución uniforme en (a, b) se abrevia por $X \sim U(a, b)$.

La función de distribución es

$$F_X(x) = \frac{x-a}{b-a} I_{(a,b)}(x) + I_{[b,\infty)(x)},$$

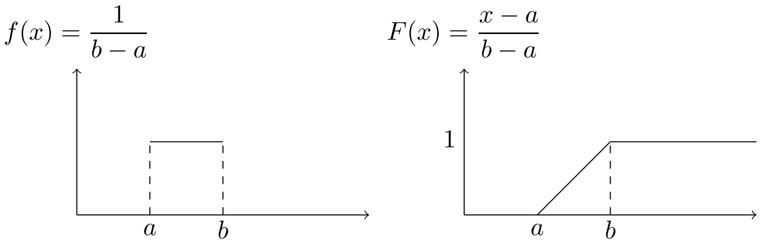
ya que

$$F_X(x) = \int_a^x f_X(y)\, dy.$$

En particular, si $a = 0$ y $b = 1$ entonces, $P(0 < X < \frac{1}{2}) = \frac{1}{2}$ y $P(\frac{1}{3} < X \leqslant \frac{3}{4}) = \frac{3}{4} - \frac{1}{3} = \frac{5}{12}$.

En general, si $a \leqslant \alpha \leqslant \beta \leqslant b$, entonces $P(\alpha < X \leqslant \beta) = \frac{\beta-\alpha}{b-a}$ (ver definición 2.12).

La gráfica de f es constante en (a, b) y la de F es una función continua creciente, como se muestra en la figura 4.1.



(a) Funcion de densidad (b) Función de distribución

Figura 4.1: Función de densidad y de distribución uniforme

Teorema 4.13. Si X tiene una distribución uniforme en (a, b) entonces,

$$E(X) = \frac{(a+b)}{2}, \ V(X) = \frac{(b-a)^2}{12}, \ \text{y } M_X(t) = \frac{e^{bt} - e^{at}}{t(b-a)}.$$

Demostración. Si $X \sim U(a, b)$ entonces,

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\ dx = \int_a^b \frac{x}{b-a}\ dx = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

$$E(X^2) = \int_a^b \frac{x^2}{b-a}\ dx = \frac{x^3}{3(b-a)}\Big|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3} \quad \text{y}$$

$$V(X) = \frac{b^2+ab+a^2}{3} - \frac{(a+b)^2}{4} = \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}.$$

Además,

$$M_X(t) = E(e^{tX}) = \int_a^b \frac{e^{tx}}{b-a}\ dx = \frac{e^{tx}}{t(b-a)}\Big|_a^b = \frac{e^{tb}-e^{ta}}{t(b-a)}.$$

Ejemplo 4.9. Si X tiene distribución uniforme en (-3,7) calcular
a) $P(X > 5)$, b) $P(-1 < X < 3)$.

Solución. Si $X \sim U(-3, 7)$, entonces $F(x) = \frac{x+3}{10}$ y por tanto

a) $P(X > 5) = 1 - F(5) = 1 - \frac{8}{10} = \frac{1}{5}$
b) $P(-1 < X < 3) = F(3) - F(-1) = \frac{6}{10} - \frac{2}{10} = \frac{2}{5}$.

4.2.2. Distribución normal

La gran mayoría de las técnicas usadas en estadística aplicada están basadas en la distribución normal. Por eso, se considera esta distribución como una de las más importantes. El "teorema central del límite" muestra su gran importancia.

Definición 4.9. Una variable aleatoria X se dice que es normal, o que está distribuida normalmente con parámetros μ y σ^2, si su función de densidad está dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \tag{4.10}$$

donde $-\infty < x < \infty$ y los parámetros μ y σ satisfacen $-\infty < \mu < \infty$ y $\sigma > 0$.

La gráfica de la ecuación 4.10 es una curva "acampanada" que es simétrica alrededor de μ, como se muestra en la figura 4.2.

Los valores de μ y σ^2 representan, en algún sentido, el valor promedio y la variación posible de X. El teorema 4.16 confirma esto.

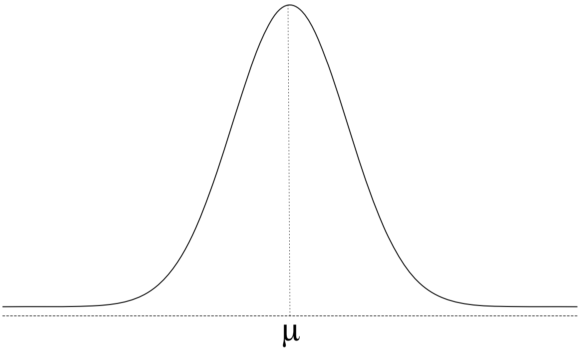


Figura 4.2. Funcion de densidad normal

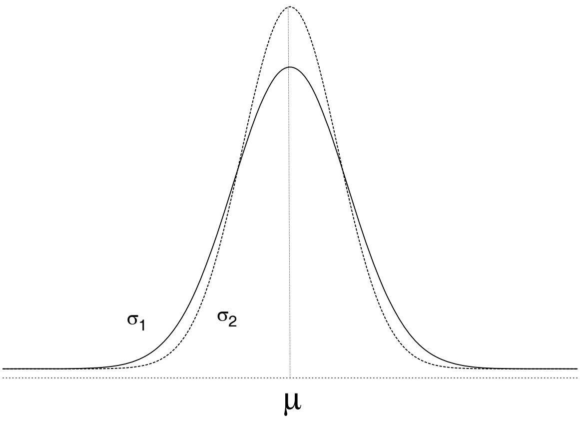


Figura 4.3. Densidades normales con σ diferentes

Una situación importante sobre estos valores es la siguiente: si se fija μ y se cambia la varianza, la gráfica de la función de densidad cambia en la forma, es más achatada una que otra como se muestra en la gráfica 4.3.

Si en cambio, se mantiene σ^2 fija y se varía μ, la gráfica de f se traslada, como se observa de la figura 4.4.

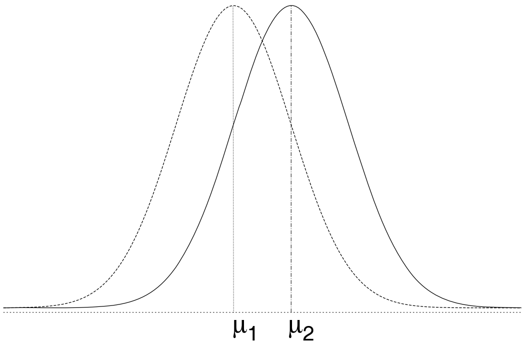


Figura 4.4. Densidades normales con μ diferentes

La distribución normal fue introducida por Abraham de Moivre en 1733 [8] y la usó para aproximar probabilidades asociadas con la distribución binomial cuando el parámetro n es grande. Este resultado lo extendieron más tarde Laplace y Moivre en lo que se conoce hoy como el teorema central del límite.

Se abrevia $X \sim N(\mu, \sigma^2)$ para decir que X tiene una distribución normal con parámetros μ y σ^2.

Teorema 4.14. Si $X \sim (\mu, \sigma^2)$ y si $Y = aX + b$, donde a y b son constantes con $a \neq 0$, entonces $Y \sim N(a\mu + b, a^2\sigma^2)$.

Demostración. Sea $Y = aX + b$ y $a > 0$. Entonces

$$\begin{aligned} F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) \\ = F_X\left(\frac{y-b}{a}\right). \end{aligned}$$

Luego,

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right)\cdot\frac{1}{a}.$$

Como $X \sim N(\mu,\sigma^2)$ se tiene que

$$f_Y(y) = \frac{e^{-\frac{1}{2\sigma^2}\left(\frac{y-b}{a}-\mu\right)^2}}{\sqrt{2\pi}\sigma}\cdot\frac{1}{a} = \frac{1}{\sqrt{2\pi}(a\sigma)}e^{-\frac{1}{2(a\sigma)^2}[y-(a\mu+b)]^2}. \tag{4.11}$$

Por analogía de la ecuación 4.11 con la 4.10 se concluye que

$$Y \sim N(a\mu+b, a^2\sigma^2).$$

Corolario 4.1. Si $X \sim N(\mu,\sigma^2)$ entonces $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$.

Tal variable Z se dice que tiene una distribución normal estándar o unitaria y sus funciones de densidad y de distribución se denotan por $\phi(z)$ y $\Phi(z)$, respectivamente. Es decir,

$$\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}, \quad \text{y} \quad \Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{\infty}^{z} e^{-\frac{y^2}{2}}\,dy \tag{4.12}$$

esta última se encuentra tabulada en la mayoría de los libros de estadística (ver apéndice D) y en el dominio de Internet.

Para valores no negativos de Z se tiene que $\Phi(-z) = 1-\Phi(z)$. Esto se debe a la simetría de la distribución normal.

Teorema 4.15. Si $X \sim N(\mu,\sigma^2)$ entonces,

$$P(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

En particular

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Demostración. Si $X \sim N(\mu,\sigma^2)$ entonces,

$$\begin{aligned}P(a < X < b) = P(a-\mu < X-\mu < b-\mu) &= P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right)\\ &= P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).\end{aligned}$$

En particular,

$$F_X(x) = P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Ejemplo 4.10. Si X tiene distribución normal con media 10 y varianza 36 calcular $P(X < 5)$, $P(4 < X < 16)$ y $P(X > 20)$.

Solución. Como $X \sim N(10, 36)$ entonces $F(x) = \Phi\left(\frac{x-10}{6}\right)$ y,

$$\begin{aligned} P(X < 5) &= \Phi\left(\frac{5-10}{6}\right) = \Phi\left(\frac{-5}{6}\right) = \Phi(-0.83) = 1 - \Phi(0.83) \\ &= 1 - 0.7967 = 0.2033. \end{aligned}$$

$$\begin{aligned} P(4 < X < 16) &= \Phi\left(\frac{16-10}{6}\right) - \Phi\left(\frac{4-10}{6}\right) = \phi(1) - \phi(-1) \\ &= 2\Phi(1) - 1 = 0.6826. \end{aligned}$$

$$\begin{aligned} P(X > 20) &= 1 - P(X \leq 20) = 1 - \Phi\left(\frac{20-10}{6}\right) = 1 - \Phi(1.66) \\ &= 1 - 0.9515 = 0.0485. \end{aligned}$$

Los valores de $\Phi(\cdot)$ se tomaron de la tabla de la distribución normal estándar (apéndice D).

Ejemplo 4.11. Un testigo experto en pleitos de paternidad testifica que, el período de embarazo está distribuido aproximadamente normal con parámetros $\mu = 270$ y $\sigma^2 = 100$. El acusado en el pleito es capaz de probar que estuvo fuera del país durante el período que comenzó 290 días antes del nacimiento del niño y finalizó 240 días antes del nacimiento. Si el acusado es el padre del niño, ¿cuál es la probabilidad que la madre haya tenido un embarazo muy largo o muy corto indicado por el testimonio?(Tomado de Ross[22].)

Solución. Sea X la variable aleatoria que denota el período de embarazo y suponga que el acusado es el padre. Luego, la probabilidad que haya tenido un embarazo muy largo o muy corto es

$$\begin{aligned}
P[(X>290)\cup(X<240)] &= P(X>290)+P(X<240)\\
&= 1-\Phi\left(\frac{290-270}{10}\right)+\Phi\left(\frac{240-270}{10}\right)\\
&= 1-\Phi(2)+\Phi(-3) = 0.0228+1-\Phi(3)\\
&= 0.0241
\end{aligned}$$

Teorema 4.16. Si X tiene distribución normal con parámetros μ y σ^2 entonces,

$$E(X)=\mu, V(X)=\sigma^2 \text{ y } M_X(t)=e^{t\mu+\frac{t^2\sigma^2}{2}}.$$

Demostración. Si $X\sim N(\mu,\sigma^2)$ entonces,

$$\begin{aligned}
M_X(t) &= E(e^{tX}) = E(e^{(tX-t\mu)}e^{t\mu}) = e^{t\mu}E[e^{t(X-\mu)}]\\
&= e^{t\mu}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}[(x-\mu)^2-2\sigma^2t(x-\mu)]}\,dx\\
&= e^{t\mu}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}[(x-\mu)^2-2\sigma^2t(x-\mu)+\sigma^4t^2]}e^{\frac{\sigma^2t^2}{2}}\,dx\\
&= e^{t\mu+\frac{\sigma^2t^2}{2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}[(x-\mu)-\sigma^2t]^2}\,dx.
\end{aligned}$$

La integral de la expresión anterior es 1 porque es la integral en los reales de una función de densidad y por lo tanto, $M_X(t)=e^{t\mu+\frac{\sigma^2t^2}{2}}$

Ahora bien,

$$M_X'(t)=(\mu+\sigma^2t)e^{t\mu+\frac{\sigma^2t^2}{2}} \text{ y } M_X''(t)=[\sigma^2+(\mu+\sigma^2t)^2]e^{t\mu+\frac{\sigma^2t^2}{2}}.$$

De esta manera,

$$M_X'(0)=E(X)=\mu \text{ y } M_X''(0)=\sigma^2+\mu^2.$$

Como,

$$V(X)=M_X''(0)-[E(X)]^2 \text{ entonces } V(X)=\sigma^2+\mu^2-\mu^2=\sigma^2.$$

Si $Z \sim N(0,1)$ entonces,

$$\mu_r = \mu_r' = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^r e^{-\frac{z^2}{2}} dz.$$

Cuando r es impar, $\mu_r = 0$ y cuando r es par $\mu_r = (r-1)(r-3)\cdots 3\cdot 1$
Así es que,

$$E(Z) = \mu_1 = 0,\ V(Z) = \mu_2 = 1,\ \mu_3 = 0 \ y\ \mu_4 = 3.$$

El siguiente teorema es una aproximación normal a la distribución binomial y es un caso especial del teorema central límite, el cuál se demostrará en el capítulo 6. Esta aproximación es buena cuando $np(1-p) \geq 10$.

Teorema 4.17. Teorema límite de Moivre - Laplace. Si X_n denota el número de éxitos que ocurren cuando se desarrollan n pruebas independientes, cada una con probabilidad de éxito p entonces para $a < b$,

$$P\left(a \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq b\right) \approx \Phi(b) - \Phi(a),\ \text{cuando } n \to \infty.$$

Ejemplo 4.12. Suponga que se lanzan dos dados legales 400 veces. Si X denota el número de veces que la suma 7 ocurre, encontrar $P(70 < X < 100)$.

Solución. Al lanzar dos dados legales la probabilidad que la suma sea 7 es 6/36, como se observó en el capítulo 3. Luego X, el número de veces que la suma 7 ocurre, tiene una distribución binomial con parámetros $n = 400$ y $p = 1/6$.

Puesto que $np(1-p) = 55,55 > 10$ entonces se puede obtener la probabilidad requerida usando el teorema 4.17, es decir,

$$\begin{aligned} P(70 < X < 100) &= P\left(\frac{70-np}{\sqrt{np(1-p)}} \leq \frac{X-np}{\sqrt{np(1-p)}} \leq \frac{100-np}{\sqrt{np(1-p)}}\right) \\ &= P\left(\frac{70-66.66}{\sqrt{55.55}} \leq \frac{X-66.66}{\sqrt{55.55}} \leq \frac{100-66.66}{\sqrt{55.55}}\right) \\ &\approx \Phi(4.47) - \Phi(0.45) = 1 - 0.6736 = 0.3264. \end{aligned}$$

Para finalizar esta distribución se enuncia el siguiente lema el cual es útil entre otras cosas para obtener momentos de alto orden.

Teorema 4.18. Lema de Stein. Si $X \sim N(\mu, \sigma^2)$ y si g es una función diferenciable tal que $E[|g'(X)|] < \infty$, entonces

$$E[g(X)(X-\mu)] = \sigma^2 E[g'(X)].$$

Ejemplo 4.13. Si $X \sim N(\mu, \sigma^2)$, entonces el cuarto momento μ_4, es por el lema de Stein,

$$\begin{aligned}
\mu_4 = E[(X-\mu)^4] &= E[(X-\mu)^3(X-\mu)] \\
&= \sigma^2 E[3(X-\mu)^2] \\
&= 3\sigma^2 E[(X-\mu)^2] \\
&= 3\sigma^2 \cdot \sigma^2 \\
&= 3\sigma^4.
\end{aligned}$$

Análogamente, usando el lema se puede obtener que, el segundo momento respecto al origen, μ_2' es

$$\begin{aligned}
\mu_2' = E(X^2) &= E[X(X-\mu+\mu)] \\
&= E[X(X-\mu)] + \mu E(X) \\
&= \sigma^2 E(1) + \mu^2 \\
&= \sigma^2 + \mu^2.
\end{aligned}$$

4.2.3. Distribuciones exponencial, gamma y chi-cuadrado

Estas tres distribuciones juegan papeles importantes en estadística y la razón por la que se consideran en la misma sección es, entre otras, que la exponencial y la chi-cuadrado son casos especiales de la gamma. La distribución gamma fue utilizada en 1947 por Brown y Flood [13] para describir la vida de los vasos de vidrio que circulan en una cafetería y desde 1958 como un modelo para problemas de confiabilidad industrial.

Definición 4.10. Distribución exponencial. Una variable aleatoria X se dice que tiene una distribución exponencial con parámetro λ, $\lambda > 0$, si su función de densidad está dada por

$$f_X(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x). \tag{4.13}$$

Es de anotar que $f(x) = 0$ cuando $x < 0$, y para decir que X tiene distribución exponencial con parámetro λ se abrevia $X \sim E(\lambda)$.

La función de distribución F de una variable aleatoria exponencial con parámetro λ es

$$F_X(x) = 1 - e^{-\lambda x}, \ x \geq 0.$$

En efecto, si $x > 0$

$$F_X(x) = P(X \leq x) = \int_0^x \lambda e^{-\lambda y} \ dy = -e^{-\lambda y}\Big|_0^x = 1 - e^{-\lambda x}.$$

La distribución exponencial aparece, a menudo, como la distribución del tiempo hasta que algún evento específico ocurre.

Definición 4.11. Una variable aleatoria X no negativa, se dice que no tiene memoria si

$$P(X > s + t | X > t) = P(X > s), \text{ para todo } s, \ t \geq 0.$$

Teorema 4.19. Si $X \sim E(\lambda)$ entonces X no tiene memoria.

Demostración. Si X tiene una distribución exponencial con parámetro λ entonces

$$F_X(x) = 1 - e^{-\lambda x} \text{ y por tanto, } P(X > x) = e^{-\lambda x}.$$

Luego,

$$\begin{aligned} P(X > s + t | X > t) &= \frac{P(X > s + t; X > t)}{P(X > t)} = \frac{P(X > s + t)}{P(X > t)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = \frac{e^{-\lambda s - \lambda t}}{e^{-\lambda t}} \\ &= e^{-\lambda s} = P(X > s). \end{aligned}$$

La media, la varianza y la función generatriz de momentos de la distribución exponencial, son una consecuencia inmediata del teorema 4.20 de la distribución gama.

Definición 4.12. Distribución gamma. Una variable aleatoria X se dice que tiene una distribución gamma con parámetros (r, λ)[la cual se abrevia $X \sim G(r, \lambda)$], si su función de densidad está dada por

$$f_X(x) = \frac{\lambda e^{-\lambda x}(\lambda x)^{r-1}}{\Gamma(r)}\ I_{(0,\infty)}. \tag{4.14}$$

donde $\Gamma(\cdot)$ es la función gamma(ver apéndice B).

La distribución gamma con parámetros (r, λ) se conoce como la distribución *de Erlang* cuando $r = n$ y n es un entero positivo, representa la distribución del tiempo que se tiene que esperar hasta que un total de n eventos ocurren. Cuando $r = 1$, la distribución gamma se reduce a la exponencial, y cuando $\lambda = 1/2$ y $r = n/2$, para n entero positivo se le conoce con el nombre de distribución chi-cuadrado (χ_n^2) con n grados de libertad. En este caso la función de densidad se transforma en

$$f_X(x) = \frac{e^{-\frac{x}{2}} x^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})},\ x > 0. \tag{4.15}$$

Teorema 4.20. Si X tiene una distribución gamma con parámetros (r, λ) entonces

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^r,\ E(X) = \frac{r}{\lambda}\ \text{ y } V(X) = \frac{r}{\lambda^2}.$$

Demostración.

$$\begin{aligned} M_X(t) &= \int_0^\infty \frac{e^{tx}\lambda e^{-\lambda x}(\lambda x)^{r-1}}{\Gamma(r)}\ dx = \int_0^\infty \frac{\lambda e^{-(\lambda-t)x}(\lambda x)^{r-1}}{\Gamma(r)}\ dx \\ &= \frac{\lambda^r}{(\lambda-t)^r}\int_0^\infty \frac{(\lambda-t)^r}{\Gamma(r)}\ x^{r-1}e^{-(\lambda-t)x}\ dx \end{aligned}$$

Sea $u = (\lambda - t)x$ entonces $dx = \dfrac{du}{\lambda - t}$. Luego,

$$M_X(t) = \frac{\lambda^r}{(\lambda-t)^r}\int_0^\infty \frac{u^{r-1}e^{-u}}{\Gamma(r)}\ du = \frac{\lambda^r}{(\lambda-t)^r}.$$

De este resultado se tiene que

$$M_X'(t) = r\lambda^r(\lambda-t)^{-r-1} \text{ y } M_X''(t) = r(r+1)\lambda^r(\lambda-t)^{-r-2}.$$

Luego,

$$E(X) = M_X'(0) = \frac{r}{\lambda} \quad \text{y} \quad V(X) = M_X''(0) - [E(X)]^2 = \frac{r}{\lambda^2}.$$

Corolario. Si $X \sim G(r, \lambda)$ y $r = 1$ entonces,

$$E(X) = \frac{1}{\lambda},\ V(X) = \frac{1}{\lambda^2} \text{ y } M(X) = \frac{\lambda}{\lambda - t}.$$

que son la media, la varianza y la función generatriz de momentos de la distribución exponencial.

Corolario. Si $X \sim G(r, \lambda)$, $\lambda = 1/2$ y $r = n/2$, $n \in \mathbb{Z}^+$ entonces,

$$E(X) = n,\ V(X) = 2n \text{ y } M_X(t) = \left(\frac{1}{1-2t}\right)^{n/2},$$

que son la media, la varianza y la función generatriz de momentos de la distribución chi-cuadrado con n grados de libertad.

Se enuncia a continuación un teorema que permite hallar cualquier probabilidad para una distribución gamma cuando r es un entero positivo.

Teorema 4.21.. Si $X \sim G(n, \lambda)$, donde $n \in Z^+$, entonces

$$F(x) = 1 - \sum_{i=0}^{n-1} \frac{e^{-\lambda x}(\lambda x)^i}{i!}.$$

Ejemplo 4.14.. Si $X \sim G(7, 3)$ entonces, utilizando el teorema anterior

$$P(X > 4) = 1 - F(4) = \sum_{i=0}^{6} \frac{e^{-12}12^i}{i!} = 0.0458.$$

4.2.4. Distribución Weibull

Esta distribución es ampliamente usada en ingeniería debido a su versatilidad y también en análisis de vida. Toma su nombre del físico sueco Waloddi Weibull, quién en 1939 (A Statistical Theory of the Strenght of Materials) [13] la utilizó para representar la distribución de la ruptura de la resistencia de los materiales. Ha sido utilizada exitosamente en teoría de la confiabilidad, puede representar la distribución del tiempo de vida de un objeto consistente de muchas partes, en la cual el objeto falla cuando una de las partes falla.

Definición 4.13. Distribución Weibull. Una variable aleatoria X se dice que tiene una distribución Weibull con parámetros a y b[abreviada $X \sim W(a, b)$], si su función de densidad de probabilidad está dada por

$$f_X(x) = abx^{b-1}e^{-ax^b}, \ x > 0, \ a > 0, \ b > 0. \tag{4.16}$$

Puede mostrarse que la función de distribución de una variable aleatoria X con distribución Weibull es

$$F_X(x) = 1 - e^{-ax^b}, \ x > 0. \tag{4.17}$$

Teorema 4.22. Si X tiene una distribución Weibull con parámetros a y b entonces,

$$E(X) = \frac{\Gamma(1+\frac{1}{b})}{a^{1/b}}, \ V(X) = \frac{\Gamma(1+\frac{2}{b}) - \Gamma^2(1+\frac{1}{b})}{a^{2/b}} \text{ y } \mu_k' = \frac{\Gamma(1+\frac{k}{b})}{a^{k/b}}.$$

Demostración. Primero se prueba la igualdad para μ_k' y con base en esta se obtienen la media y la varianza.

$$\begin{aligned}
\mu_k' = E(X^k) &= \int_0^\infty x^k abx^{b-1}e^{-ax^b} \ dx \\
&= \int_0^\infty abx^{k+b-1}e^{-ax^b} \ dx \\
&= \frac{1}{a^{k/b}} \int_0^\infty a^{(k/b)+1}bx^{k+b-1}e^{-ax^b} \ dx.
\end{aligned}$$

Si $y = ax^b$, entonces $dy = abx^{b-1}dx$ y

$$\mu_k' = \frac{1}{a^{k/b}} \int_0^\infty y^{k/b}e^{-y} \ dy = \frac{\Gamma(1+\frac{k}{b})}{a^{k/b}}.$$

De esta manera,

$$E(X) = \mu_1' = \frac{\Gamma(1+\frac{1}{b})}{a^{1/b}} \text{ y } E(X^2) = \mu_2' = \frac{\Gamma(1+\frac{2}{b})}{a^{2/b}}.$$

Por lo tanto,

$$V(X) = \frac{\Gamma(1+\frac{2}{b})}{a^{2/b}} - \left[\frac{\Gamma(1+\frac{1}{b})}{a^{1/b}}\right]^2 = \frac{\Gamma(1+\frac{2}{b}) - \Gamma^2(1+\frac{1}{b})}{a^{2/b}}.$$

Cuando el parámetro b es grande $E(X)$ y $V(X)$ se pueden aproximar (McEwen y Parresol 1991)[13] por

$$E(X) = 1 - \frac{\gamma}{b} + \frac{\pi/6 + \gamma^2}{2b^2}, \text{ y } V(X) = \frac{\pi^2}{6b^2},$$

donde γ es la constante de Euler(ver apéndice B).

Observación.
La distribución Weibull con parámetros a y b se reduce a la exponencial con parámetro a cuando $b = 1$.

4.2.5. Distribución beta

Definición 4.14. Una variable X se dice que tiene una distribución beta con parámetros a y b [abreviada $X \sim Be(a, b)$], si su función de densidad está dada por

$$f_X(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \; 0 < x < 1, a > 0, \; b > 0 \tag{4.18}$$

donde

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} \, dx,$$

es la función beta que tiene la propiedad $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Teorema 4.23. Si X tiene una distribución beta con parámetros a y b entonces,

$$E(X) = \frac{a}{a+b} \quad \text{y} \quad V(X) = \frac{ab}{(a+b+1)(a+b)^2}.$$

Demostración.

$$\begin{aligned} E(X) &= \int_0^1 \frac{x^a(1-x)^{b-1}}{B(a,b)} \, dx = \frac{B(a+1,b)}{B(a,b)} = \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)B(a+b)} \\ &= \frac{a\Gamma(a)\Gamma(b)}{(a+b)\Gamma(a+b)B(a,b)} = \left(\frac{a}{a+b}\right)\left[\frac{B(a,b)}{B(a,b)}\right] = \frac{a}{a+b}. \end{aligned}$$

Ahora,

$$E(X^2) = \int_0^1 \frac{x^{a+1}(1-x)^{b-1}}{B(a,b)}\, dx = \frac{B(a+2,b)}{B(a,b)}$$
$$= \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)B(a,b)} = \frac{a(a+1)}{(a+b)(a+b+1)}.$$

Luego,

$$V(X) = \frac{a(a+1)}{(a+b)(a+b+1)} - \frac{a^2}{(a+b)^2} = \frac{ab}{(a+b+1)(a+b)^2}.$$

4.2.6. Distribución lognormal

El origen se remonta al año 1879, cuando McAlister[13] describió explicitamente una teoría de esta distribución. En 1945 Gaddum dio una aplicación en biología y luego se han mostrado aplicaciones en medicina y economía.

Se dice que si existe un número θ tal que $W = \ln(X-\theta)$ está normalmente distribuida, entonces la distribución de X es lognormal. Para esto es necesario que $X > \theta$ pero que tenga probabilidad 0 cuando $X \leqslant \theta$.

Si $W = \ln(X-\theta) \sim N(\mu, \sigma^2)$, la distribución de X se obtiene de la transformación

$$Z = \frac{ln(X-\theta)-\mu}{\sigma} \tag{4.19}$$

donde $Z \sim N(0,1)$, μ, σ y θ son parámetros.

De 4.19 se sigue(ver teorema 6.1) que la función de densidad de probabilidad de X es

$$f_X(x) = \frac{1}{(x-\theta)\sqrt{2\pi}\sigma} e^{-\frac{1}{2}[\frac{ln(x-\theta)-\mu}{\sigma}]^2}, x > \theta, \sigma > 0. \tag{4.20}$$

El valor del parámetro θ afecta solo la localización de la distribución y no afecta la varianza ni la forma. Por esto, si $\theta = 0$, la distribución se conoce como “distribución lognormal de dos parámetros ”, así que $Z = \frac{lnX-\mu}{\sigma} \sim N(0,1)$ y

$$f_X(x) = \frac{1}{x\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left[\frac{lnx-\mu}{\sigma}\right]^2}, x > 0. \tag{4.21}$$

Teorema 4.24. Si X tiene una distribución lognormal con dos parámetros y con función de densidad definida por 4.21 entonces,

$$\mu_r' = e^{r\mu+\frac{1}{2}r^2\sigma^2}; E(X) = e^{\mu+\frac{\sigma^2}{2}}, \text{ y } V(X) = e^{2\mu+\sigma^2}[e^{\sigma^2} - 1].$$

Demostración. Sea $Y = lnX$ entonces por definición $Y \sim N(\mu, \sigma^2)$ y por consiguiente $M_Y(t) = e^{t\mu+\frac{t^2\sigma^2}{2}}$. Luego,

$$\mu_r' = E(X^r) = E(e^{rY}) = M_Y(r) = e^{r\mu+\frac{r^2\sigma^2}{2}}.$$

De este resultado se tiene que,

$$E(X) = \mu_1' = e^{\mu+\frac{\sigma^2}{2}} \text{ y } E(X^2) = \mu_2' = e^{2\mu+2\sigma^2}.$$

Por tanto,

$$V(X) = E(X^2) - [E(X)]^2 = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1).$$

Nota

Si X tiene distribución lognormal con parámetros μ y σ^2 se abrevia por $X \sim$ logn(μ, σ^2).

4.2.7. Distribuciones *t* de Student y *F* de Snedecor

Estas dos últimas distribuciones que se presentan son fundamentales en la inferencia estadística. Junto con la distribución chi-cuadrado, son base para la construcción de intervalos de confianza y las pruebas de hipótesis (capítulos 9 y 10).

Definición 4.15. Distribución *t*. Una variable aleatoria X se dice que tiene una distribución t de Student (o simplemente t) con n grados de libertad, [abreviada $X \sim t_n$], si su función de densidad está dada por

$$f_X(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}}\left(\frac{1+x^2}{n}\right)^{-(n+1)/2}, \ -\infty < x < \infty. \tag{4.22}$$

Cuando $n = 1$ la distribución t recibe el nombre de distribución de *Cauchy*, cuya función de densidad es

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \ -\infty < x < \infty. \tag{4.23}$$

Esta distribución se caracteriza porque no tiene ni media ni varianza.

La distribución t es simétrica y su función de distribución $F_X(x)$ se encuentra tabulada en el apéndice D, tabla D.3.

Teorema 4.25. Si X tiene una distribución t de Student con n grados de libertad entonces,

$$E(X) = 0, \quad \text{para } n > 1 \ \text{ y } \quad V(X) = \frac{n}{n-2}, \ n > 2.$$

La demostración de este teorema es evidente para $E(X)$, sin embargo para el cálculo de la varianza el trabajo es de cuidado.

Definición 4.16. Distribución *F*. Una variable aleatoria X se dice que tiene una distribución F de Snedecor (o simplemente F) con m y n grados de libertad [abreviada $X \sim F(m,n)$], si su función de densidad está dada por

$$f_X(x) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} \left[\frac{x^{\frac{m}{2}-1}}{[1+\frac{m}{n}x]^{(m+n)/2}}\right], \ x \geq 0. \tag{4.24}$$

Cuando $m = n = 1$ la distribución F se convierte en una distribución con función de densidad dada por

$$f_X(x) = \frac{1}{\pi(1+x)x^{1/2}}, \ x \geq 0. \tag{4.25}$$

Lo cual no tiene media ni varianza.

Teorema 4.26. Si X tiene una distribución F con m y n grados de libertad entonces,

$$E(X) = \frac{n}{n-2}, \ n > 2 \quad \text{y} \quad V(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, \ n > 4.$$

Esta demostración también se hará más sencilla, cuando en el capítulo 7 se demuestre que la distribución F es el cociente entre dos chi-cuadrado, cada una de ellas dividida por sus grados de libertad.

Otras distribuciones continuas con nombres específicos y que también tienen su importancia son: la doble exponencial, la logística, Pareto y Gumbel. Estas se pueden ver en [20].

4.3. Ejercicios

1. Si X es una variable aleatoria que tiene una distribución de Poisson con parámetro λ, probar que si λ no es entero, $p_X(x)$ tiene una moda única en $x = \lambda$. Si λ es entero, probar que $p_X(x)$ tiene dos modas obtenidas en $x = \lambda$ y $x = \lambda - 1$.

2. Si X es una variable aleatoria que tiene una distribución binomial con parámetros 25 y 0.2 evaluar,
$$P\big[X < E(X) - 2\sqrt{V(X)}\big].$$

3. En un cruce de cierta vía ocurren en promedio 5 accidentes de tránsito por mes. ¿Cuál es la probabilidad que para cualquier mes dado en este cruce ocurran
   a) exactamente 5 accidentes?
   b) al menos 8 accidentes?
   c) menos de 4 accidentes?

4. La probabilidad que un paciente se recupere después de una delicada operación de corazón es 0.9. ¿Cuál es la probabilidad que
   a) exactamente 7 de los siguientes 10 pacientes intervenidos sobrevivan?
   b) al menos 3 y a lo mas 8 de los siguientes 10 pacientes intervenidos sobrevivan?
   c) ¿Cuál es el promedio de sobrevida de estos 10 pacientes?

5. Se lanza un dado hasta que un 3 aparece. ¿Cuál es la probabilidad que este deba ser lanzado más de 6 veces?

6. Si la variable aleatoria X tiene una distribución geométrica con parámetro p, demostrar que
$$P\Big(X \geq n - 1 + k | X \geq n - 1\Big) = P(X > k).$$

7. Si la variable aleatoria X tiene una distribución binomial con parámetros n y p, ¿cuál es la distribución de Y, si $Y = n - X$? (Sugerencia: Use la función generatriz de momentos).

8. Si la variable aleatoria X tiene una distribución hipergeométrica, demostrar que

$$\frac{\binom{K}{k}\binom{M-K}{n-k}}{\binom{N}{n}} \longrightarrow \binom{n}{k} p^k (1-p)^{n-k}$$

cuando $M \to \infty$ y $p = \dfrac{K}{M}$.

9. Si la variable aleatoria X tiene una distribución uniforme en $(-\delta, \delta)$. Hallar los valores de δ para los cuales

   a) $P(-1 < X < 2) = 0.75$.

   b) $P(|X| < 1) = P(|X| > 2)$.

10. El diámetro de las piezas que fabrica una maquina es una variable aleatoria X que tiene distribución uniforme.Si el diámetro más pequeño posible es 10 y el diámetro medio es 13 mm,

    a) Obtenga la distribución de X.

    b) Una pieza es apta si su diámetro está entre 11 y 15 mm. Calcular la probabilidad que una pieza sea defectuosa.

    c) Si en un día se fabrican 12 piezas, obtener la probabilidad que a lo más 5 sean defectuosas.

11. Los buses llegan a una parada específica con intervalos de 15 minutos iniciándose desde las 6:30 a.m. Si un pasajero llega a la parada en un tiempo que está distribuido uniformemente entre las 6:30 y las 7:30, encontrar la probabilidad que él espere,

    a) menos de 7 minutos por un bus.

    b) más de 12 minutos por un bus.

12. Si X tiene una distribución normal con parámetros $(12, 16)$, hallar

    a) $P(X < 8)$.

    b) $P(X > 11)$.

    c) $P(7 < X < 20)$.

13. Un dado legal se lanza mil veces. Calcular una aproximación a la probabilidad que el número 6 aparezca entre 150 y 200 veces.

14. Un cierto tipo raro de sangre se encuentra en solo el 0.05 % de la gente. Si la muestra de un grupo seleccionado al azar es 30.000, ¿cuál es la probabilidad (exacta y aproximada) que al menos 5 personas en el grupo tengan este tipo raro de sangre?

15. Si la variable aleatoria X tiene una distribución exponencial con parámetro λ, hallar la mediana de X. Además, si $P(X > t) = \delta$, expresar el parámetro λ en términos de δ y t.

16. Suponga que el tiempo de llegada T, entre dos clientes que entran a una oficina postal satisface,

$$P(T > t) = \alpha e^{-\lambda t} + \beta e^{-\mu t}, t \geq 0$$

donde $\alpha + \beta = 1$, $\alpha \geq 0$, $\beta \geq 0$, $\lambda > 0$ y $\mu > 0$. Calcule $E(T)$ y $V(T)$.

17. Si la variable aleatoria Y tiene una distribución uniforme en $(-3, 5)$, ¿cuál es la probabilidad que las raíces de la ecuación $2x^2 + 2Yx + \frac{Y}{2} + 1 = 0$ sean ambas reales?

18. Si la variable aleatoria X tiene una distribución normal con media μ y varianza 9. Halle aproximadamente μ si $P(X > 4) = 0.1$

19. Una variable aleatoria X tiene función de densidad $f_X(x) = 3(1-x)^2 I_{[0,1]}(x)$.

    a) ¿Que distribución tiene X?
    b) Halle $E(X)$, $V(X)$ y $M_X(t)$.

20. Sea X una variable aleatoria con distribución chi-cuadrado con 10 grados de libertad, usando la tabla D.2 del apéndice D determinar a y b tales que,

$$P(X < a) = P(X > b) \text{ y } P(a < X < b) = 0.90$$

21. Para la variable aletoria X con función de densidad dada por

$$f_X(x) = \frac{2}{\theta} x e^{-\frac{x^2}{\theta}} I_{(0,\infty)}(x),\ \theta > 0 \tag{4.26}$$

conocida como la distribución de *Rayleig*, hallar $F(x)$, $E(X)$ y $V(X)$.

22. Para la variable aleatoria X con función de densidad dada por

$$f_X(x) = \left(\frac{4}{\theta^3 \pi}\right) x^2 e^{-\frac{x^2}{\theta^2}} I_{(0,\infty)}(x),\ \theta > 0$$

conocida como la distribución de *Maxwell*, hallar $F(x)$, $E(X)$ y $V(X)$.

23. Sean a y b constantes positivas, verificar que la función

$$f(x) = \begin{cases} \frac{a}{b}(\frac{b}{x})^{a+1}, & x \geq b \\ 0, & x < b \end{cases}$$

es la función de densidad de una variable aleatoria X. Calcular $F(x)$, $E(X)$ y $V(X)$.

24. Si $Z \sim N(0,1)$ probar que $\mu_r = 0$ si r es impar y $\mu_r = (r-1)(r-3)\cdots 3 \cdot 1$ si r es par.

25. Si $X \sim W(a,b)$ probar que $F(x) = 1 - e^{-ax^b}$.

26. Demostrar el teorema 4.6

27. Demuestre el teorema de Hwang.

28. Demostrar el teorema 4.10.

29. Demostrar el teorema 4.11.

30. Demostrar el teorema 4.17.

31. Demuestre el lema de Stein.

32. Demuestre el teorema 4.21

33. Demuestre el teorema 4.24

Capítulo 5

Variables aleatorias distribuidas conjuntamente

En la realización de un experimento aleatorio, es común estar interesado en más de una variable aleatoria, de manera que se hace necesario extender la definición de distribución de una variable aleatoria a casos más generales.

5.1. Funciones de distribución, de masa y de densidad de probabilidad conjunta

Sean $X_1,\ X_2,\ \ldots,\ X_n$ variables aleatorias definidas en un espacio de probabilidad $(\Omega, \boldsymbol{\mathcal{F}}, P)$ y sean $x_1, x_2, ..., x_n$ números reales. Como $\{w \in \Omega : X_k(w) \leq x_k\} \in \boldsymbol{\mathcal{F}}$, también $\cap_{k=1}^{n}\{w \in \Omega : X_k(w) \leq x_k\} \in \boldsymbol{\mathcal{F}}$.

Definición 5.1. Se define la función de distribución conjunta de $X_1,\ X_2,\ \ldots,\ X_n$ o la función de distribución del vector $X = (X_1,\ X_2,\ \ldots,\ X_n)$ por

$$F(x_1, x_2, ..., x_n) = P\big(\cap_{k=1}^{n}\{w \in \Omega : X_k(w) \leq x_k\}\big).$$

Esta función se denota también por $P(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n)$, o sea que,

$$F(x_1, x_2, ..., x_n) = P(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n). \tag{5.1}$$

Teorema 5.1. Propiedades de F. Sean $X_1,\ X_2,\ \ldots,\ X_n$ variables aleatorias definidas en un espacio de probabilidad $(\Omega, \boldsymbol{\mathcal{F}}, P)$ entonces, a.) Para todo $x_1, x_2, ..., x_n$ se cumple que $0 \leq F(x_1, x_2, ..., x_n) \leq 1$.

b.) $F(x_1, x_2, ..., x_n)$ es no decreciente en cada uno de sus argumentos.

c.) $\lim\limits_{x_n \to \infty} F(x_1, x_2, ..., x_n) = F(x_1, x_2, ..., x_{n-1})$ y
$\lim\limits_{x_n \to -\infty} F(x_1, x_2, ..., x_n) = 0$.

La demostración se propone como ejercicio.

La propiedad c establece que la función de distribución conjunta de $n-1$ variables se obtiene de la función de distribución conjunta de n de esas variables.

En particular, si X y Y son variables aleatorias definidas en $(\Omega, \boldsymbol{\mathcal{F}}, P)$,

$$F(x, y) = P(X \leq x, Y \leq y) \tag{5.2}$$

$$F_X(x) = \lim_{y \to \infty} F(x, y) = F(x, \infty) \quad \text{y} \tag{5.3}$$

$$F_Y(y) = F(\infty, y). \tag{5.4}$$

Las funciones de distribución F_X y F_Y se llaman “funciones de distribución marginales” de X y Y respectivamente.

Observación.
Todas la probabilidades acerca de X y Y pueden obtenerse de la función de distribución conjunta, así

Para $a_1 < a_2$ y $b_1 < b_2$,

$$P(a_1 \leq X \leq a_2, b_1 \leq Y \leq b_2) = F(a_2, b_2) + F(a_1, b_1) - F(a_1, b_2) - F(a_2, b_1).$$

En particular,

$$P(X > a, Y > b) = 1 - F_X(a) - F_Y(b) + F(a, b).$$

Definición 5.2. Sean $X_1,\ X_2,\ \ldots,\ X_n$ variables aleatorias discretas. Se define la función de masa de probabilidad conjunta de $X_1, X_2, ..., X_n$, denotada como $p(x_1, x_2, ..., x_n)$, por

$$p(x_1, x_2, ..., x_n) = P(X_1 = x, X_2 = x_2, ..., X_n = x_n). \tag{5.5}$$

En particular, si X y Y son variables aleatorias discretas, la función de masa de probabilidad conjunta de X y Y se define por

$$p(x,y) = P(X = x, Y = y) \tag{5.6}$$

Observación.
La función de masa de probabilidad de X se obtiene de $p(x,y)$, así

$$p_X(x) = P(X = x) = P(X = x, Y < \infty) = \sum_{\{y:p(x,y)>0\}} p(x,y). \tag{5.7}$$

Análogamente,

$$p_Y(y) = \sum_{\{x:p(x,y)>0\}} p(x,y). \tag{5.8}$$

Ejemplo 5.1. Supóngase que se seleccionan 3 bolas aleatoriamente de una urna que contiene 3 bolas blancas, 4 negras y 5 rojas. Si X denota el número de bolas blancas y Y el número de bolas negras obtenidas, encontrar la distribución conjunta de X y Y. Además, encontrar las funciones de probabilidad marginales de X y de Y.

Solución. Las variables aleatorias X y Y toman valores $0, 1, 2$ y 3 y la tabla 5.1 muestra la distribución conjunta de X y Y,

Y\X	0	1	2	3	$p_Y(y)$
0	$\frac{10}{220}$	$\frac{30}{220}$	$\frac{15}{220}$	$\frac{1}{220}$	$\frac{56}{220}$
1	$\frac{40}{220}$	$\frac{60}{220}$	$\frac{12}{220}$	0	$\frac{112}{220}$
2	$\frac{30}{220}$	$\frac{18}{220}$	0	0	$\frac{48}{220}$
3	$\frac{4}{220}$	0	0	0	$\frac{4}{220}$
$p_X(x)$	$\frac{84}{220}$	$\frac{108}{220}$	$\frac{27}{220}$	$\frac{1}{220}$	1

Tabla 5.1: Distribución Conjunta de las variables X y Y

donde por ejemplo

$$\begin{aligned}
P(0,0) &= P(\text{obtener tres bolas rojas}) = \frac{\binom{5}{3}}{\binom{12}{3}} = \frac{10}{220} \\
P(1,1) &= P(\text{obtener una bola blanca, una negra y una roja}) \\
&= \frac{\binom{3}{1}\binom{4}{1}\binom{5}{1}}{\binom{12}{3}} = \frac{60}{220} \\
P(2,1) &= P(\text{obtener dos bolas blancas y una negra}) \\
&= \frac{\binom{3}{2}\binom{4}{1}}{\binom{12}{3}} = \frac{12}{220}.
\end{aligned}$$

La última fila y la última columna de la tabla dan la función de probabilidad marginal de X y de Y, respectivamente. Estas se obtuvieron usando las ecuaciones 5.7 y 5.8, por ejemplo

$$p_X(0) = \sum_{y=0}^{3} p(0,y) = \frac{84}{220} \text{ y } p_Y(1) = \sum_{x=1}^{3} p(x,1) = \frac{112}{220}.$$

Considere ahora la siguiente situación: un experimento tiene N resultados posibles, sea $\Omega = \{w_1, w_2, ..., w_N\}$ y $p_j = P(\{w_j\})$ para $j = 1, 2, ..., N$. Si el experimento se repite n veces y X_j representa el número de veces que ocurre el resultado w_j, cuando las repeticiones son independientes, entonces

$$p(x_1, x_2, ..., x_n) = \frac{n!}{x_1!x_2!...x_n!} p_1^{x_1} p_2^{x_2} ... p_N^{x_n}, \tag{5.9}$$

para $x_i = 0, 1, .., n$ y, $\sum_{i=1}^{n} x_i = n$.

Definición 5.3. Una variable aleatoria $X = (X_1, X_2, ..., X_n)$ cuya función de masa de probabilidad está dada por la expresión anterior se dice que tiene una distribución *multinomial* con parámetros $p_1, p_2, ..., p_N$.

Definición 5.4. Las variables aleatorias $X_1,\ X_2,\ \ldots,\ X_n$ son continuas conjuntamente si existe una función $f(x_1, x_2, \ldots, x_n)$, no negativa tal que,

$$F(a_1, a_2, \ldots, a_n) = \int_{-\infty}^{a_n} \cdots \int_{-\infty}^{a_1} f(x_1, x_2, \ldots, x_n)\, dx_1 \ldots dx_n \tag{5.10}$$

Las variables aleatorias X y Y son continuas conjuntamente si existe una función $f(x, y) \geq 0$ tal que

$$F(a,b)=\int_{-\infty}^{b}\left[\int_{-\infty}^{a} f(x,y)\ dx\right]dy \tag{5.11}$$

La función $f(x,y)$ se llama la función de densidad conjunta de X y Y. Además

$$f(a,b)=\frac{\partial^2}{\partial a\partial b}F(a,b) \tag{5.12}$$

Teorema 5.2. Si X y Y son continuas conjuntamente entonces,

$$f_X(x)=\int_{-\infty}^{\infty} f(x,y)\ dy\ \text{ y } f_Y(y)=\int_{-\infty}^{\infty} f(x,y)\ dx.$$

Demostración.

$$\begin{aligned}f_X(x)&=\frac{d}{dx}F_X(x)=\frac{d}{dx}F(x,\infty)=\frac{d}{dx}\left(\int_{\infty}^{x}\int_{-\infty}^{\infty} f(u,y)\ dydu\right)\\&=\frac{d}{dx}\int_{-\infty}^{x}\left(\int_{-\infty}^{\infty} f(u,y)\ dy\right)du=\int_{-\infty}^{\infty} f(x,y)\ dy.\end{aligned}$$

Similarmente se muestra que $f_Y(y)=\int_{-\infty}^{\infty} f(x,y)\ dx$.

Ejemplo 5.2. Sea $f(x,y)=e^{-(x+y)}$, $0<x<\infty$, $0<y<\infty$, la función de densidad conjunta de X y Y. Hallar $P(X>1,Y<2), P(X>1)$ y $P(X>Y)$.

Solución. En este y otros ejemplos se calculan integrales impropias de la forma $\int_{a}^{\infty} f(x)dx$, las cuales se resuelven aplicando limite de la siguiente manera

$$\int_{a}^{\infty} f(x)dx=\lim_{b\to\infty}\int_{a}^{b} f(x)dx$$

$$
\begin{aligned}
P(X>1, Y<2) &= \int_0^2 \int_1^\infty e^{-(x+y)}\ dxdy = \int_0^2 \Big(\int_1^\infty e^{-x} e^{-y}\ dx \Big) dy \\
&= \int_0^2 e^{-y} \Big[-e^{-x} \Big|_1^\infty \Big] dy = e^{-1} \int_0^2 e^{-y}\ dy \\
&= e^{-1} \Big[-e^{-y} \Big|_0^2 \Big] = e^{-1} [1 - e^{-2}] = 0,3181.
\end{aligned}
$$

Para hallar $P(X > 1)$ se encuentra primero $f_X(x)$.

$$
f_X(x) = \int_{-\infty}^{\infty} f(x,y)\ dy = \int_0^\infty e^{-x} e^{-y}\ dy = e^{-x} \Big[-e^{-y} \Big|_0^\infty \Big] = e^{-x}
$$

Por lo tanto,

$$
\begin{aligned}
P(X>1) &= \int_1^\infty f_X(x)\ dx = \int_1^\infty e^{-x} dx = -e^{-x} \Big|_1^\infty \\
&= e^{-1} = 0,3679
\end{aligned}
$$

y

$$
\begin{aligned}
P(X>Y) &= \iint_{x>y} f(x,y) dxdy = \int_0^\infty \int_0^x e^{-x} e^{-y}\ dxdy \\
&= \int_0^\infty e^{-x} \Big(\int_0^x e^{-y} dy \Big) dx = \int_0^\infty e^{-x} [1 - e^{-x}]\ dx \\
&= -e^{-x} \Big|_0^\infty + \frac{e^{-2x}}{2} \Big|_0^\infty = \frac{1}{2}.
\end{aligned}
$$

5.2. Distribuciones condicionales e independencia estadística

Definición 5.5. Sean X y Y variables aleatorias discretas con función de masa de probabilidad conjunta $p(x,y)$, la función de masa de probabilidad condicional de X dado $Y = y$, denotada por $p_{X|Y}(x|y)$, se define por

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)},\ p_Y(y) > 0. \tag{5.13}$$

Similarmente, se define la función de masa de probabilidad condicional de Y dado $X = x$ por

$$p_{Y|X}(y|x) = \frac{p(x,y)}{p_X(x)},\ p_X(x) > 0. \tag{5.14}$$

Definición 5.6. Sean X y Y variables aleatorias discretas con función de masa de probabilidad conjunta $p(x, y)$. La función de distribución condicional de X dado $Y = y$, denotada $F_{X|Y}(\cdot|y)$, se define por

$$F_{X|Y}(x|y) = P\Big(X \leq x|Y = y\Big) = \sum_{a \leq x} p_{X|Y}(a|y). \tag{5.15}$$

Análogamente,

$$F_{Y|X}(y|x) = \sum_{b \leq y} p_{Y|X}(b|x). \tag{5.16}$$

Ejemplo 5.3. La función de masa de probabilidad conjunta de X y Y está dada por $p(1, 1) = 1/8,\ p(1, 2) = 1/4,\ p(2, 1) = 1/8$ y $p(2, 2) = 1/2$. Calcular la función de masa de probabilidad condicional de X dado $y = i$, $i = 1, 2$.

Solución. Se debe encontrar $p_{X|Y}(x|1)$ y $p_{X|Y}(x|2)$. Para esto se halla primero $p_Y(1)$ y $p_Y(2)$.

$$p_Y(1) = \sum_{x=1}^{2} p(x,1) = \frac{1}{4},\ \text{y}\ p_Y(2) = \sum_{x=1}^{2} p(x,2) = \frac{3}{4}.$$

Luego,

$$p_{X|Y}(x|1) = \frac{p(x,1)}{p_Y(1)} = 4p(x,1) \quad \text{y} \quad p_{X|Y}(x|2) = \frac{p(x,2)}{p_Y(2)} = \frac{4}{3}p(x,2).$$

En particular,

$$p_{X|Y}(1|1) = 4p(1,1) = \frac{1}{2} \quad \text{y} \quad p_{X|Y}(2|1) = 4p(2,1) = \frac{1}{2}.$$

$$p_{X|Y}(1|2) = \frac{4}{3}p(1,2) = \frac{1}{3} \quad \text{y} \quad p_{X|Y}(2|2) = \frac{4}{3}p(2,2) = \frac{2}{3}.$$

Definición 5.7. Sean X y Y variables aleatorias continuas distribuidas conjuntamente con función de densidad conjunta $f(x,y)$, la función de distribución condicional de X dado $Y = y$, denotada por $F_{X|Y}(\cdot|y)$, se define por

$$F_{X|Y}(a|y) = \int_{-\infty}^{a} f_{X|Y}(x|y)\ dx. \tag{5.17}$$

Análogamente,

$$F_{Y|X}(b|x) = \int_{-\infty}^{b} f_{Y|X}(y|x)\ dy. \tag{5.18}$$

Ejemplo 5.4. La función de densidad conjunta de X y Y está dada por $f(x,y) = xe^{-x(y+1)}$, para $x > 0$ e $y > 0$. Encontrar

a) La función de densidad condicional de X dado $Y = y$.

b) $P(X < 1|Y = y)$.

Solución. Se utilizarán las ecuaciones $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$ y 5.17 para resolver el problema.

$$f_Y(y) = \int_0^{\infty} f(x,y)\ dx = \int_0^{\infty} xe^{-x(y+1)}\ dx = \frac{1}{y+1}\int_0^{\infty} e^{-x(y+1)}\ dx = \frac{1}{(y+1)^2}.$$

Luego,

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = (y+1)^2 xe^{-x(y+1)},\ \text{lo que responde a).}$$

Con base en este resultado,

$$\begin{aligned} P(X < 1|Y = y) &= \int_0^1 f_{X|Y}(x|y)\ dx = \int_0^1 (y+1)^2 xe^{-x(y+1)}\ dx \\ &= (y+1)^2\left(-\frac{1}{(y+1)e^{y+1}} - \frac{1}{(y+1)^2 e^{y+1}} + \frac{1}{(y+1)^2}\right) \\ &= 1 - \frac{y+2}{e^{y+1}},\ \text{lo que responde b).} \end{aligned}$$

Definición 5.8. Sean X y Y variables aleatorias distribuidas conjuntamente. Se dice que X y Y son independientes si para A y B conjuntos de números reales se tiene que,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B). \tag{5.19}$$

De esta definición se deduce que X y Y son independientes si

$$F(a, b) = F_X(a)F_Y(b). \tag{5.20}$$

En general, $X_1, X_2, \ldots, X_n$ son independientes si

$$F(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n). \tag{5.21}$$

La definición de independencia tiene su equivalencia en términos de la función de masa de probabilidad conjunta en el caso discreto, y de la función de densidad de probabilidad conjunta en el caso continuo.

Definición 5.9. Sean X y Y variables aleatorias distribuidas conjuntamente. Se dice que X y Y son independientes si y solo si:

$$p(x, y) = p_X(x)p_Y(y), \text{para todo } (x, y) \quad \text{[caso discreto]} \tag{5.22}$$

$$f(x, y) = f_X(x)f_Y(y), \text{para todo } (x, y) \quad \text{[caso continuo].} \tag{5.23}$$

En cada caso, si para algún (x, y) las igualdades 5.22 y 5.23 no se cumplen, entonces se dice que X y Y son dependientes.

Ejemplo 5.5. ¿Son las variables del ejemplo 5.3 independientes?

Solución. Con los datos del ejemplo se construye la siguiente distribución:

$Y \setminus X$	1	2	$p_Y(y)$
1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
2	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$
$p_X(x)$	$\frac{3}{8}$	$\frac{5}{8}$	1

Tabla 5.2: Distribución conjunta de las variables X y Y del ejemplo 5.3

Se puede observar que $p(1,1) = 1/8$, y $p_X(1)p_Y(1) = (3/8)(1/4)$, por tanto, $p(1,1) \neq p_X(1)p_Y(1)$, es decir, X y Y no son independientes.

Ejemplo 5.6. La función de densidad conjunta de X y Y está dada por $f(x,y) = xe^{-(x+y)}$, $x > 0$ y $y > 0$. ¿Son X y Y independientes?

Solución. Para verificar si X y Y son independientes se deben encontrar las densidades marginales de X y Y.

$$f_X(x) = \int_0^\infty xe^{-(x+y)}\ dy = xe^{-x}\int_0^\infty e^{-y}\ dy = xe^{-x},\ x > 0.$$
$$f_Y(y) = \int_0^\infty xe^{-(x+y)}\ dx = e^{-y}\int_0^\infty xe^{-x}\ dx = e^{-y},\ y > 0.$$

Como $f_X(x)f_Y(y) = xe^{-(x+y)} = f(x,y)$ para todo $x > 0$ e $y > 0$, entonces X y Y son independientes.

El siguiente ejemplo muestra cómo resolver determinados tipos de problemas dada la condición de independencia.

Ejemplo 5.7. Calcular la función de masa de probabilidad condicional de X dado que $X + Y = n$ cuando X y Y son variables aleatorias independientes e idénticamente distribuidas como variables geométricas.

Solución. Si $X \sim G(p)$ y $Y \sim G(p)$, entonces $P(X = k) = p(1-p)^{k-1}$ y $P(Y = t) = p(1-p)^{t-1}$. Se debe encontrar $P(X|X + Y = n)$.

Para esto se debe hallar primero $P(X + Y = n)$, puesto que

$$P(X|X+Y=n) = \frac{P(X=k, X+Y=n)}{P(X+Y=n)}.$$

El evento $\{X + Y = n\}$ se puede escribir como la unión de los eventos disjuntos $\{X = k, Y = n-k\},\ k = 1, 2, \ldots, n-1$, es decir,

$$\{X+Y=n\} \equiv \bigcup_{k=1}^{n-1}\{X=k, Y=n-k\}.$$

Luego,

$$P(X+Y=n)=\sum_{k=1}^{n-1} P(X=k, Y=n-k)=\sum_{k=1}^{n-1} P(X=k)P(Y=n-k)$$

por ser X y Y independientes. Es decir,

$$\begin{aligned} P(X+Y=n) &= \sum_{k=1}^{n-1} p(1-p)^{k-1}p(1-p)^{n-k-1} = p^2\sum_{k=1}^{n-1}(1-p)^{n-2} \\ &= (n-1)p^2(1-p)^{n-2}. \end{aligned}$$

Por tanto,

$$P(X|X+Y=n)=\frac{P(X=k)P(Y=n-k)}{P(X+Y=n)}$$

$$=\frac{p(1-p)^{k-1}p(1-p)^{n-k-1}}{(n-1)p^2(1-p)^{n-2}}=\frac{1}{n-1},$$

$n=2,3,\ldots$

Se observa en este ejemplo que la distribución condicional de X dado $X+Y=n$ es independiente del parámetro p.

5.3. Esperanzas

En esta sección se estudian algunas propiedades del valor esperado de funciones de variables aleatorias distribuidas conjuntamente.

Definición 5.10. Sean $X_1, X_2, \ldots, X_k$ variables aleatorias distribuidas conjuntamente y sea g una función de las k variables. El valor esperado de g se define por

$$E[g(X_1, X_2, \ldots, X_k)] = \begin{cases} \sum\limits_{x_k} \sum\limits_{x_{k-1}} \cdots \sum\limits_{x_1} g(x_1, \ldots, x_k) p_{X_1,\ldots,X_k}(x_1, \ldots, x_k) \\ \text{si las variables son discretas.} \\ \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} g(x_1, \ldots, x_k) f(x_1, \ldots, x_k) dx_1 \cdots dx_k \\ \text{si las variables son continuas.} \end{cases} \tag{5.24}$$

Los teoremas que siguen se demostraran para el caso continuo, el caso discreto debe considerarse como ejercicio, ya que las pruebas son similares.

Teorema 5.3. Si $g(x_1, \ldots, x_k) = x_i$, entonces $E[g(X_1, X_2, \ldots, X_k)] = E(X_i) = \mu_{X_i}$.

Demostración. Sean $X_1, X_2, \ldots, X_k$ variables aleatorias continuas con función de densidad conjunta $f(x_1, \ldots, x_k)$ entonces,

$$\begin{aligned} E[g(X_1, X_2, \ldots, X_k)] &= \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} x_i f(x_1, \ldots, x_k)\ dx_1 \cdots dx_k \\ &= \int\limits_{-\infty}^{\infty} x_i f_{X_i}(x_i) dx_i = E(X_i) = \mu_{X_i}. \end{aligned}$$

Se usó el hecho de que

$$f_{X_i}(x_i) = \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} f_{X_1,\ldots,X_k}(x_1, \ldots, x_k)\ dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_k.$$

Teorema 5.4. Si $g(x_1, \ldots, x_k) = [x_i - E(X_i)]^2$ entonces, $E[g(X_1, X_2, \ldots, X_k)] = V(X_i)$.

Demostración.

$$\begin{aligned} E[g(X_1, X_2, \ldots, X_k)] &= \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} [x_i - E(X_i)]^2 f(x_1, \ldots, x_k)\ dx_1 \cdots dx_k \\ &= \int\limits_{-\infty}^{\infty} [x_i - E(X_i)]^2 f_{X_i}(x_i) dx_i = V(X_i). \end{aligned}$$

Ejemplo 5.8. Si X y Y tienen función de densidad conjunta dada por $f(x, y) = 2$, para $0 < x < y$, $0 < y < 1$. Hallar $E(XY)$.

Solución.

$$E(XY) = \int_0^1 \int_0^y 2xy \, dxdy = \int_0^1 yx^2 \Big|_0^y \, dy = \int_0^1 y^3 \, dy = \frac{1}{4}.$$

Teorema 5.5. Si X y Y son variables aleatorias independientes, entonces para dos funciones g y h cualesquiera,

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)].$$

Demostración. Suponiendo que X y Y son continuas con función de densidad conjunta $f_{X,Y}(x, y)$ entonces,

$$\begin{aligned} E[g(X)h(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_{XY}(x, y) \, dxdy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y) \, dxdy \end{aligned}$$

por ser X y Y independientes. Luego,

$$\begin{aligned} E[g(X)h(Y)] &= \left[\int_{-\infty}^{\infty} g(x)f_X(x) \, dx \right] \left[\int_{-\infty}^{\infty} h(y)f_Y(y) \, dy \right] \\ &= E[g(X)]E[h(Y)]. \end{aligned}$$

Definición 5.11. Sean X y Y variables aleatorias definidas en el mismo espacio de probabilidad. La covarianza de X y Y, denotada $Cov(X, Y)$ ó σ_{XY} se define por

$$Cov(X, Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]. \tag{5.25}$$

Teorema 5.6. Si X y Y son variables aleatorias definidas en el mismo espacio de probabilidad, $Cov(X, Y) = E(XY) - E(X)E(Y)$. Además, si X y Y son independientes, $Cov(X, Y) = 0$.

Demostración.

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y)$$

$$= E(XY) - E(X)\mu_Y - E(Y)\mu_X + \mu_X\mu_Y = E(XY) - E(X)E(Y).$$

Además, si X y Y son independientes, $E(XY) = E(X)E(Y)$ y por tanto $Cov(X,Y) = 0$. En este caso se dice que las variables X y Y no están correlacionadas.

El recíproco de la última parte de este teorema no es cierto, es decir, si $Cov(X,Y) = 0$, en general no es cierto que X y Y sean independientes. El siguiente ejemplo ilustra esa situación.

Ejemplo 5.9. Sean X y Y variables aleatorias con distribución conjunta dada en la tabla 5.3. Mostrar que $Cov(X,Y) = 0$ y sin embargo X y Y son dependientes.

$Y \setminus X$	-1	0	1	$p_Y(y)$
0	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{2}{3}$
1	0	$\frac{1}{3}$	0	$\frac{1}{3}$
$p_X(x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

Tabla 5.3: Distribución conjunta de las variables X y Y del ejemplo 5.9

Solución. De la distribución de probabilidades de la tabla 5.3, $\sigma_{XY} = 0$ por que $E(X) = (-1)(1/3) + (0)(1/3) + (1)(1/3) = 0$, de manera que $E(X)E(Y) = 0$, y además porque

$$\begin{aligned} E(XY) &= \sum_{y=0}^{1} \sum_{x=-1}^{1} xyp(x,y) \\ &= \sum_{y=0}^{1} [-yp(-1,y) + yp(1,y)] = -p(-1,1) + p(1,1) = 0 \end{aligned}$$

Pero $p(-1,0) = 1/3 \neq p_X(-1)p_Y(0) = (1/3)(2/3) = 2/9$. Luego, $\sigma_{XY} = 0$ y sin embargo X y Y son dependientes.

Definición 5.12. El coeficiente de correlación de las variables aleatorias X y Y, denotado por ρ_{XY} se define por

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}, \ \sigma_X > 0, \ \sigma_Y > 0. \tag{5.26}$$

La covarianza y el coeficiente de correlación de las variables aleatorias X y Y son medidas de una relación lineal de X y Y. La primera tiende a medir la relación lineal de X y Y, pero su magnitud real no tiene mucha importancia puesto que esta depende de la variabilidad de X y Y. El coeficiente de correlación remueve esta variabilidad y es por eso una mejor medida de la relación lineal entre X y Y.

Una expresión útil es el resultado del siguiente teorema, donde se usa la definición de covarianza y la de varianza.

Teorema 5.7. Si X y Y son variables aleatorias, entonces

$$V(X \pm Y) = V(X) + V(Y) \pm 2Cov(X,Y).$$

Demostración. La prueba se hará solamente para la suma.

$$\begin{aligned} V(X+Y) &= E\{[X+Y-E(X+Y)]^2\} = E\{[(X-E(X)) + (Y-E(Y))]^2\} \\ &= E\{(X-E(X))^2 + (Y-E(Y))^2 + 2[X-E(X)][Y-E(Y)]\} \\ &= E[(X-E(X))^2] + E[(Y-E(Y))^2] + 2E[(X-E(X))(Y-E(Y))] \\ &= V(X) + V(Y) + 2Cov(X,Y). \end{aligned}$$

En general,

$$V(X_1 + X_2 + \cdots + X_n) = \Sigma_{i=1}^n V(X_i) + 2\Sigma\Sigma_{i<j} Cov(X_i, X_j).$$

Además, si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes, entonces

$$V(\Sigma_{i=1}^n X_i) = \Sigma_{i=1}^n V(X_i).$$

Teorema 5.8. Si ρ_{XY} es el coeficiente de correlación de X y Y entonces,

$$-1 \leq \rho_{XY} \leq 1.$$

Demostración. Sean σ_X^2 y σ_Y^2 las varianzas de X y Y respectivamente entonces,

$$\begin{aligned} 0 \leq V\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) &= \frac{V(X)}{\sigma_X^2} + \frac{V(Y)}{\sigma_Y^2} + 2Cov\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) \\ &= 2 + 2\left(\frac{\sigma_{XY}}{\sigma_X \sigma_Y}\right) \\ &0 \leq 2(1 + \rho_{XY}). \end{aligned}$$

Luego, $\rho_{XY} \geq -1$. (i)

Además,

$$\begin{aligned} 0 \leq V\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) &= \frac{V(X)}{\sigma_X^2} + \frac{V(Y)}{\sigma_Y^2} - 2Cov\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) \\ &= 2 - 2\rho_{XY} \\ &0 \leq 2(1 - \rho_{XY}), \end{aligned}$$

es decir que $\rho_{XY} \leq 1$. (ii)

De (i) y (ii), se tiene que

$$-1 \leq \rho_{XY} \leq 1.$$

En la demostración de este teorema se emplearon las propiedades que se invitan a probar en el ejercicio 9 de este capítulo.

Se puede mostrar que cuando $\rho = \pm 1$, hay una perfecta relación lineal entre X y Y, y cuado $\rho = 0$ o se aproxima a cero, la relación no es lineal.

Así como se definió el condicionamiento de variables aleatorias, es natural definir el valor esperado condicional de una variable aleatoria dado el valor de otra.

Definición 5.13. Sean X y Y variables aleatorias distribuidas conjuntamente y sea $g(x, y)$ una función de las variables aleatorias X y Y. La esperanza condicional de $g(X, Y)$ dado $Y = y$, se define por

$$E[g(X,Y) \mid Y = y] = \begin{cases} \sum\limits_{x_j} g(x_j, y) p_{X|Y}(x_j|y), & \text{si } X \text{ y } Y \text{ son discretas} \\ \\ \int\limits_{-\infty}^{\infty} g(x,y) f_{X|Y}(x|y)\, dx, & \text{si } X \text{ y } Y \text{ son continuas.} \end{cases}$$

En particular, si $g(x, y) = x$, se define $E[X|Y = y]$ por

$$E[X|Y=y] = \begin{cases} \sum\limits_x x p_{X|Y}(x|y), & \text{si } X \text{ y } Y \text{ son discretas.} \\ \\ \int\limits_{-\infty}^{\infty} x f_{X|Y}(x|y)\ dx, & \text{si } X \text{ y } Y \text{ son continuas.} \end{cases}$$

Notación

$E(X|Y = y)$ se denota por $E(X|Y)$ y $E[g(X,Y)|Y = y]$ por $E[g(X,Y)|Y]$

Puede probarse además que, $E(X|Y)$ y $E[g(X,Y)|Y]$ son funciones de y.

Ejemplo 5.10. Dada la distribución conjunta de X y Y del ejemplo 5.9, encontrar $E[X|Y = 0]$ y $E[Y|X = 0]$.

Solución. En el ejemplo 5.9 se tiene que

$$p_{X|Y}(x|0) = \begin{cases} \frac{1}{2}, & x = -1 \\ 0, & x = 0 \\ \frac{1}{2}, & x = 1 \end{cases} \quad ; \quad p_{Y|X}(y|0) = \begin{cases} 0, & y = 0 \\ 1, & y = 1 \end{cases}$$

Luego,

$$E(X|Y = 0) = \sum_x x p_{XY}(x|0) = (-1)\left(\frac{1}{2}\right) + 0(0) + 1\left(\frac{1}{2}\right) = 0$$

y

$$E(Y|X = 0) = \sum_y y p_{Y|X}(y|0) = 0(0) + 1(1) = 1.$$

Ejemplo 5.11. Dada la función de densidad conjunta de X y Y del ejemplo 5.8, hallar $E[XY|Y = y]$

Solución.

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{2}{\int\limits_0^y 2\ dx} = \frac{1}{y},\ 0 < y < 1.$$

Luego,

$$E(XY|Y = y) = \int\limits_0^y xy\left(\frac{1}{y}\right)\ dx = \frac{y^2}{2},\ 0 < y < 1.$$

Anteriormente se anotó que $E[g(X)|Y]$ es una función de y. Sea $h(y) = E[g(X)|Y = y]$, se mostrará en el teorema que sigue que $E[h(Y)] = E\{E[g(X)|Y]\} = E[g(X)]$.

Teorema 5.9. Si X y Y son variables aleatorias distribuidas conjuntamente, entonces

$$E[\{E[g(X)|y]\}] = E[g(X)].$$

Demostración. Suponga que X y Y son continuas y que $h(y) = E[g(X)|y]$, entonces,

$$\begin{aligned}
E\{E[g(X)|y]\} = E[h(Y)] &= \int_{-\infty}^{\infty} h(y) f_Y(y)\ dy = \int_{-\infty}^{\infty} E[g(X)|y] f_Y(y)\ dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y) f_Y\ dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x) f(x,y)\ dxdy = \int_{-\infty}^{\infty} g(x)\left(\int_{-\infty}^{\infty} f(x,y)\ dy\right)dx \\
&= \int_{-\infty}^{\infty} g(x) f_X(x)\ dx = E[g(X)].
\end{aligned}$$

Como consecuencia del teorema se tiene que, $E[E(X|Y)] = E(X)$.

Definición 5.14. Sean X y Y variables aleatorias distribuidas conjuntamente, la varianza condicional de X dado $Y = y$, notada $V[X|Y]$, se define por

$$V(X|Y) = E[\{X - E(X|Y)\}^2|Y].$$

De esta manera, la fórmula análoga a $V(X) = E(X^2) - [E(X)]^2$ está dada por

$$V(X|Y = y) = E(X^2|Y = y) - [E(X|Y = y)]^2.$$

Teorema 5.10. Si X y Y son variables aleatorias distribuidas conjuntamente, entonces

$$V(X) = E[V(X|Y)] + V[E(X|Y)]$$

Demostración.

$$E[V(X|Y)] = E[E(X^2|Y)] - E[\{E(X|Y)\}^2] = E(X^2) - E[\{E(X|Y)\}^2]$$

y

$$V[E(X|Y)] = E[\{E(X|Y)\}^2] - \{E[E(X|Y)]\}^2 = E[\{E(X|Y)\}^2] - [E(X)]^2$$

luego,

$$E[V(X|Y)] + V[E(X|Y)] = E(X^2) - [E(X)]^2 = V(X).$$

Para finalizar este capítulo se define la función generatriz de momentos conjunta, la cual es de gran utilidad para llegar a algunos resultados interesantes que aparecen posteriormente.

Definición 5.15. La función generatriz de momentos conjunta de las variables aleatorias $X_1, X_2, \ldots, X_n$, denotada por $M_{X_1,\ldots,X_k}(t_1, \ldots, t_k)$, se define por

$$M_{X_1,\ldots,X_k}(t_1, \ldots, t_k) = E\left[e^{\sum_{i=1}^{k} t_i X_i}\right]. \tag{5.27}$$

Si la esperanza existe para todos los valores de $t_1, t_2, \ldots, t_k$ y para algún $h > 0$, tal que $-h < t_i < h$, $i = 1, 2, \ldots, k$.

Teorema 5.11. Las funciones generatrices de momentos marginales se obtienen de la función generatriz conjunta.

Demostración. La prueba se hace para $k = 2$. Si X y Y son variables aleatorias conjuntas con función generatriz de momentos conjunta $M_{X,Y}(t_1, t_2)$. Entonces

$$M_X(t_1) = E[e^{t_1 X}] = E[e^{t_1 X + 0Y}]$$

$$= M_{X,Y}(t_1, 0) = \lim_{t_2 \to 0} M_{X,Y}(t_1, t_2).$$

Análogamente,

$$M_Y(t_2) = \lim_{t_1 \to 0} M_{X,Y}(t_1, t_2).$$

Ejemplo 5.12. Si la función generatriz de momentos conjunta de X y Y es $M_{X,Y}(t_1, t_2) = e^{\frac{1}{2}(t_1^2 + t_2^2)}$, ¿cuál es la distribución de Y?

Solución. Por el teorema anterior

$$M_Y(t_2) = \lim_{t_1 \to 0} M_{X,Y}(t_1, t_2).$$

Luego,

$$M_Y(t_2) = \lim_{t_1 \to 0} e^{\frac{1}{2}(t_1^2+t_2^2)} = e^{\frac{1}{2}t_2^2}.$$

Pero esta es la función generatriz de momentos de una variable aleatoria con distribución normal $(0,1)$, es decir, $Y \sim N(0,1)$

Teorema 5.12. Las variables aleatorias X y Y, distribuidas conjuntamente, son independientes si y solo si $M_{X,Y}(t_1,t_2) = M_X(t_1)M_Y(t_2)$.

Demostración. ($\Longrightarrow$)Si X y Y son independientes, se tiene que

$$\begin{aligned} M_{X,Y}(t_1,t_2) &= E[e^{t_1X+t_2Y}] = E[e^{t_1X}e^{t_2Y}] \\ &= E[e^{t_1X}]E[e^{t_2Y}] \end{aligned}$$

Es decir,

$$M_{X,Y}(t_1,t_2) = M_X(t_1)M_Y(t_2).$$

($\Longleftarrow$)(Ejercicio)

Observación
Los teoremas 5.11 y 5.12 se generalizan para k variables aleatorias, con $k > 2$.

5.4. Ejercicios

1. Se lanzan dos dados legales. Encontrar la función de masa de probabilidad conjunta de X y Y cuando X es el valor del primer dado y Y es el mayor valor de los dos números obtenidos.

2. Sean X y Y variables aleatorias distribuidas conjuntamente con función de probabilidad conjunta dada por $f(x,y) = 2/k^2$, $0 < x < y$, $0 < y < 1$.

   a) Determinar k.

   b) Determinar las densidades marginales de X y de Y.

   c) Encontrar la densidad condicional de X dado $Y = y$, y el valor esperado condicional de Y dado $X = x$.

3. Si X y Y son variables aleatorias binomiales independientes, ambas con parámetros n y p, demostrar que la distribución condicional de X dado que $X + Y = n$ es una distribución hipergeométrica. Posteriormente, encontrar $E[X|X + Y = n]$.

4. Se escoge un número X al azar entre los digitos 5, 6, 7 y 8. Después, también al azar se escoge un número Y entre los pares menores que X.

   a) Halle la función de masa de probabilidad conjunta de X y Y.

   b) Halle $p_{Y|X}(y|X = 7)$.

5. La función de densidad de probabilidad conjunta de X y Y está dada por

$$f(x, y) = \begin{cases} \left(\frac{6}{7}\right)\left(x^2 + \frac{xy}{2}\right), \ 0 < x < 1, \ 0 < y < 2. \\ 0, \ \text{en otro caso} \end{cases}$$

   a) Calcular la función de distribución de X.

   b) Hallar $P(X > Y)$ y $P[(Y > \frac{1}{2})|(X < \frac{1}{2})]$.

6. La función de densidad de probabilidad conjunta de X y Y está dada por

$$f(x, y) = \begin{cases} 4x^3(1 - x), 0 < x < 1, \ 3 < y < 8 \\ 0, \ \text{en otro caso} \end{cases}$$

   a) Obtenga $f_X(x)$ y $f_Y(y)$. ¿Son X y Y independientes?

   b) ¿Que distribución tienen estas variables aleatorias?

   c) Comparar $P(2 < Y < 4)$ con $P(2 < Y < 4|X < 1/2)$.

7. Un hombre y una mujer, se ponen de acuerdo para encontrarse en un lugar a las 2:00 p.m. Si el hombre llega en un tiempo distribuido uniformemente entre las 2:15 p.m. y las 2:45 p.m. y, si la mujer llega independientemente en un tiempo distribuido uniformemente entre las 1:30 m. y las 2:30 p.m., encontrar la probabilidad que el primero en llegar no espere más de 5 minutos. ¿Cuál es la probabilidad que el hombre llegue primero?

8. Las variables aleatorias X y Y se dice que tienen una distribución normal bivariada si su función de densidad conjunta está dada por

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \ e^{-\frac{1}{2}(1-\rho^2)(Fac)}, \ \text{donde}$$

$$Fac = \left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left[\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_X}\right]$$

a) Demostrar que la función de densidad condicional de X dado $Y = y$ es la densidad normal con parámetros

$$\mu_X - \left(\frac{\rho\sigma_X}{\sigma_Y}\right)(y - \sigma_Y), \text{ y } \sigma_X^2(1-\rho^2).$$

b) Demostrar que X y Y son ambas variables aleatorias normales con parámetros (μ_X, σ_X^2) y (μ_Y, σ_Y^2), respectivamente.

c) Demostrar que X y Y son independientes cuando $\rho = 0$.

9. Si las variables aleatorias X y Y tienen función de densidad conjunta

$$f(x,y) = \begin{cases} \frac{1}{8}(6-x-y), 0 < x < 2 < y < 4, \\ 0, \text{ en otro caso} \end{cases}$$

a) Hallar $V[Y|X = x]$.

b) Verificar que $E(Y) = E[E(Y|x)]$.

c) Encontrar $E[XY|X = x]$.

10. Lance tres monedas. Sea X el número de caras en las primeras dos lanzadas, y Y el número de caras en las dos últimas.

a) Encontrar la distribución conjunta de X y Y.

b) Hallar $E[Y|X = 1]$.

c) ¿Qué es ρ_{XY}?

11. Si X y Y son variables aleatorias distribuidas conjuntamente, demostrar que,

a) $Cov(a + bX, c + dY) = bdCov(X, Y)$.

b) $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$.

c) Si $Y = a + bX$ entonces, $\rho_{XY} = 1$ si $b > 0$ y $\rho_{XY} = -1$ si $b < 0$.

12. Si $X \sim Be(5, 3)$ y Y es otra variable aleatoria con función de densidad

$$f_Y(y) = \begin{cases} \frac{4}{3}(1-y^3), 0 < y < 1 \\ 0, \quad \text{en otro caso} \end{cases}$$

Obtenga la $Cov(3X + 5Y)$ si $Cov(X, Y) = -\frac{2}{75}$.

13. Si X y Y son variables aleatorias con función de densidad conjunta

$$f(x,y) = \begin{cases} kxy^2, 0 < x < 2, 0 < y < 1 \\ 0, \text{ en otro caso} \end{cases}$$

   a) Obtenga k y determine si X y Y son independientes.
   b) ¿Que distribución tiene cada una de ellas?
   c) Encuentre $V(2X + 3Y)$.

14. La función de densidad conjunta de X y Y esta dada por

$$f(x,y) = \begin{cases} kye^{-x}, \; x > 0, \; 0 < y < 3 \\ 0, \quad \text{en otro caso} \end{cases}$$

   Halle k, f_X, f_Y y $cov(X,Y)$.

15. Sean X y Y dos variables aleatorias con función generatriz de momentos dada por

$$M_{X,Y}(t_1,t_2) = \frac{1}{3}\left[e^{(t_1+t_2)} + \frac{1}{6}(e^{t_1} + e^{t_2})\right]^2$$

   Calcular $E(X)$ y $V(X)$.

16. Considere una muestra aleatoria de tamaño 2 tomada sin reemplazamiento de una urna que contiene 3 bolas numeradas por $1,2,3$. Sea X el número de la primera bola seleccionada y Y el número mayor de las bolas seleccionadas,

   a) Encontrar la distribución conjunta de X y Y.
   b) Hallar $P[X = 1|Y = 3]$.
   c) Encontrar la $Cov(X,Y)$.

17. Si $X \sim U(0,1)$ y la distribución condicional de Y dado $X = x$ tiene una distribución binomial con parámetros n y $p = x$, para $y = 0,1,\cdots,n$. Encuentre $E(Y)$ y la distribución de Y.

18. Suponga que se desarrollan pruebas independientes con probabilidad de *éxito* p y sean X_i el número de pruebas para que i éxitos ocurran, para $i = 1,2,\cdots,r$. Si $W_i = X_i - X_{i-1}$, $i = 1,2,\cdots,r$ y donde X_0 se define como 0, pruebe que $E(Y) = \frac{r}{p}$ y $V(Y) = \frac{r(1-p)}{p^2}$, para $Y = \sum_{i=1}^{r} W_i$.

19. Demuestre el teorema 5.1

20. Demuestre el recíproco del teorema 5.12.

Capítulo 6

Funciones de variables aleatorias

6.1. Introducción

En este capítulo se darán algunas técnicas para encontrar distribuciones de funciones de variables aleatorias. Aunque en el capítulo anterior se discutieron tópicos sobre el valor esperado de funciones de variables aleatorias, no se dijo nada sobre la función de densidad como tampoco de la función de distribución de estas nuevas variables aleatorias.

Existen varias técnicas para encontrar la distribución de funciones de variables aleatorias. A continuación se presentan tres: técnica de la función de distribución, tećnica de la función generatriz de momentos y la técnica de las transformaciones. Además, como consecuencia de ellas, se demuestran resultados muy importantes como *la ley de los grandes números* y el *teorema central del límite*.

De los capítulos anteriores se pueden inferir los siguientes resultados:

a) Si $X_1, X_2, \ldots, X_n$, son variables aleatorias, entonces

$$E(\Sigma_{i=1}^{n} X_i) = \Sigma_{i=1}^{n} E(X_i), \text{ y}$$
$$V(\Sigma_{i=1}^{n} X_i) = \Sigma_{i=1}^{n} V(X_i) + 2\Sigma\Sigma_{i<j} Cov(X_i, X_j). \tag{6.1}$$

Además, si $X_1, X_2, \ldots, X_n$ son variables aleatorias no correlacionadas entonces,

$$V(\Sigma_{i=1}^{n} X_i) = \Sigma_{i=1}^{n} V(X_i). \tag{6.2}$$

b) Sean $X_1, X_2, \ldots, X_n$ y $Y_1, Y_2, \ldots, Y_m$ dos conjuntos de variables aleatorias, y $a_1, a_2, \ldots, a_n$ y $b_1, b_2, \ldots, b_m$ dos conjuntos de constantes, entonces del ejercicio 9, del capítulo anterior, se tiene que

$$Cov(\Sigma_{i=1}^{n} a_i X_i, \Sigma_{j=1}^{m} b_i Y_i) = \Sigma_{i=1}^{n} \Sigma_{j=1}^{m} a_i b_j Cov(X_i, Y_j). \tag{6.3}$$

c) Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes e idénticamente distribuidas con media μ_X y varianza σ_X^2 y si $\bar{X} = \frac{\sum X_i}{n}$, entonces

$$E(\bar{X}) = \mu_X \quad \text{y} \quad V(\bar{X}) = \frac{\sigma_X^2}{n}. \tag{6.4}$$

6.2. Técnica de la función de distribución

Inicialmente se considera una variable aleatoria X y sea $g(.)$ una función real, así que $Y = g(X)$ es una variable aleatoria. El objetivo está en encontrar la distribución de Y, conocida la distrubución de X. El siguiente teorema muestra como encontrarla.

Teorema 6.1. Sea X una variable aleatoria continua y $g(.)$ una función de X. Entonces la función de distribución de $Y = g(X)$ se obtiene de la función de distribución de X.

Demostración. Si g es una función estrictamente creciente y $F_X(\cdot)$ es la función de distribución de X entonces,

$$F_Y(y) = P(Y \leq y) = P[g(X) \leq y] = P[(X \leq g^{-1}(y)] = F_X[g^{-1}(y)].$$

Si g es una función estrictamente decreciente entonces,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P[g(X) \leq y] = P[X \geq g^{-1}(y)] = 1 - P[X < g^{-1}(y)] \\ &= 1 - F_X[g^{-1}(y)]. \end{aligned}$$

Estas dos situaciones prueban el teorema.

Corolario. Si X es una variable aleatoria continua con función de densidad de probabilidad f_X, y si $g(.)$ es una función estrictamente monótona (creciente o decreciente) diferenciable, para $Y = g(X)$ entonces

$$f_Y(y) = f_X[g^{-1}(y)] \left| \frac{d}{dy}\, g^{-1}(y) \right|. \tag{6.5}$$

En el caso en que X es una variable aleatoria discreta que toma los valores x_i, $i = 1, 2, \ldots$ y si $Y = g(X)$, entonces Y toma valores $y_i = g(x_i)$, $i = 1, 2, \ldots$ y puede mostrarse que

$$\begin{aligned} p_Y(y) &= \sum_{x_i \,\in\, A} p_X(x_i), \quad \text{donde } p_X(x_i) = P(X = x_i) \text{ y} \\ A &= \{x_i : g(x_i) = y_i\}. \end{aligned} \tag{6.6}$$

Ejemplo 6.1. Sea X una variable aleatoria con distribución de Poisson con parámetro λ y sea $Y = X^2 - 5$. Hallar la función de masa de probabilidad de Y.

Solución. Obsérve que $\{y : y = x^2 - 5,\ x = 1, 2, \ldots\} = \{-5, -4, -1, 4, 11, 20 \ldots\}$ De $y = x^2 - 5$ se tiene que $x = \pm\sqrt{5 + y}$. Pero x sólo toma valores enteros positivos, así que $x = \sqrt{5 + y}$. Luego, $A = \{x : y = x^2 - 5\} = \{x : x = \sqrt{5 + y}\}$.
De manera que,

$$\begin{aligned} p_Y(y) = P(Y = y) &= \sum_{x \in A} p_X(x) = P(X = \sqrt{5 + y}) \\ &= \frac{e^{-\lambda}\lambda^{\sqrt{5+y}}}{(\sqrt{5 + y})!}, y = -5, -4, -1, 4, 11, ... \end{aligned}$$

Así para $y = 4$, $p_Y(4) = \dfrac{e^{-\lambda}\lambda^3}{3!} = p_X(3)$.

Ejemplo 6.2. Si X es una variable aleatoria continua con función de densidad f_X, y función de distribución F_X, hallar la función de distribución y de densidad de $Y = X^2$.

Solución.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

Luego,

$$f_Y(y) = \frac{1}{2\sqrt{y}}\ [f_X(\sqrt{y}) + f_X(-\sqrt{y})]$$

En particular, si X tiene una distribución normal $(0, 1)$ entonces

$$f_Y(y) = \frac{1}{2\sqrt{y}}\left(\frac{1}{\sqrt{2\pi}}\ e^{-y/2} + \frac{1}{\sqrt{2\pi}}\ e^{-y/2}\right) = \frac{y^{-1/2}e^{-y/2}}{\sqrt{2\pi}} = \frac{y^{-1/2}e^{-y/2}}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})},$$

que es la función de densidad de una variable aleatoria con distribución chi-cuadrado con un grado de libertad (ver 4.15).

Ejemplo 6.3. Encontrar la función de densidad de la variable aleatoria $Y = e^X$, si X es una variable aleatoria con distribución uniforme $(0, 1)$.

Solución. Sea $Y = e^X$, puesto que esta es una función creciente por el corolario del teorema 6.1, se tiene que:

$$f_Y(y) = f_X[g^{-1}(y)]\ \frac{d}{dx}\ g^{-1}(y) = f_X(\ln y)\frac{1}{y} = \frac{1}{y},\ 1 < y < e.$$

En particular,

$$E(Y) = \int_1^e y.\frac{1}{y}dy = e - 1.$$

Resultado que coincide con la solución del ejemplo 3.16.

Supóngase ahora que la distribución conjunta de las variables aleatorias $X_1, X_2, \ldots,$ X_n está dada, entonces teóricamente, la distribución conjunta de las variables aleatorias $Y_1, Y_2, ..., Y_k$ pueden ser determinadas, donde $Y_j = g_j(X_1, X_2, \ldots, X_n)$, $j = 1, 2, \ldots, k$.

En efecto, la función de distribución conjunta de $Y_1, Y_2, \ldots, Y_k$ es

$$F_{Y_1,Y_2,\ldots,Y_k}(y_1, y_2, \ldots, y_k) = P(Y_1 \leq y_1, \ldots, Y_k \leq y_k).$$

Ahora, para cada $(y_1, y_2, \ldots, y_k)$, el evento $\{Y_1 \leq y_1, \ldots, Y_k \leq y_k\}$ es equivalente al evento $\{g_1(X_1, \ldots, X_n) \leq y_1, \cdots, g_k(X_1, \ldots, X_n) \leq y_k\}$ y puesto que la distribución conjunta de $X_1, \ldots, X_n$ es conocida, posiblemente se puede obtener $P[g_1(X_1, \ldots, X_n) \leq y_1, \cdots, g_k(X_1, \ldots, X_n) \leq y_k]$ y consecuentemente determinar $F_{Y_1,Y_2,\ldots,Y_k}(y_1, y_2, \ldots, y_k)$.

A continuación, se encontrará la distribución de la suma, diferencia, producto y cociente de dos variables aleatorias, usando esta técnica; además se encontrará también la distribución de las variables mínimo y máximo de un conjunto de variables aleatorias.

Teorema 6.2. Distribución de la suma y diferencia. Sean X y Y variables aleatorias distribuidas conjuntamente, con función de densidad conjunta $f(x, y)$ y sean $Z = X + Y$ y $V = Y - X$ entonces,

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\ dx = \int_{-\infty}^{\infty} f_{X,Y}(z - y, y)\ dy, \tag{6.7}$$

$$f_V(v) = \int_{-\infty}^{\infty} f_{X,Y}(x, v + x)\ dx = \int_{-\infty}^{\infty} f_{X,Y}(y - v, y)\ dy. \tag{6.8}$$

Demostración. Se probará la segunda parte de f_Z y la primera de f_V. Las restantes se prueban de forma análoga y se dejan como ejercicio.

$$\begin{aligned} F_Z(z) = P(Z \leq z) = P(X + Y \leq z) &= \iint_{x+y\leq z} f_{X,Y}(x, y)\ dxdy \\ &= \int_{-\infty}^{\infty} \Big(\int_{-\infty}^{z-y} f_{X,Y}(x, y)\ dx \Big) dy. \end{aligned}$$

Sea $x = u - y$, entonces $dx = du$, y cuando $x \to -\infty$, $u \to -\infty$; además cuando $x = z - y$, $u = z$. Por tanto,

$$F_Z(z) = \int_{-\infty}^{\infty} \Big(\int_{-\infty}^{z} f_{X,Y}(u - y, y)\ du \Big) dy = \int_{-\infty}^{z} \Big(\int_{-\infty}^{\infty} f_{X,Y}(u - y, y)\ dy \Big) du.$$

De manera que,

$$f_Z(z) = \frac{d}{dz}\ F_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(z - y, y)\ dy.$$

Esto prueba la segunda parte de la ecuación 6.7. Ahora,

$$\begin{aligned} F_V(v) = P(V \leq v) = P(Y - X \leq v) &= \iint_{y-x\leq v} f_{X,Y}(x, y)\ dxdy \\ &= \int_{-\infty}^{\infty} \Big(\int_{-\infty}^{v+x} f_{X,Y}(x, y)\ dy \Big) dx. \end{aligned}$$

Sea $y = u + x$, entonces $dy = du$, y cuando $y \to -\infty$, $u \to -\infty$; igualmente cuando $y = v + x$, $u = v$. Luego,

$$F_V(v) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{v} f_{X,Y}(x, u + x)\ du \right) dx = \int_{-\infty}^{v} \left(\int_{-\infty}^{\infty} f_{X,Y}(x, u + x)\ dx \right) du.$$

Por tanto,

$$f_V(v) = \frac{d}{dv}\ F_V(v) = \int_{-\infty}^{\infty} f_{X,Y}(x, v + x)\ dx.$$

Corolario. Sean X y Y son variables aleatorias independientes y continuas. Si $Z = X + Y$, entonces

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(z - x) f_X(x)\ dx = \int_{-\infty}^{\infty} f_X(z - y) f_Y(y)\ dy. \tag{6.9}$$

En este caso se dice que f_Z es la *convolución* de f_X y f_Y.

Ejemplo 6.4. Si X y Y son variables aleatorias independientes, cada una de ellas con distribución uniforme $(0, 1)$, hallar la función de densidad de $X + Y$.

Solución. Sea $Z = X + Y$, puesto que $f_X(x) = I_{(0,1)}(x)$ y $f_Y(y) = I_{(0,1)}(y)$, entonces Z asume los valores en $(0, 2)$.

Del corolario del teorema 6.2, se tiene que

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(z - x) f_X(x)\ dx = \int_{0}^{1} I_{(0,1)}(z - x) I_{(0,1)}(x)\ dx.$$

Se encontrará una expresión equivalente a $I_{(0,1)}(z - x) I_{(0,1)}(x)$.

Puesto que $I_{(0,1)}(z-x) = 1$ si $x < z$ y $x > z-1$ entonces,

$$\begin{aligned}
I_{(0,1)}(z-x)I_{(0,1)}(x) &= I_{(0,z)}(x)I_{(z-1,1)}(x)I_{(0,1)}(x)\\
&= I_{(0,z)\cap(z-1,1)\cap(0,1)}(x). \text{ ver apéndice B}\\
&= \begin{cases} I_{(0,z)\cap(0,1)}(x), & \text{si } z < 1\\ I_{(z-1,1)\cap(0,1)}(x), & \text{si } z \geq 1 \end{cases}\\
&= I_{(0,z)}(x)I_{(0,1)}(z) + I_{(z-1,1)}(x)I_{[1,2)}(z).
\end{aligned}$$

Luego,

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} I_{(0,z)}(x)I_{(0,1)}(z)\ dx + \int_{-\infty}^{\infty} I_{(z-1,1)}(x)I_{[1,2)}(z)\ dx\\
&= I_{(0,1)}(z)\int_0^z dx + I_{[1,2)}(z)\int_{z-1}^1 dx = zI_{(0,1)}(z) + (2-z)I_{[1,2)}(z).\\
&= \begin{cases} z, & 0 < z < 1\\ 2-z, & 1 \leq z < 2. \end{cases}
\end{aligned}$$

Ejemplo 6.5. Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas, como distribuciones exponenciales de parámetro λ. Demostrar que $\sum_{i=1}^{n} X_i$ tiene una distribución gama con parámetros n y λ.

Solución. Se probará por inducción.
Para $n = 1$, X_1 tiene distribución exponecial con parámetro λ, que es una distribución gamma con parámetros 1 y λ.

Supóngase que para $n = k-1$ el resultado es el de una distribución gamma con parámetros $k-1$ y λ. Se probará que para $n = k$, $\sum_{i=1}^{k} X_i$ tiene una distribución gamma con parámetros k y λ, lo cual prueba el ejercicio.

$$F_{X_1,\ldots,X_k}(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{z-y} f_{X_1}(x_1)f_Y(y)\ dx_1dy,\ \text{donde } Y = (X_2,\ldots,X_k)$$

$$F_{X_1+\cdots+X_k}(z) = \int_{-\infty}^{\infty} F_{X_1}(z-y)f_Y(y)\ dy. \text{ Por convolución.}$$

Pero,

$$f_Y(y) = f_{X_2,\cdots,X_k}(x_2,\ldots,x_k) = \frac{\lambda e^{-\lambda y}(\lambda y)^{k-2}}{(k-2)!},$$

luego

$$\begin{aligned}
F_{\sum\limits_{i=1}^{k} X_i}(z) &= \int_0^z \{1 - e^{-\lambda(z-y)}\}\left(\frac{\lambda e^{-\lambda y}(\lambda y)^{k-2}}{(k-2)!}\right) dy \\
&= \int_0^z \left(\frac{\lambda e^{-\lambda y}(\lambda y)^{k-2}}{(k-2)!}\right) dy - \lambda e^{-\lambda z}\int_0^z \frac{(\lambda y)^{k-2}}{(k-2)!}\ dy \\
&= \int_0^z \left(\frac{\lambda e^{-\lambda y}(\lambda y)^{k-2}}{(k-2)!}\right) dy - \frac{e^{-\lambda z}(\lambda z)^{k-1}}{(k-1)!}.
\end{aligned}$$

Por tanto, si $u = e^{-\lambda y}$ y $dv = y^{k-1}$ entonces,

$$\begin{aligned}
F_{\sum\limits_{i=1}^{k} X_i}(z) &= \frac{e^{-\lambda z}(\lambda z)^{k-1}}{(k-1)!} + \int_0^z \frac{\lambda e^{-\lambda y}(\lambda y)^{k-1}}{(k-1)!}dy - \frac{e^{-\lambda z}(\lambda z)^{k-1}}{(k-1)!} \\
&= \int_0^z \frac{\lambda e^{-\lambda y}(\lambda y)^{k-1}}{(k-1)!}dy.
\end{aligned}$$

O sea que,

$$f_{\sum\limits_{i=1}^{k} X_i}(z) = \frac{\lambda e^{-\lambda z}(\lambda z)^{k-1}}{(k-1)!}$$

que es la función de densidad de una distribución gamma con parámetros k y λ.

Teorema 6.3. Distribución del producto y cociente. Sean X y Y variables aleatorias continuas con función de densidad conjunta $f(x,y)$, y sean $Z = XY$ y $V = \frac{X}{Y}$ entonces,

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|x|}\ f_{X,Y}(x, z/x)\ dx = \int_{-\infty}^{\infty} \frac{1}{|y|}\ f_{X,Y}(z/y, y)\ dy, \tag{6.10}$$

y

$$f_V(v) = \int_{-\infty}^{\infty} |y| f_{X,Y}(vy, y)\ dy. \tag{6.11}$$

Demostración. Se demostrará la segunda parte de la ecuación 6.10 y la ecuación 6.11.

$$F_Z(z) = P(Z \leq z) = P(XY \leq z) = \iint\limits_{xy \leq z} f_{X,Y}(x,y)\ dxdy.$$

De $xy \leq z$ se tiene que $x \leq \frac{z}{y}$ si $y > 0$ y $x \geq \frac{z}{y}$ si $y < 0$. Luego,

$$F_Z(z) = \int_0^{\infty} \int_{-\infty}^{z/y} f_{X,Y}(x,y)\ dxdy + \int_{-\infty}^{0} \int_{z/y}^{\infty} f_{X,Y}(x,y)\ dxdy.$$

Si $u = xy$ entonces $dx = du/y$ y además para

a) $y > 0$, cuando $x \to -\infty$, $u \to -\infty$ y cuando $x = \frac{z}{y}$, $u = z$.

b) $y < 0$, cuando $x \to \infty$, $u \to -\infty$ y cuando $x = \frac{z}{y}$, $u = z$.

Por tanto,

$$\begin{aligned}
F_Z(z) &= \int_0^{\infty} \bigg(\int_{-\infty}^{z} f_{X,Y}(u/y, y)\ \frac{du}{y} \bigg) dy + \int_{-\infty}^{0} \bigg(\int_{z}^{-\infty} f_{X,Y}(u/y, y)\ \frac{du}{y} \bigg) dy \\
&= \int_{-\infty}^{z} \bigg[\int_0^{\infty} \bigg(\frac{1}{y} \bigg) f_{X,Y}(u/y, y)\ dy \bigg] du + \int_{-\infty}^{z} \bigg[\int_{-\infty}^{0} \bigg(-\frac{1}{y} \bigg) f_{X,Y}(u/y)\ dy \bigg] du \\
&= \int_{-\infty}^{z} \bigg[\int_{-\infty}^{\infty} \bigg(\frac{1}{|y|} \bigg) f_{X,Y}(x/y, y)\ dy \bigg] du.
\end{aligned}$$

O sea que,

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|y|}\ f_{X,Y}(z/y, y)\ dy.$$

A continuación se prueba la ecuación 6.11.

$$F_V(v) = P(V \leq v) = P(\frac{X}{Y} \leq v) = \iint\limits_{x/y \leq v} f_{X,Y}(x,y)\ dxdy.$$

De $\frac{x}{y} \leq v$ se tiene que $x \leq vy$ si $y > 0$, y $x \geq vy$ si $y < 0$. Luego,

$$F_V(v) = \int_0^{\infty} \int_{-\infty}^{vy} f_{X,Y}(x,y)\ dxdy + \int_{-\infty}^{0} \int_{vy}^{\infty} f_{X,Y}(x,y)\ dxdy.$$

Si $u = \frac{x}{y}$ entonces $dx = y\ du$ y además para

a) $y > 0$, cuando $x \to -\infty$, $u \to -\infty$ y cuando $x = vy$, $u = v$.

b) $y < 0$, cuando $x \to \infty$, $u \to -\infty$ y cuando $x = vy$, $u = v$.

Por tanto,

$$\begin{aligned}
F_V(v) &= \int_0^{\infty} \int_{-\infty}^{v} y\ f_{X,Y}(uy,y)\ dudy + \int_{-\infty}^{0} \int_{v}^{-\infty} y\ f_{X,Y}(uy,y)\ dudy \\
&= \int_0^{\infty} \int_{-\infty}^{v} y\ f_{X,Y}(uy,y)\ dudy + \int_{-\infty}^{0} \int_{-\infty}^{v} (-y)\ f_{X,Y}(uy,y)\ dudy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{v} |y|\ f_{X,Y}(uy,y)\ dudy = \int_{-\infty}^{v} \int_{-\infty}^{\infty} |y|\ f_{X,Y}(uy,y)\ dydu.
\end{aligned}$$

O sea que,

$$f_V(v) = \int_{-\infty}^{\infty} |y| f_{X,Y}(vy,y)\ dy.$$

Corolario. Si X y Y son variables independientes, entonces

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} \frac{1}{|x|}\ f_X(x) f_Y(z/x)\ dx \\ &= \int_{-\infty}^{\infty} \frac{1}{|y|}\ f_X(z/y) f_Y(y)\ dy, x \neq 0, y \neq 0. \end{aligned}$$

Ejemplo 6.6. Si X y Y son variables aleatorias independientes, ambas distribuidas uniformemente en $(0,1)$, hallar $f_Z(z)$, donde $Z = XY$.

Solución. Como $f_X(x) = I_{(0,1)}(x)$ y $f_Y(y) = I_{(0,1)}(y)$, por el corolario anterior, se tiene que

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} \frac{1}{|y|}\ f_X(z/y) f_Y(y)\ dy = \int_{-\infty}^{\infty} \frac{1}{|y|}\ I_{(0,1)}(z/y) I_{(0,1)}\ dy \\ &= \int_{0}^{1} \frac{1}{|y|} I_{(0,1)}(z/y)\ dy. \end{aligned}$$

Pero,

$$\begin{aligned} I_{(0,1)}(z/y) &= 1,\ \text{si } 0 < z/y < 1 \\ &= 1,\ \text{si } z > 0 \text{ y } z < y \\ &= I_{(0,1)}(z) I_{(z,1)}(y). \end{aligned}$$

Luego,

$$f_Z(z) = \int_{0}^{1} \frac{1}{y} I_{(0,1)}(z/y) I_{(z,1)}(y)\ dy = I_{(0,1)}(z) \int_{z}^{1} \frac{dy}{y}.$$

O sea que,

$$f_Z(z) = (-\ln z) I_{(0,1)}(z).$$

Para finalizar esta sección, se hallará la distribución de algunas variables aleatorias llamadas estadísticas de orden.

Definición 6.1. Estadísticas de orden. Si $X_1, X_2, \ldots, X_n$ variables aleatorias y $Y_1 \leq Y_2 \leq \cdots \leq Y_n$ son las mismas variables aleatorias en orden ascendente de magnitud así que $Y_1 = min(X_1, X_2, \ldots, X_n)$, Y_2 es el segundo valor más pequeño de $(X_1, X_2, \ldots, X_n)$,..., $Y_n = máx(X_1, X_2, \ldots, X_n)$, entonces

$Y_1, Y_2, \cdots, Y_n$ se llaman las estadísticas de orden correspondientes a las variables aleatorias $X_1, X_2, \ldots, X_n$. Otra notación para las estadísticas de orden son $X_{(1)}$,$X_{(2)}$,..., $X_{(n)}$ donde $X_{(i)} = Y_i$, para $i = 1, 2, ..., n$.

Teorema 6.4. Distribución del mínimo. Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes. Si $Y_1 = X_{(1)}$, es decir, el mínimo de las variables aleatorias $X_1, X_2, \ldots, X_n$ entonces,

$$F_{Y_1}(y) = 1 - \prod_{i=1}^{n}[1 - F_{X_i}(y)].$$

En particular, si $X_1, X_2, \ldots, X_n$ tienen la misma función de distribución F_X entonces,

$$\begin{aligned} F_{Y_1}(y) &= 1 - [1 - F_X(y)]^n, \\ &\text{y} \\ f_{Y_1}(y) &= n[1 - F_X(y)]^{n-1} f_X(y) \end{aligned} \tag{6.12}$$

Demostración. Si $Y_1 = \text{mín}(X_1, X_2, \cdots, X_n)$ entonces,

$$F_{Y_1}(y) = P(Y_1 \leq y) = 1 - P(Y_1 > y) = 1 - P(X_1 > y, \ldots, X_n > y),$$

ya que si Y_1 es mayor que y entonces cada $X_i > y$.

Además, como $X_1, X_2, \ldots, X_n$ son independientes entonces,

$$F_{Y_1}(y) = 1 - \prod_{i=1}^{n} P(X_i > y) = 1 - \prod_{i=1}^{n}[1 - P(X_i \leq y)] = 1 - \prod_{i=1}^{n}[1 - F_{X_i}(y)].$$

En particular, si cada X_i tiene una función de distribución F_X entonces,

$$F_{Y_1}(y) = 1 - [1 - F_X(y)]^n.$$

Derivando esta expresión,

$$f_{Y_1}(y) = \frac{d}{y}\, F_{Y_1}(y) = -n[1 - F_X(y)]^{n-1}[-f_X(y)] = n[1 - F_X(y)]^{n-1} f_X(y).$$

Teorema 6.5. Distribución del máximo. Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes y si $Y_n = X_{(n)}$, es decir, el máximo de las $X_1, X_2, \ldots, X_n$ entonces,

$$F_{Y_n}(y) = \prod_{i=1}^{n} F_{X_i}(y).$$

En particular, si $X_1, X_2, \ldots, X_n$ tienen la misma distribución $F_X(\cdot)$ entonces,

$$\begin{aligned} F_{Y_n}(y) &= [F_X(y)]^n, \\ &\text{y} \\ f_{Y_n}(y) &= n[F_X(y)^{n-1}]f_X(y). \end{aligned} \tag{6.13}$$

Demostración. Si $Y_n = máx(X_1, X_2, \ldots, X_n)$ entonces,

$$F_{Y_n}(y) = P(Y_n \leq y) = P(X_1 \leq y, \ldots, X_n \leq y)$$

ya que si Y_n, el máximo de todas las X_i, es menor o igual que y, entonces todas las X_i, son menores o iguales a y. Además, como $X_1, X_2, \ldots, X_n$ son independientes entonces,

$$F_{Y_n}(y) = \prod_{i=1}^{n} F_{X_i}(y).$$

En particular, si cada X_i tiene función de distibución F_X entonces,

$$F_{Y_n}(y) = [F_X(y)]^n$$

y además

$$f_{Y_n}(y) = \frac{d}{dy} F_{Y_n}(y) = n[F_X(y)^{n-1}]f_X(y).$$

Ejemplo 6.7. Si $X_1, X_2, \ldots, X_5$ son variables aleatorias independientes e idénticamente distribuidas como variables aleatorias exponenciales con parámetro λ hallar,

$$P[máx(X_1, X_2, \ldots, X_5) \leq a].$$

Solución. Sea $Y_5 = máx(X_1, X_2, \ldots, X_5)$. Como cada X_i tiene $F_X(x) = 1 - e^{-\lambda x}$, entonces

$$P(Y_5 \leq a) = [F_X(a)]^5 = [1 - e^{-\lambda a}]^5.$$

En general, se puede obtener también la distribución conjunta de Y_j y Y_k y a partir de esta la marginal de cualquiera de las dos; este tema se trata en el próximo capítulo.

6.3. Técnica de la función generatriz de momentos

Este método para determinar la distribución de funciones de variables aleatorias es particularmente útil cuando se puede reconocer que la función generatriz de momentos de estas, es una función generadora de momentos conocida.

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias con función de densidad conjunta

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f(x_1, x_2, \cdots, x_n).$$

Se quiere encontrar la distribución conjunta de las variables $Y_1 = g_1(X_1, \ldots, X_n)$, $\ldots$, $Y_k = g_k(X_1, \ldots, X_n)$ donde $g_j(\cdot)$ son funciones de variables aleatorias para $j = 1, 2, \ldots, k$.

Si la función generatriz de momentos conjunta de $Y_1, Y_2, \ldots, Y_k$ existe entonces,

$$M_{Y_1,\ldots,Y_k}(t_1, \ldots, t_k) = E[e^{t_1 Y_1 + \cdots + t_k Y_k}] = \int \cdots \int e^{t_1 g_1(x_1,\ldots,x_n) + \cdots + t_k g_k(x_1,\ldots,x_n)} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n.$$

Si la función de $t_1, \ldots, t_k$ resultante de esta integral se reconoce como la función generatriz de momentos conjunta de alguna distribución conjunta conocida, entonces se sigue que $Y_1, \ldots, Y_k$ tiene una distribución conjunta.

Para $k = 1$, la función generatriz de momentos es una función de argumentos simple y se tendrá una mejor ocasión de reconocer la función generatriz de momentos resultante que cuando $k > 1$.

La aplicación más importante de esta técnica se da cuando se quiere encontrar la distribución de sumas de variables aleatorias independientes.

Ejemplo 6.8. Si X tiene una distribución normal $(0, 1)$ y si $Y = X^2$, encontrar la distribución de Y.

Solución. En el ejemplo 6.2 se encontró que Y tiene una distribución chi-cuadrado con un grado de libertad. Se mostrará que aplicando la técnica de la función generatriz de momentos, se llega al mismo resultado,

$$
\begin{aligned}
M_Y(t) = E[e^{tY}] = E[e^{tX^2}] &= \int_{-\infty}^{\infty} e^{tx^2}\left(\frac{1}{\sqrt{2\pi}}\right)e^{-\frac{x^2}{2}}\,dx \\
&= \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}(1-2t)x^2}}{\sqrt{2\pi}}\,dx.
\end{aligned}
$$

Sea $u = (\sqrt{1-2t})x$, entonces $dx = \dfrac{du}{\sqrt{1-2t}}$. Luego,

$$
M_Y(t) = \frac{1}{(1-2t)^{1/2}}\int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}u^2}}{\sqrt{2\pi}}\,du = \left(\frac{1}{1-2t}\right)^{1/2},
$$

que es la función generatriz de momentos de una variable aleatoria con distribución chi-cuadrado con 1 grado de libertad (ver corolario del teorema 4.20).

Ejemplo 6.9. Sean $X_1 \sim N(0,1)$ y $X_2 \sim N(0,1)$, con X_1 y X_2 independientes. Sean $Y_1 = X_1 + X_2$ y $Y_2 = X_2 - X_1$, hallar la función generatriz de momentos conjunta de Y_1 y Y_2. Luego, encontrar la de Y_1 y la de Y_2.

Solución. Como $M_{X_1}(t_1) = e^{t_1^2/2}$ y $M_{X_2}(t_2) = e^{t_2^2/2}$, entonces

$$
\begin{aligned}
M_{Y_1,Y_2}(t_1,t_2) &= E[e^{(t_1Y_1+t_2Y_2)}] = E[e^{t_1X_1+t_1X_2+t_2X_2-t_2X_1}] \\
&= E[e^{(t_2-t_1)X_1}e^{(t_1+t_2)X_2}] \\
&= E[e^{(t_2-t_1)X_1}]E[e^{(t_1+t_2)X_2}] \\
&= M_{X_1}(t_1-t_2)M_{X_2}(t_1+t_2) \\
&= e^{\frac{(t_1-t_2)^2}{2}}e^{\frac{(t_1+t_2)^2}{2}} \\
&= e^{\frac{t_1^2-2t_1t_2+t_2^2+t_1^2+2t_1t_2+t_2^2}{2}} \\
&= e^{t_1^2+t_2^2}.
\end{aligned}
$$

Además,

$$
\begin{aligned}
M_{Y_1}(t_1) &= E[e^{t_1(X_1+X_2)}] = E[e^{t_1X_1}]E[e^{t_1X_2}] \\
&= M_{X_1}(t_1)M_{X_2}(t_1) = e^{t_1^2/2}e^{t_1^2/2} = e^{t_1^2}, \quad \text{y} \\
M_{Y_2}(t_2) &= E[e^{t_2(X_1-X_2)}] = E[e^{t_2X_2}]E[e^{-t_2X_1}] \\
&= M_{X_2}(t_2)M_{X_1}(-t_2) = e^{t_2^2/2}e^{t_2^2/2} = e^{t_2^2}.
\end{aligned}
$$

Se observa que,

$$M_{Y_1,Y_2}(t_1,t_2) = M_{Y_1}(t_1)M_{Y_2}(t_2)$$

Este resultado implica que Y_1 y Y_2 son también variables aleatorias independientes.

Teorema 6.6. Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes y si la función generatriz de momentos de cada una de ellas existe para todo t, $-h < t < h$, entonces

$$M_{\sum\limits_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

Demostración. Suponga que $M_{X_i}(t)$ existe para todo t, $-h < t < h$, $h > 0$. Luego

$$\begin{aligned} M_{\sum\limits_{i=1}^{n} X_i}(t) &= E[e^{t\sum_{i=1}^{n} X_i}] = E[e^{tX_1+tX_2+\cdots+tX_n}] \\ &= E[e^{tX_1}e^{tX_2}\cdots e^{tX_n}] = E[e^{tX_1}]\cdots E[e^{tX^n}] \\ &= \prod_{i=1}^{n} M_{X_i}(t). \end{aligned}$$

Un resultado importante que se desprende de este teorema, es que si $X_1, X_2, \ldots, X_n$ son independientes e idénticamente distribuidas con función generatriz de momentos $M_X(t)$ entonces

$$M_{\sum\limits_{i=1}^{n} X_i}(t) = [M_X(t)]^n. \tag{6.14}$$

Los ejemplos que se dan a continuación ilustran estos hechos.

Ejemplo 6.10. Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes de Bernoulli, cada una con probabilidades de éxito p, hallar la distribución de $\sum\limits_{i=1}^{n} X_i$.

Solución. Como X_i tiene una distribución de Bernoulli con parámetro p, para todo i, entonces

$$M_{X_i}(t) = pe^t + 1 - p.$$

En consecuencia,

$$M_{\sum\limits_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n}[pe^t + 1 - p] = [pe^t + 1 - p]^n,$$

que es la función generatriz de momentos de la distribución binomial con parámetros n y p. Es decir que, $\sum\limits_{i=1}^{n} X_i \sim B(n,p)$.

Ejemplo 6.11. Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes tales que $X_i \sim N(\mu_i, \sigma_i^2)$, demostrar que

$$\sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right).$$

Demostracíon. Primero se prueba que $a_iX_i \sim N(a_i\mu_i, a_i^2\sigma_i^2)$.

Puesto que $X_i \sim N(\mu_i, \sigma_i^2)$ entonces,

$$M_{X_i}(t) = e^{t\mu_i + \frac{t^2\sigma_i^2}{2}}$$

y

$$M_{a_iX_i}(t) = E[e^{ta_iX_i}] = M_{X_i}(ta_i) = e^{(ta_i)\mu_i + \frac{(ta_i)^2\sigma_1^2}{2}} = e^{t(a_i\mu_i) + \frac{t^2(a_i\sigma_i)^2}{2}}.$$

Es decir, $a_iX_i \sim N(a_i\mu_i, a_i^2\sigma_i^2)$ y por tanto,

$$\begin{aligned} M_{\sum\limits_{i=1}^{n} a_iX_i}(t) &= \prod_{i=1}^{n} M_{a_iX_i}(t) = e^{t(a_1\mu_1)+t^2(a_1\sigma_1)^2/2+\cdots+t(a_n\mu_n)+t^2(a_n\sigma_n)^2/2} \\ &= e^{t\sum_{i=1}^{n} a_i\mu_i + \frac{t^2\sum_{i=1}^{n} a_i^2\sigma_i^2}{2}}. \end{aligned}$$

Esta es la función generatriz de momentos de una distribución normal con media $\sum_{i=1}^{n} a_i\mu_i$ y varianza $\sum_{i=1}^{n} a_i^2\sigma_i^2$.

En particular,

a) Si $X \sim N(\mu_X, \sigma_X^2)$ y $Y \sim N(\mu_Y, \sigma_Y^2)$ y, X y Y son independientes entonces,

$$\begin{aligned} X + Y &\sim N(\mu_X + \mu_y, \sigma_X^2 + \sigma_Y^2) \text{ y} \\ X - Y &\sim N(\mu_X - \mu_y, \sigma_X^2 + \sigma_Y^2). \end{aligned}$$

b) Si $X_1, X_2, \ldots, X_n$ son independientes e idénticamente distribuidas como $N \sim (\mu, \sigma^2)$ entonces,

$$\bar{X} = \frac{\sum\limits_{i=1}^{n} X_i}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Los dos teoremas que se dan a continuación usan el resultado b) y son de los teoremas de más importancia en teoría de probabilidad.

Teorema 6.7. *"Ley de los grandes números"*. Sea $X_1, X_2, \ldots$, una sucesión de variables aleatorias independientes e idénticamente distribuidas, cada una con media $E(X_i) = \mu$, entonces para cualquier $\varepsilon > 0$, se tiene que,

$$P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| > \varepsilon\right) \to 0 \text{ cuando } n \to \infty \tag{6.15}$$

La prueba de este teorema se hará en la sección 7.3.

Este teorema permite encontrar el tamaño deseado de una muestra para problemas específicos.

Ejemplo 6.12. ¿Cuántas veces debe lanzarse una moneda legal de manera que la probabilidad de que $\bar{X}_n$ esté entre 0.4 y 0.6 sea al menos 0.90?

Solución. Se debe encontrar un n tal que $P(0.4 \leq \bar{X}_n \leq 0.6) \geq 0.90$.

Sea $X_i = 0$ ó 1 según salga sello o cara. Como la moneda es legal, $P(X = 0) = P(X = 1) = 1/2$, y por tanto, $E(X_i) = 1/2$ y $V(X_i) = 1/4$. Entonces,

$$\begin{aligned} P(0.4 \leq \bar{X}_n \leq 0.6) &= P(-0.1 \leq \bar{X}_n - 0.5 \leq 0.1) \\ &= P(|\bar{X}_n - \mu| \leq 0.1) \\ &= 1 - P(|\bar{X}_n - \mu| > 0.1) \end{aligned}$$

Pero por la desigualdad de Markov,

$$P(|\bar{X}_n - \mu| > 0.1) \leq \frac{\sigma^2}{n(0.1)^2} = \frac{1}{4n(0.01)} = \frac{25}{n}.$$

Luego,

$$P(0.4 \leq \bar{X}_n \leq 0.6) \geq 1 - \frac{25}{n}.$$

Además, por hipótesis

$$0.90 \leq 1 - \frac{25}{n}, \text{ de donde } n \geq 250.$$

Teorema 6.8. *"Teorema central del límite"*. Sea $X_1, X_2, \ldots$, una sucesión de variables aleatorias independientes e idénticamente distribuidas, cada una con media

μ y varianza σ^2, entonces la distribución de

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \text{ tiende a ser normal estándar cuando } n \to \infty.$$

Demostración. Se debe probar que $M_{Z_n}(t) = e^{t^2/2}$ cuando $n \to \infty$, y $Z_n = \dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$.

$$\begin{aligned} M_{Z_n}(t) = E(t^{tZ_n}) &= E\big[e^{t(\bar{X}_n-\mu)/(\sigma/\sqrt{n})}\big] \\ &= E\big[e^{(t/\sqrt{n})\sum_{i=1}^{n}(X_i-\mu)/\sigma}\big]. \end{aligned}$$

Sea $U_i = \frac{X_i-\mu}{\sigma}$ entonces,

$$\begin{aligned} M_{Z_n}(t) = E\left[e^{\frac{t}{\sqrt{n}}\sum\limits_{i=1}^{n} U_i}\right] &= E\big\{e^{(t/\sqrt{n})U_1}e^{(t/\sqrt{n})U_2}\cdots e^{(t/\sqrt{n})U_n}\big\} \\ = \prod_{i=1}^{n} E\{e^{(t/\sqrt{n})U_i}\} &= \prod_{i=1}^{n} M_{U_i}(t/\sqrt{n}). \end{aligned}$$

Puesto que las X_i están idénticamente distribuidas, entonces las U_i están también idénticamente distribuidas, es decir,

$$M_{Z_n}(t) = [M_U(t/\sqrt{n})]^n, \text{ donde } U = (X-\mu)/\sigma.$$

Ahora, una expansión en series de $M_U(t/\sqrt{n})$ produce:

$$M_U(t/\sqrt{n}) = M_U(0) + M_U'(0)\left(\frac{t}{\sqrt{n}}\right) + M_U''(0)\left(\frac{t^2}{2n}\right) + \cdots$$

Pero,

$$M_U(0) = 1, M_U'(0) = E(U) = 0 \text{ y } M_U''(0) = E(U^2) = V(U) = 1, \text{ por tanto,}$$

$$M_U(t/\sqrt{n}) = 1 + \frac{t^2}{2n} + M_U'''(0)\left(\frac{t^3}{2(3)n^{3/2}}\right) + \cdots$$

Luego,

$$M_{Z_n}(t) = \left[1 + \left(\frac{t^2}{2n}\right)(1 + M_U'''(0))\left(\frac{t}{3n^{1/2}}\right) + \cdots\right]^n = \left(1 + \frac{\lambda}{n}\right)^n,$$

donde $\lambda = \left(\dfrac{t^2}{2}\right)\left[1 + M_U'''(0)\left(\dfrac{t}{3n^{1/2}}\right) + \cdots\right]$

Pero λ tiende a $t^2/2$ cuando $n \to \infty$ y también $(1 + \lambda/n)^n \to e^\lambda$ cuando $n \to \infty$. Es decir,

$$M_{Z_n}(t) = e^{t^2/2} \text{ cuando } n \to \infty.$$

Este teorema, probado incialmente por Abraham De Moivre alrededor de 1733 para el caso en que las X_i son variables aleatorias de Bernoulli con $p = 1/2$, fue extendido en su prueba por Laplace para el caso de p arbitrario, y además descubrió la forma más general, que es la dada en este trabajo. La prueba de De Moivre no fue completamente rigurosa. La primera prueba rigurosa la presentó el matemático ruso Liapunoff [8] entre 1901 y 1902.

Ejemplo 6.13. Se lanzan 25 dados legales, encontrar la probabilidad aproximada que la suma obtenida esté entre 65 y 80.

Solución. Sea X_i el valor del i-ésimo dado, para $i = 1, 2, \ldots, 25$. Entonces $E(X_i) = 7/2 = 3.5$ y $V(X_i) = 35/12$. Se debe calcular $P(65 \leq \sum_{i=1}^{25} X_i \leq 80)$.

$$\begin{aligned}
P(65 \leq \sum_{i=1}^{25} X_i \leq 80) &= P(2.6 \leq \bar{X} \leq 3.2) \\
&= P\left(\frac{2.6 - 3.5}{0.3416} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{25}} \leq \frac{3.2 - 3.5}{0.3416}\right) \\
&\equiv \Phi(-1.01) - \Phi(-2.63) = 0.1519.
\end{aligned}$$

6.4. Técnica de la transformación

Se tratará el caso en que los Y_i son funciones de variables aleatorias, pues cuando Y es una función de una sola variable, esta técnica se convierte en la técnica de la función de distribución.

Primero se considera el caso en que las variables aleatorias $X_1, X_2, \ldots, X_n$ son discretas con función de masa de probabilidad conjunta de $p_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ y sea $S = \{(X_1, X_2, \ldots, X_n) : p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) > 0\}$. Se quiere encontrar la función de masa de probabilidad conjunta de $Y_1 = g_1(X_1, \ldots, X_n), \ldots, Y_k = g_k(X_1, \ldots, X_n)$.

Puesto que $Y_1, \ldots, Y_k$ son conjuntamente discretas, se puede probar que

$$p_{Y_1,\ldots,Y_k}(y_1, \ldots, y_k) = P(Y_1 = y_1, \ldots, Y_k = y_k) = \sum p_{X_1,\ldots,X_n}(x_1, \ldots, x_n);$$

donde la suma se toma sobre aquellos valores que están en S para los cuales

$$(y_1, \ldots, y_k) = [g_1(x_1, \ldots, x_n), \ldots, g_k(x_1, \ldots, x_n)].$$

Mediante un ejemplo se mostrará como se hace esto.

Ejemplo 6.14. Sean X_1, X_2, X_3 variables aleatorias que tienen función de masa de probabilidad conjunta dada por la tabla 6.1. Hallar la función de masa de probabilidad conjunta de las variables $Y_1 = X_1 + X_2 + X_3$ y $Y_2 = |X_2 - X_1|$.

(X_1, X_2, X_3)	(0,0,0)	(0,0,1)	(0,1,1)	(1,0,1)	(1,1,0)	(1,1,1)
$p(x_1, x_2, x_3)$	1/8	3/8	1/8	1/8	1/8	1/8

Tabla 6.1: Función de Masa de Probabilidad conjunta de las variables X_1, X_2, X_3 del ejemplo 6.14

Solución. Aquí $S = \{(0,0,0), (0,0,1), (0,1,1), (1,0,1), (1,1,0), (1,1,1)\}$, Y_1 toma valores posibles 0,1,2 y 3 y Y_2 0 y 1.

Ahora bien,
$(y_1, y_2) = (0,0)$ si y sólo si $(x_1, x_2, x_3) = (0,0,0)$, luego $p_{Y_1,Y_2}(0,0) = 1/8$.
$(y_1, y_2) = (1,0)$ si y sólo si $(x_1, x_2, x_3) = (0,0,1)$, luego $p_{Y_1,Y_2}(1,0) = 3/8$.
$(y_1, y_2) = (2,0)$ si y sólo si $(x_1, x_2, x_3) = (1,1,0)$, luego $p_{Y_1,Y_2}(2,0) = 1/8$.
$(y_1, y_2) = (2,1)$ si y sólo si $(x_1, x_2, x_3) = (0,1,1)$ y $(x_1, x_2, x_3) = (1,0,1)$, luego $p_{Y_1,Y_2}(2,1) = 2/8$.
$(y_1, y_2) = (3,0)$ si y sólo si $(x_1, x_2, x_3) = (1,1,1)$, luego $p_{Y_1,Y_2}(3,0) = 1/8$.

La distribución conjunta de Y_1 y Y_2 se da en la tabla 6.2

$Y_2 \setminus Y_1$	0	1	2	3
0	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
1	0	0	$\frac{2}{8}$	0

Tabla 6.2: Función de masa de probabilidad conjunta de Y_1 y Y_2 del ejemplo 6.14

Suponga ahora que $X_1, \ldots, X_n$ son variables aleatorias continuas con función de densidad conjunta $f(x_1, \ldots, x_n)$.

Sea $S = \{(x_1, \ldots, x_2) : f(x_1, \ldots, x_n) > 0\}$. Se desea encontrar la densidad conjunta de $Y_1 = g_1(X_1, \ldots, X_n), \ldots, Y_k = g_k(X_1, \ldots, X_n)$. Donde $g_j(.)$ es una función real para todo j.

El siguiente teorema, que no se demuestra pues requiere algunos elementos de Cálculo Avanzado, permite encontrar esta función.

Teorema 6.9. Sean X_1 y X_2 variables aleatorias continuas con función de densidad $f_{X_1,X_2}(x_1, x_2)$ y sea $S = \{(x_1, x_2) : f_{X_1,X_2}(x_1, x_2) > 0\}$ si

a) $y_1 = g_1(x_1, x_2)$ y $y_2 = g_2(x_1, x_2)$ define una transformación uno a uno de S en el plano Y_1Y_2.

b) $\partial x_1/\partial y_1$, $\partial x_1/\partial y_2$, $\partial x_2/\partial y_1$, $\partial x_2/\partial y_2$ son continuas, donde $x_1 = h_1^{-1}(y_1, y_2)$ y $x_2 = h_2^{-1}(y_1, y_2)$.

c) El jacobiano de la transformación, J es diferente de cero, donde

$$J = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} \\ \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} \end{vmatrix}$$

Entonces,

$$f_{Y_1,Y_2}(y_1, y_2) = |J| f_{X_1,X_2}[h_1^{-1}(y_1, y_2), h_2^{-1}(y_1, y_2)] \tag{6.16}$$

Ejemplo 6.15. Sean X_1, X_2 variables independientes exponenciales con parámetro λ. Encontrar la distribución de $X_1 + X_2$.

Solución. Sea $y_1 = x_1 + x_2$ y $y_2 = x_2$. Entonces $x_1 = y_1 - y_2$ y $x_2 = y_2$.
De esto se tiene que

a) $0 < y_1 < \infty$ y $0 < y_2 < y_1$.

b) $\frac{\partial x_1}{\partial y_1} = 1, \frac{\partial x_1}{\partial y_2} = -1, \frac{\partial x_2}{\partial y_1} = 0$ y $\frac{\partial x_2}{\partial y_2} = 1$. Luego $J = 1$.

Además, $f_{X_1,X_2}(x_1, x_2) = \lambda e^{-\lambda x_1} \lambda e^{-\lambda x_2} = \lambda^2 e^{-\lambda(x_1+x_2)}$. Por tanto, usando 6.16,

$$f_{Y_1,Y_2}(y_1, y_2) = \lambda^2 e^{-\lambda(y_1 - y_2 + y_2)} = \lambda^2 e^{-\lambda y_1},\ 0 < y_1 < \infty,\ 0 < y_2 < y_1.$$

De esta manera, $f_{Y_1}(y_1) = f_{X_1+X_2}(y_1) = \int_0^{y_1} \lambda^2 e^{-\lambda y_1} dy_2 = \lambda^2 y_1 e^{-\lambda y_1}, 0 < y_1 < \infty$.
Esta densidad es la función de densidad de una distribución gamma con parámetros 2 y λ, como era de esperarse (ejemplo 6.5).

Una generalización del teorema anterior es el siguiente:

Teorema 6.10. Sean $X_1, X_2, \cdots, X_n$ variables aleatorias continuas con función densidad de probabilidad conjunta $f(x_1, x_2, \cdots, x_n)$ y sea $S = \{(x_1, x_2, \cdots, x_n) : f(x_1, x_2, \cdots, x_n) > 0\}$. Se define $Y_i = g_i(X_1, X_2, \cdots, X_n)$ para $i = 1, 2, \cdots, k$. Si $S_1, S_2, \cdots, S_k$ forman una partición de S tales que Y_i es una transformación uno a uno de S_i sobre un conjunto $\mathcal{D}$, con $i = 1, 2, \cdots, k$. Sea $x_1 = h_{1i}^{-1}(y_1, y_2, \cdots, y_n), \cdots$, $x_n = h_{ni}^{-1}((y_1, y_2, \cdots, y_n)$ la transformación de $\mathcal{D}$ sobre S_i, $i = 1, 2, \cdots, k$. Defínase

$$J_i = \begin{vmatrix} \frac{\partial h_{1i}^{-1}}{\partial y_1} & \frac{\partial h_{1i}^{-1}}{\partial y_2} & \cdots \frac{\partial h_{1i}^{-1}}{\partial y_n} \\ \frac{\partial h_{2i}^{-1}}{\partial y_1} & \frac{\partial h_{2i}^{-1}}{\partial y_2} & \cdots \frac{\partial h_{2i}^{-1}}{\partial y_n} \\ \vdots & \vdots & \cdots \vdots \\ \frac{\partial h_{ni}^{-1}}{\partial y_1} & \frac{\partial h_{ni}^{-1}}{\partial y_2} & \cdots \frac{\partial h_{ni}^{-1}}{\partial y_n}. \end{vmatrix}$$

Si todas las derivadas parciales en J_i son continuas sobre $\mathcal{D}$ y $J_i \neq 0$ para $i = 1, 2, \cdots, k$, entonces

$f_{Y_1, Y_2, \cdots, Y_n}(y_1, y_2, \cdots, y_n) =$
$\sum_{i=1}^{k} |J_i| f_{X_1, X_2, \cdots, X_n}\big(h_{1i}^{-1}(y_1, y_2, \cdots, y_n), \cdots, h_{ni}^{-1}(y_1, y_2, \cdots, y_n)\big)$.

6.5. Ejercicios

1. Resuelva el ejemplo 6.2 usando la fórmula 6.5.

2. Sean $X \sim B(n, p)$ y $Y \sim B(m, p)$. Si X y Y son independientes, hallar la función de masa de probabilidad conjunta de $X + Y$.

3. Si X y Y son variables aleatorias independientes de Poisson con parámetros λ_1 y λ_2 respectivamente, demostrar que $X + Y$ tiene una distribución de Poisson con parámetro $\lambda_1 + \lambda_2$.

4. Si X y Y son variables aleatorias independientes, ambas distribuidas uniformemente en $(0,1)$, hallar $f_V(v)$, donde $V = X/Y$.

5. Sean X y Y varables aleatorias con función de masa de probabilidad conjunta

$$f(x,y) = \begin{cases} \frac{x(4-y)}{36}, & x = 1,2,3, \ y = 1,2,3. \\ 0, & \text{en otro caso .} \end{cases}$$

   Hallar la función de masa de probabilidad de $V = Y - X$.

6. Dadas las variables aleatorias del ejemplo 6.6, encontrar $P(Y_1 \leq a)$, donde $Y_1 = min(X_1, \ldots, X_5)$.

7. Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes, cada una de ellas con distribución exponencial con parámetro λ, demostrar que $\sum_{i=1}^{n} X_i$ tiene una distribución gamma con parámetros n y λ. Use la técnica de la fgm.

8. Si X y Y son variables aleatorias independientes con distribución exponencial con parámetros λ_1 y λ_2, respectivamente, encontrar $E[m\acute{a}x(X,Y)]$.

9. Sean X y Y variables aleatorias independientes con distribución geométrica. Hallar la función generatriz de momentos conjunta de X y $X+Y$.

10. a) Si $f(x,y) = e^{-(x+y)} I_{(0,\infty)}(x) I_{(0,\infty)}(y)$, encontrar la función de densidad conjunta de X y $X-Y$.

    b) Si $f_X(x) = I_{(0,1)}(x)$, hallar la densidad de $Y = 3X - 1$

11. Si X_1 y X_2 son variables aleatorias independientes, ambas con distribución exponencial con parámetro λ, encontrar la densidad conjunta de $Y_1 = X_1/X_2$ y $Y_2 = X_1 + X_2$, y las distribuciones marginales de Y_1 y Y_2.

12. Si X está distribuida uniformemente en $(0,1)$ y Y como una exponencial con parámetro λ, encontrar la distribución de:

    a) $Z = X + Y$.

    b) $Z = X/Y$.

13. Si X_1 y X_2 son variables aleatorias independientes exponenciales con parámetros λ_1 y λ_2 respectivamente, hallar la distribución de $Z = X_1/X_2$. Encontrar $P(X_1 < X_2)$.

14. La función de densidad de probabilidad conjunta de X y Y esta dada por

$$f(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}},-\infty<x<\infty,-\infty<y<\infty.$$

Determinar la función de densidad de probabilidad conjunta de $R=\sqrt{X^2+Y^2}$ y $\Phi=tan^{-1}\frac{Y}{X}$. ¿Son R y Φ independientes?

15. Sean X_1, X_2, X_3 variables aleatorias independientes y $N(\mu,\sigma^2)$. Defínase

$$Y_1=X+\delta X_3 \text{ y } Y_2=X_2+\delta X_3.$$

a) Encuentre la media y varianza de Y_1 y Y_2. Halle ρ_{XY}.

b) Obtenga $M_{Y_1,Y_2}(t_1,t_2)$.

16. Si X es una variable aleatoria continua con función de densidad

$$f(x)=\frac{1}{\pi(1+x^2)},-\infty<x<\infty.$$

Halle la distribución de $Y=\frac{1}{X}$. ¿Qué observa?

17. Sean X y Y variables aleatorias independientes y normales $(0,1)$.

a) Demostrar que $\frac{X}{X+Y}$ tiene una distribución de Cauchy.

b) Encontrar la distribución de $\frac{X}{|Y|}$.

c) Hallar $P(X^2+Y^2<1)$.

18. Si X y Y son variables aleatorias independientes que tienen la misma distribución geométrica, encuentre la distribución de $Y-X$.

19. Sean X y Y variables aleatorias con función de densidad de probabilidad conjunta $f(x,y)$. Sean $U=aX+b$ y $V=cY+d$, donde a, b, c, y d son constantes fijas con $a>0$ y $c>0$. Probar que

$$f_{U,V}(u,v)=\frac{1}{ac}f\left(\frac{u-b}{a},\frac{v-d}{c}\right).$$

20. Sea $\bar{X}_n$ la media de una muestra de tamaño n y varianza $\sigma^2=36$. Encuentre n tal que $P(\mu-3<X_n<\mu+3)=0.954$.

Capítulo 7

Estadísticos y distribuciones muestrales

En el capítulo 1 se definió el concepto de población como el conjunto total de medidas de un área particular y el de muestra como el de una parte de esta para hacer las inferencias pertinentes. En este capítulo se redefine la noción de muestra, se proporciona la definición de estadístico y se obtienen las distribuciones de algunos estadísticos especiales que juegan un papel importante en los últimos capítulos.

De ahora en adelante, cuando se refiera a función de masa o de densidad de probabilidad se abreviará fmp o fdp respectivamente y se denotará por f. También cuando se escriba $\sum$ y $\prod$ denotarán respectivamente $\sum_{i=1}^{n}$ y $\prod\limits_{i=1}^{n}$.

7.1. Muestra aleatoria

En general, la colección de datos en un experimento consiste en tomar varias observaciones en una variable de interés. Un modelo para recolectar datos es el de una muestra aleatoria.

Definición 7.1. Muestra aleatoria. Las variables aleatorias $X_1, X_2, ..., X_n$ se dicen que es una muestra aleatoria de tamaño n de una población con fmp o fdp $f(\cdot)$ si $X_1, X_2, ...,$ X_n son mutuamente independientes y la fdp o fmp marginal de cada X_i es la misma función $f(\cdot)$. Es decir que, $X_1, X_2, ..., X_n$ es una muestra aleatoria de $f(\cdot)$ si son variables aleatorias independientes y están idénticamente distribuidas (abreviada $v.a.i.i.d$) con fdp o fmp, $f(\cdot)$.

De la definición de fmp o fdp conjunta se tiene que si $X_1, X_2, ..., X_n$ es una muestra aleatoria entonces,

$$f_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n) = \prod_{i=1}^{n} f(x_i). \tag{7.1}$$

Si la fdp o fmp de la población es un miembro de una familia parametrizada por θ, $f(x)$ se escribe como $f(x;\theta)$ y entonces la fdp o fmp conjunta es

$$f(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} f(x_i, \theta). \tag{7.2}$$

Ejemplo 7.1. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta) = \theta e^{-\theta x} I_{(0,\infty)}(x)$ entonces,

$$f(x_2, x_2, \cdots, x_n; \theta) = \prod f(x_1; \theta) = \theta^n e^{-\theta \sum x_i}.$$

Si X es el tiempo hasta que un fusible falla (en años), entonces $X_1, X_2, ..., X_n$ del ejemplo puede corresponder al tiempo de falla para que n fusibles idénticos fallen, y la probabilidad de que todos los fusibles duren a lo sumo un año es

$$\begin{aligned} P(X_1 \leq 1, X_2 \leq 1, ..., X_n \leq 1) &= \int_0^1 \cdots \int_0^1 \prod \theta e^{-\theta x_i} dx_1 \cdots dx_n \\ &= (1 - e^{-\theta}) \int_0^1 \cdots \int_0^1 \prod_{i=2}^{n} \theta e^{-\theta x_i} dx_2 \cdots dx_n \\ &= (1 - e^{-\theta})^n. \end{aligned}$$

Por las propiedades de independencia y variables aleatorias identicamente distribuidas esta probabilidad se puede hallar directamente, pues

$$\begin{aligned} P(X_1 \leq 1, X_2 \leq 1, ..., X_n \leq 1) &= P(X_1 \leq 1)P(X_2 \leq 1) \cdots P(X_n \leq 1) \\ &= \Big[P(X_i \leq 1)\Big]^n = (1 - e^{-\theta})^n. \end{aligned}$$

El modelo de muestreo dado en la definición 7.1 se llama a veces *muestreo de una población infinita*, pensando en obtener los valores de $X_1, X_2, ..., X_n$ en forma secuencial. Primero el experimento se realiza y se obseva $X_1 = x_1$, luego el experimento se repite y se observa $X_2 = x_2$. El supuesto de independencia en el muestreo aleatorio implica que la distribución de probabilidad de X_2 no está afectada por el hecho de que X_1 se observó primero. La eliminación de x_1 de la población infinita no cambia la población, así que $X_2 = x_2$ es todavía una observación aleatoria de la misma población.

Cuando el muestreo es de una población finita como $\{x_1, x_2, \cdots, x_N\}$ la definición 7.1 puede o no ser relevante dependiendo de cómo se realiza la recolección de datos.

Si se toma una muestra $X_1, X_2, \cdots, X_n$ de esta población se puede hacer de dos maneras:

a) Con remplazamiento, en la cual la definición 7.1 se mantiene. Suponga que un valor es seleccionado de la población de manera que cada uno de los N valores son igualmente probables de ser escogidos, $P(\{x_i\}) = \frac{1}{N}$. Este valor se registra como $X_1 = x_1$. Este se **regresa** a la población y de nuevo cada uno de los N valores son igualmente probables de ser escogidos. Se selecciona el segundo valor y se registra como $X_2 = x_2$ (si el mismo valor es escogido, entonces $x_1 = x_2$). Este proceso de sacar individuos de los N valores se repite n veces obteniéndose la muestra $X_1, X_2, \cdots, X_n$. Cada X_i es discreta y toma uno de los valores $\{x_1, x_2, \cdots, x_N\}$ con igual probabilidad. Las variables aleatorias $X_1, X_2, \cdots, X_n$ son independientes porque el proceso de seleccionar cualquier X_i es el mismo.

b) Sin remplazamiento, este se realiza como sigue: Un valor se selecciona del conjunto $\{x_1, x_2, \cdots, x_N\}$ de manera que cada uno de los N valores tiene probabilidad $\frac{1}{N}$ de ser escogido. Este valor se consigna como $X_1 = x_1$. Luego un segundo valor se selecciona de los $N - 1$ valores restantes, donde cada uno de estos tiene probabilidad $\frac{1}{N-1}$. Este valor se registra como $X_2 = x_2$. La escogencia de los valores restantes se continúa de esta manera, obteniéndose $X_1, X_2, \cdots, X_n$, pero una vez que un valor es elegido ya no esta disponible para la próxima escogencia.

Una muestra tomada de una población finita sin reemplazamiento no satisface todas las condiciones de la definición 7.1, pues las variables aleatorias $X_1, X_2, \cdots, X_n$ no son independientes, sin embargo, están identicamente distribuidas (verifíquelo). Tomar muestras sin reemplazamiento de una población finita se llama a veces *muestreo aleatorio simple.*

Es de anotar que cuando el tamaño N de la población es grande comparado con el tamaño n de la muestra se puede asumir independencia.

7.2. Estadísticos y momentos muestrales

Considere las variables aleatorias observables $X_1, X_2, ..., X_n$ y sea $t : R^n \longrightarrow R$ una función.

Definición 7.2. Estadístico. Una función $T = t(X_1, X_2, ..., X_n)$ que no depende de parámetros desconocidos se dice que es un estadístico.

Ejemplo 7.2. Si X tiene distribución normal con parámetros μ y σ^2 entonces, $X - \mu$, $\frac{X}{\sigma}$ y $\frac{X-\mu}{\sigma}$ no son estadísticos.

Ejemplo 7.3. Si $X_1, X_2, ..., X_n$ es una muestra aleatoria de $f(., \theta)$ entonces,

$$\bar{X} = \frac{\Sigma_{i=1}^n X_i}{n} \text{ y } S^2 = \frac{\Sigma_{i=1}^n (X_i - \bar{X})^2}{n-1,}$$

la media y la varianza muestral son estadísticos.

Cuando una muestra aleatoria se observa, los valores de $\bar{X}$ y S^2 se denotan por $\bar{x}$ y s^2. Estos valores son útiles como *estimaciones* de $\mu = E(X)$ y de $\sigma^2 = V(X)$.

Definición 7.3. Momentos muestrales. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población con función de probabilidad $f(\cdot)$. Se definen:

a.) El r- ésimo momento muestral alrededor de 0 por

$$M_r^{'} = \frac{1}{n}\Sigma_{i=1}^n X_i^r. \tag{7.3}$$

Si $r = 1$, entonces $M_1^{'} = \bar{X}$.

b.) El r-ésimo momento muestral alrededor de $\bar{X}$ por

$$M_r = \frac{\sum_{i=1}^n (X_i - \bar{X})^r}{n}. \tag{7.4}$$

Si $r = 2$, entonces $M_2 = \frac{n-1}{n}S^2$

Se observa que los momentos muestrales son ejemplos de estadísticos.

Teorema 7.1. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población con función de probabilidad $f(\cdot)$ entonces,

$$E(M_r^{'}) = \mu_r^{'}, \quad \text{si } \mu_r^{'} \text{ existe y,} \tag{7.5}$$

$$V(M_r^{'}) = \frac{1}{n}\left[\mu_{2r}^{'} - (\mu_r^{'})^2\right] = \frac{1}{n}\left[E(X^{2r}) - E(X^r)^2\right]. \tag{7.6}$$

Demostración. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(.)$ y supongase que μ_r' existe, entonces

$$E(M_r^{'}) = E\left(\frac{1}{n}\Sigma_{i=1}^n X_i^r\right) = \frac{1}{n}\Sigma_{i=1}^n E(X_i^r) = \frac{1}{n}\Sigma_{i=1}^n \mu_r^{'} = \mu_r^{'}$$

$$\begin{aligned} V(M_r^{'}) &= V\left(\frac{1}{n}\Sigma_{i=1}^n X_i^r\right) = \frac{1}{n^2}\Sigma_{i=1}^n V(X_i^r) \\ &= \frac{1}{n^2}\Sigma_{i=1}^n \{E(X_i^{2r}) - [E(X_i^r)]^2\} \\ &= \frac{1}{n^2}\Sigma_{i=1}^n [\mu_{2r}^{'} - (\mu_r^{'})^2] = \frac{1}{n}[\mu_{2r}^{'} - (\mu_r^{'})^2]. \end{aligned}$$

Corolario. Si $X_1, X_2, ..., X_n$ es una muestra aleatoria de una población X con $E(X) = \mu$ y $V(X) = \sigma^2$ entonces,

$$E(\bar{X}) = \mu \text{ y } V(\bar{X}) = \frac{\sigma^2}{n}. \tag{7.7}$$

La propiedad $E(\bar{X}) = \mu$ indica que si la media muestral se usa para *estimar* la media poblacional μ, entonces los valores de las estimaciones estan, en el promedio, en la media poblacional. Un estadístico con esta propiedad se dice que es *estimador insesgado* para el parámetro que se intenta estimar.

Teorema 7.2. Si $X_1, X_2, ..., X_n$ es una muestra aleatoria de una población X con $E(X) = \mu$ y $V(X) = \sigma^2$ entonces,

$$E(S^2) = \sigma^2 \text{ y } V(S^2) = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right),\ n > 1. \tag{7.8}$$

Demostración. Aquí se prueba la primera parte, la otra se deja como ejercicio.

$$\begin{aligned} E(S^2) &= E[\frac{1}{n-1}\Sigma_{i=1}^n (X_i - \bar{X})^2] = \frac{1}{n-1}E(\Sigma_{i=1}^n X_i^2 - n\bar{X}^2) \\ &= \frac{1}{n-1}[\Sigma_{i=1}^n E(X_i^2) - nE(\bar{X}^2)] = \frac{1}{n-1}[\Sigma_{i=1}^n(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2)] \\ &= \frac{1}{n-1}(n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2) = \frac{1}{n-1}(n\sigma^2 - \sigma^2) = \sigma^2. \end{aligned}$$

La propiedad $E(S^2) = \sigma^2$ indica que la varianza muestral es otro ejemplo de un estadístico *insesgado*, y esta es la principal razón para usar el divisor $n-1$ en lugar de n en la definición de varianza muestral.

Un estadístico es también una variable aleatoria y su distribución se conoce como la distribución muestral del estadístico.

7.3. Distribución muestral de $\bar{X}$

En esta sección se presentan dos de los teoremas más importantes de la teoría de probabilidad y se considera la distribución de la media muestral para muestras de algunas poblaciones diferentes a la distribución normal.

El corolario del teorema 7.1 establece que si $X_1, X_2, ..., X_n$ es una muestra aleatoria de $f(\cdot)$ que tiene $E(X) = \mu$ y $V(X) = \sigma^2$ entonces, $E(\bar{X}) = \mu$ y $V(\bar{X}) = \sigma^2/n$. El primer resultado significa que $\bar{X}$ está centrada alrededor de μ. El resultado $V(\bar{X}) = \sigma^2/n$ significa que la dispersión de los valores de $\bar{X}$ alrededor de μ es pequeña para tamaños de muestra grande comparada a un tamaño de muestra pequeño. Por ejemplo, la varianza de la distribución de $\bar{X}_{60}$ es $\frac{1}{3}$ de la varianza de la distribución de $\bar{X}_{20}$. De esta manera, para una muestra de tamaño grande los valores de $\bar{X}$ tienden a estar más "cerca" de μ que para una muestra pequeña. Esto lo establece la "ley de los grandes números".

Teorema 7.3. Ley débil de los grandes números.[1] Sea $X_1, X_2, ...,$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con $\mu = E(X_i)$ y $\sigma^2 = V(X_i) < \infty$ para $i = 1, 2, ...,$. Entonces para todo $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0 \tag{7.9}$$

o equivalentemente,

$$\lim_{n\to\infty} P(| \bar{X}_n - \mu | \leq \epsilon) = 1. \tag{7.10}$$

Demostración. Sea $\epsilon > 0$ entonces por la desigualdad de Chebyshev,

$$P(| \bar{X}_n - \mu | > \epsilon) \leq \frac{V(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}.$$

Luego,

$$\lim_{n\to\infty} P(| \bar{X}_n - \mu | > \epsilon) = 0.$$

[1] Este es un teorema de convergencia en probabilidad

La ley de los grandes números fue originalmente probada por Bernoulli para el caso especial cuando X_i es 0 o 1. La forma como se ha probado en este trabajo fue realizada por el matemático ruso Khintchine.

La siguiente versión de esta ley nos permite estimar el tamaño de muestra.

Teorema 7.4. Ley débil de los grandes números. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población con función de probabilidad $f(\cdot)$ que tiene media μ y varianza finita σ^2. Sean ϵ y δ dos números cualesquiera especificados que satisfacen $\epsilon > 0$ y $0 < \delta < 1$. Si n es cualquier entero mayor que $\sigma^2/\varepsilon^2\delta$ entonces,

$$P(-\epsilon < \bar{X}_n - \mu < \epsilon) \geq 1 - \delta.$$

Ejemplo 7.4. Suponga que alguna distribución con media desconocida μ tiene varianza 4. ¿De qué tamaño debe ser la muestra de manera que con probabilidad 0.9 la media muestral este al menos a 0.2 de la media poblacional?

Solución. De los datos del problema $\sigma^2 = 4$, $\epsilon = 0.2$ y se debe hallar n tal que $P(|\bar{X}_n - \mu| > 0.2) = 0.9$. Esto significa que $P(-0.2 < \bar{X}_n - \mu < 0.2) = 0.1$ y por tanto $\delta = 0.9$. Luego, por el teorema anterior existe n tal que $n > \frac{4}{(0.2)^2(0.9)} = 111.1$. Así que n debe tomarse igual o mayor a 112.

Corolario. Sea $X_1, X_2, ...,$ una sucesión de variables aleatorias independientes tales que $X_i \sim B(1, p)$. Entonces para todo $\epsilon > 0$ se tiene que

$$P(|\bar{X} - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}.$$

En efecto,

$$P(|\bar{X} - p| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2}.$$

Como $p \in (0, 1)$ entonces, el $\max p(1-p)$ ocurre para $p = \frac{1}{2}$, y se obtiene el resultado deseado.

Se asume que si hay una ley débil, tambien existe una ley fuerte de los grandes números. Se da una versión sin demostración.

Teorema 7.5. Ley fuerte de los grandes números. Sea $X_1, X_2, ...,$ una sucesion de variables aleatorias independientes con $E(X_i) = \mu$. Entonces

$$P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1. \tag{7.11}$$

A continuación se presenta otra versión del teorema central del límite el cual da la distribución asintótica de $\bar{X}$. Pero antes se enuncia un teorema que permite concluir la demostración.

Teorema 7.6. Sean X y Y dos variables aleatorias con funciones de distribución F_X y F_Y respectivamente. Si las funciones generadoras de momentos existen y $M_X(t) = M_Y(t)$ para todo $-h < t < h, h > 0$, entonces $F_X(u) = F_Y(u)$ para todo u.

Teorema 7.7. Teorema central del límite.[2] Si para cada entero positivo n, $X_1, X_2, ..., X_n$ son variables aleatorias independientes con $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ para todo $i = 1, 2, ..., n$, entonces para cada z

$$\lim_{n\to\infty} F_{Z_n} = \Phi(z) \tag{7.12}$$

donde, $Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$ y $\Phi(z)$ es la función de distribución de la normal estándar

Demostración. Sea $Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$. En el teorema 6.8 se probó que

$$\lim_{n\to\infty} M_{Z_n}(t) = e^{\frac{1}{2}t^2} = M_Z(t),$$

que es la función generadora de momentos de la distribución normal estándar. Luego, por el teorema anterior, $F_{Z_n}(t) = \Phi(t)$, cuando $n \to \infty$.

El grado de aproximación depende, a menudo, del tamaño de la muestra y de la función $f(\cdot)$ particular.

Corolario. Si para cada entero positivo n, $X_1, X_2, ..., X_n$ son variables aleatorias independientes con $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ para todo $i = 1, 2, ..., n$, entonces para cada z,

$$\lim_{n\to\infty} P\left(\frac{\Sigma X_i - n\mu}{\sqrt{n}\sigma} \leq z\right) = \Phi(z). \tag{7.13}$$

En lo que sigue se da la distribución exacta de $\bar{X}$ para muestras de algunas funciones de probabilidad conocidas.

[2]Este es un teorema de convergencia en distribución

7.3.1. Distribución de $\bar{X}$ para muestras de Bernoulli y de Poisson.

Si $X_1, X_2, ..., X_n$ es una muestra aleatoria de $X \sim B(1,p)$, se sabe del ejemplo 6.10 que $\Sigma_{i=1}^{n} X_i \sim B(n,p)$, así que

$$P\left(\sum_{i=1}^{n} X_i = k\right) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, 2, ..., n.$$

Pero,

$$P\left(\bar{X} = \frac{k}{n}\right) = P\left(\sum_{i=1}^{n} X_i = k\right) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, 2, ..., n.$$

De manera que, $\bar{X}$ toma valores $0, \frac{1}{n}, \frac{2}{n}, ..., 1$ con probabilidades binomiales $(1-p)^n, \binom{n}{1} p(1-p)^{n-1}, \binom{n}{2} p^2 (1-p)^{n-2}, ..., p^n$.

En forma análoga se concluye que $\bar{X}$ toma valores $0, \frac{1}{n}, \frac{2}{n}, ...,$ con probabilidades $e^{-n\lambda}, e^{-n\lambda}(n\lambda), \frac{e^{-n\lambda}(n\lambda)^2}{2!}, ...,$ cuando se toma una muestra de una distribución de Poisson con parámetro λ.

7.3.2. Distribución de $\bar{X}$ para una muestra de la distribución exponencial.

Si $X_1, X_2, ..., X_n$ es una muestra aleatoria de $X \sim E(\lambda)$, se sabe del ejemplo 6.5 que $\sum_{i=1}^{n} X_i \sim G(n, \lambda)$, luego

$$P\left(\sum_{i=1}^{n} X_i \leq y\right) = \int_0^y \frac{1}{\Gamma(n)} z^{n-1} \lambda^n e^{-\lambda z} dz, \qquad y > 0$$

o sea que,

$$P\left(\bar{X} \leq \frac{y}{n}\right) = \int_0^y \frac{1}{\Gamma(n)} z^{n-1} \lambda^n e^{-\lambda z} dz, \qquad y > 0.$$

Si se hace $x = \frac{y}{n}$ entonces,

$$P(\bar{X} \leq x) = \int_0^{nx} \frac{1}{\Gamma(n)} z^{n-1} \lambda^n e^{-\lambda z} dz, \qquad x > 0.$$

Ahora,si $z = nv$, entonces $dz = ndv$ y $v = x$ para $z = nx$, así es que

$$P(\bar{X} \leq x) = \int_0^x \frac{1}{\Gamma(n)} (n\lambda)^n v^{n-1} e^{-(n\lambda)v} dv.$$

Esto significa que $\bar{X}$ tiene distribución gamma con parámetros n y $n\lambda$.

7.4. Distribuciones muestrales de estadísticos de la distribución normal

Muchos estadísticos importantes, usados en pruebas, son expresados como combinación lineal de variables aleatorias independientes.

Teorema 7.8. Si $X_1, X_2, ..., X_n$ son variables aleatorias independientes tales que $X_i \sim N(\mu_i, \sigma_i^2)$, entonces

$$Y = \Sigma_{i=1}^n a_i X_i \sim N\big(\Sigma_{i=1}^n a_i \mu_i, \Sigma_{i=1}^n a_i^2 \sigma_i^2\big).$$

Este resultado se obtuvo en el ejemplo 6.11.

Corolario. Si $X_1, X_2, ..., X_n$ es una muestra aleatoria de $X \sim N(\mu, \sigma^2)$, entonces $\bar{X} \sim N(\mu, \sigma^2/n)$.

Ejemplo 7.5. Sea T el tiempo de vida, en años, de una batería. Suponga que T es aproximadamente normal con media 2 y varianza 1. Si se prueban 16 baterías hasta obtener sus tiempos de vida, ¿cuál es el tiempo correspondiente para que el promedio lo exceda con probabilidad 0.90?

Solución. Sean $T_1, T_2, ...T_{16}$ los tiempos de vida de las 16 baterias. Entonces $T_1, T_2, ...T_{16}$ es una muestra aleatoria de T y por tanto $\bar{T} \sim N(2, 1/16)$. Se desea hallar un valor t tal que $P(T > t) = 0.90$.

Pero $P(T > t) = 1 - P(T \leq t) = 1 - P\left(\frac{\bar{T} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{t-2}{1/4}\right) = 1 - \Phi(\frac{t-2}{1/4}) = 1 - \Phi(4t - 8)$.
Como $P(T > t) = 0.90$ entonces, $1 - \Phi(4t - 8) = 0.9$, lo que implica que $\Phi(4t - 8) = 0.1$, es decir, $4t - 8 = z_{0.1}$. Luego $t = 2 + \frac{1}{4} z_{0.1}$.

De la tabla de la distribución normal $z_{0.1} = -1.28$, de manera que $t = 1.68$.

En general, si $X_1, X_2, ..., X_n$ es una muestra aleatoria de $X \sim N(\mu, \sigma^2)$ para un δ dado, $0 < \delta < 1$, la $P(\bar{X} > t) = \delta$, implica que $t = \mu + \frac{\sigma}{\sqrt{n}} z_{1-\delta}$.

Corolario. Si $X_1, X_2, ..., X_n$ y $Y_1, Y_2, ..., Y_m$ son muestras aleatorias de $X \sim N(\mu_1, \sigma_1^2)$ y $Y \sim N(\mu_2, \sigma_2^2)$ respectivamente y si las dos muestras son independientes entonces,

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right).$$

Del capítulo 4 se conoce que si X tiene distribución chi-cuadrado con k grados de libertad, entonces su función generatriz de momentos es

$$M_X(t) = \left(\frac{1}{1-2t}\right)^{k/2}.$$

El siguiente resultado emplea este hecho.

Teorema 7.9. Si $X_1, X_2, ..., X_n$ son variables aleatorias independientes que tienen distribución chi-cuadrado con $k_1, k_2, \cdots, k_n$ grados de libertad respectivamente entonces,

$$V = \sum_{i=1}^{n} X_i \sim \chi^2_{\sum\limits_{i=1}^{n} k_i} \tag{7.14}$$

Demostración

$$\begin{aligned} M_V(t) &= E(e^{tV}) = E(e^{t\Sigma_{i=1}^n X_i}) = \Pi_{i=1}^n E(e^{tX_i}) = \Pi_{i=1}^n M_{X_i}(t) \\ &= \Pi_{i=1}^n \left(\frac{1}{1-2t}\right)^{k_i/2} = \left(\frac{1}{1-2t}\right)^{\sum_{i=1}^n k_i/2}. \end{aligned}$$

Esta función generatriz de momentos es la de una variable aleatoria que tiene distribución chi-cuadrado con $\Sigma_{i=1}^n k_i$ grados de libertad.

Corolario. Si $X_1, X_2, ..., X_n$ son variables aleatorias independientes tales que $X_i \sim N(\mu_i, \sigma_i^2)$ entonces,

$$Q_n = \sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \sim \chi^2_n. \tag{7.15}$$

Corolario. Si $X_1, X_2, ..., X_n$ es una muestra aleatoria de $X \sim N(\mu, \sigma^2)$ entonces,

$$V = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n \quad \text{y} \quad \frac{n(\bar{X} - \mu)^2}{\sigma^2} \sim \chi^2_1. \tag{7.16}$$

Demostración. Se conoce que si $Z \sim N(0,1)$, entonces $Z^2 \sim \chi^2_1$. Si además $Z_i = \frac{X_i - \mu}{\sigma}$, entonces $Z_i^2 \sim \chi^2_1$ y por el teorema anterior $V = \Sigma_{i=1}^n Z_i^2 \sim \chi^2_n$.

Además, $\dfrac{n(\bar{X}-\mu)^2}{\sigma^2} = \left(\dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2$ y como la expresión dentro del paréntesis es normal $(0,1)$ entonces, $\dfrac{n(\bar{X}-\mu)^2}{\sigma^2}$ es chi-cuadrado con un grado de libertad.

Teorema 7.10. Si $X_1, X_2, ..., X_n$ es una muestra aleatoria de $X \sim N(\mu, \sigma^2)$ entonces,

a) Las variables aleatorias $\bar{X}$ y $X_i - \bar{X}$ son independientes.

b) Las variables aleatorias $\bar{X}$ y S^2 son independientes.

c) La variable aleatoria $\frac{(n-1)S^2}{\sigma^2}$ tiene distribución chi-cuadrado con $n-1$ grados de libertad.

Demostración. Para probar a), como $\Sigma_{i=1}^n (x_i - \bar{x}) = 0$ se tiene que $x_1 - \bar{x} = -\Sigma_{i=2}^n (x_i - \bar{x})$. Ahora, si $y_1 = \bar{x}$ y $y_i = x_i - \bar{x}$ para $i = 2, 3, ..., n$, entonces resultan las siguientes ecuaciones:

$$\begin{aligned}
x_1 - \bar{x} &= -\Sigma_{i=2}^n y_i, \\
x_1 &= y_1 - y_2 - y_3 - ... - y_n, \\
x_2 &= y_1 + y_2, \\
x_3 &= y_1 + y_3, \\
&\vdots \\
x_n &= y_1 + y_n,
\end{aligned}$$

y además $\Sigma_{i=1}^n (x_i - \bar{x})^2 = (-\Sigma_{i=2}^n y_i)^2 + \Sigma_{i=2}^n y_i^2$.

El jacobiano de esta transformación es $J = n$, lo cual puede verificarse fácilmente. Como,

$$\begin{aligned}
f(x_n, x_2, ..., x_n) &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\Sigma_{i=1}^n (x_i-\mu)^2} \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\Sigma_{i=1}^n (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2}
\end{aligned}$$

entonces la función de densidad conjunta, g, de $Y_1, Y_2, ..., Y_n$ es,

$$g(y_1, y_2, ..., y_n) = |J| f(y_1, y_2, ..., y_n).$$

Así que,

$$g(y_1, y_2, ..., y_n) = n\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}[(-\Sigma_{i=2}^n y_i)^2 + \Sigma_{i=2}^n y_i^2 + n(y_1-\mu)^2]}$$

y de esta manera,

$$g(y_1, y_2, ..., y_n) = \left(\frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-\frac{n}{2\sigma^2}(y_1-\mu)^2} \right) \left(\frac{\sqrt{n}}{(\sqrt{2\pi}\sigma)^{n-1}} e^{-\frac{1}{2\sigma^2}[\Sigma_{i=2}^n y_i^2 + (\Sigma_{i=2}^n y_i)^2]} \right)$$

El primer paréntesis de esta expresión es la función de densidad de $Y_1 = \bar{X}$ y el segundo es la función de densidad conjunta de $Y_2, Y_3, ..., Y_n$, es decir, g es el producto de las funciones de densidad de Y_1 y de $Y_2, Y_3, ..., Y_n$, lo que muestra que $\bar{X}$ y $X_i - \bar{X}$ son independientes para $i = 2, 3, ..., n$.

Por lo que $X_1 - \bar{X} = -\sum_{i=2}^{n}(X_i - \bar{X})$, se sigue que $\bar{X}$ y $X_i - \bar{X}$ son independientes.

La prueba de b) es una consecuencia directa del resultado anterior ya que, S^2 es una función sólo de $X_i - \bar{X}$.

Para obtener c) se sabe que, $V = \Sigma_{i=1}^n \left(\frac{X_i-\mu}{\sigma}\right)^2$ tiene distribución chi-cuadrado con n grados de libertad. Pero,

$$V = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X} + \bar{X} - \mu}{\sigma} \right)^2 = \frac{\Sigma_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2}{\sigma^2}$$

$$= (n-1)\frac{S^2}{\sigma^2} + n\frac{(\bar{X}-\mu)^2}{\sigma^2} = V_1 + V_2, \text{ donde } V_1 = (n-1)\frac{S^2}{\sigma}^2 \text{ y } V_2 = n\frac{(\bar{X}-\mu)^2}{\sigma^2}.$$

De acuerdo al corolario del teorema anterior V_2 tiene distribución chi-cuadrado con un grado de libertad y como V_1 y V_2 son independientes, entonces

$$M_V(t) = M_{V_1}(t) M_{V_2}(t).$$

Es decir,

$$M_{V_1}(t) = \frac{M_V(t)}{M_{V_2}(t)} = \frac{(1-2t)^{-\frac{n}{2}}}{(1-2t)^{-\frac{1}{2}}} = \left(\frac{1}{1-2t} \right)^{\frac{n-1}{2}}.$$

Esta es la función generatriz de $V_1 = \frac{(n-1)S^2}{\sigma^2}$, que es la función generatriz de momentos de una chi-cuadrado con $n - 1$ grados de libertad.

Ciertas funciones de muestras de la distribución normal son muy importantes en análisis estadístico. Entre estas se tienen las distribuciones t y F que son útiles para hacer inferencias.

Teorema 7.11. Sea $Z \sim N(0,1)$ y $V \sim \chi^2_n$. Si Z y V son independientes, entonces

$$T = \frac{Z}{\sqrt{\frac{V}{n}}} \sim t_n. \tag{7.17}$$

Demostración. La función de probabilidad conjunta de Z y V es

$$f(z,v) = \frac{v^{n-1}e^{-v/2}e^{-z^2/2}}{\sqrt{2\pi}\Gamma(n/2)2^{n/2}}, \qquad 0 < v < \infty, \qquad -\infty < z < \infty.$$

Si $t = \dfrac{z}{\sqrt{v/n}}$ y $w = v$ entonces, $v = w$ y $z = t\sqrt{w/n}$. Esto implica que el Jacobiano es $J = \sqrt{w/n}$ y por tanto la función de probabilidad conjunta de T y W es,

$$f(t,w) = \frac{(w/n)^{1/2}w^{n/2-1}e^{-w/2}e^{-\frac{t^2w}{2n}}}{\sqrt{2\pi}\Gamma(n/2)2^{n/2}}, \quad 0 < w < \infty, \quad -\infty < t < \infty.$$

De esta manera,

$$f_T(t) = \int_0^{\infty} f(t,w)dw = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(n/2)}(1+t^2/n)^{-\frac{n+1}{2}}, \quad -\infty < t < \infty$$

que es la función de densidad de probabilidad de una variable aleatoria T que tiene distribución t con n grados de libertad(abreviadamente $T \sim t_n$).

Nota.
La distribución t es simétrica alrededor de 0 y su forma general es similar a la distibución normal estándar. Los percentiles, $t_\alpha(n)$, de esta distribución se dan en el apéndice D para valores seleccionados de α y de n.

Corolario. Si $X_1, X_2, ..., X_n$ es una muestra aleatotria de $X \sim N(\mu, \sigma^2)$ entonces,

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}. \tag{7.18}$$

Esto se sigue del hecho que $Z = \sqrt{n}\frac{\bar{X}-\mu}{\sigma} \sim N(0,1)$ y $V = (n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$, y como $\bar{X}$ y S^2 son independientes, entonces, por el teorema $T = \dfrac{Z}{\sqrt{\frac{V}{n-1}}}$ tiene distribución t con $n-1$ grados de libertad.

El siguiente teorema introduce la distribución F a partir de dos variables chi-cuadrado, sin embargo el corolario muestra que esta se puede obtener de muestras de la distribución normal.

Teorema 7.12. Sea U una variable aleatoria con distribución chi-cuadrado con m grados de libertad; sea V otra variable aleatoria con distribución chi-cuadrado con n grados de libertad y, sean U y V independientes entonces,

$$X = \frac{U/m}{V/n} \sim F(m,n). \tag{7.19}$$

Demostración. Como U y V son independientes, la función de densidad conjunta es

$$f(u,v) = \frac{1}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})2^{\frac{m+n}{2}}} u^{\frac{m-2}{2}} v^{\frac{n-2}{2}} e^{-\frac{1}{2}(u+v)}, \;\; u>0, \;\; v>0.$$

Sea $x = \dfrac{u/m}{v/n}$ y $y = v$, entonces el Jacobiano es $J = (m/n)y$ y la función de densidad de X y Y es

$$f(x,y) = \left(\frac{m}{n}\right) y \frac{1}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})2^{\frac{m+n}{2}}} \left(\frac{m}{n}xy\right)^{\frac{m-2}{2}} y^{\frac{n-2}{2}} e^{-\frac{1}{2}[(m/n)xy+y]}, x>0, y>0.$$

Luego,

$$\begin{aligned} f_X(x) &= \int_0^\infty f(x,y)dy \\ &= \frac{1}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})2^{\frac{m+n}{2}}} (\frac{m}{n})^{\frac{m}{2}} x^{\frac{m-2}{2}} \int_0^\infty y^{\frac{m+n-2}{2}} e^{-\frac{1}{2}[(m/n)x+1]y} dy \\ &= \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{m/2} \frac{x^{\frac{m-2}{2}}}{[1+(m/n)x]^{(m+n)/2}}, \;\; x>0 \end{aligned}$$

y esta es la función de densidad de probabilidad de una distribución F con m y n grados de libertad. Ver definición 4.16. (Abreviadamente $X \sim F(m,n)$).

Nota

Los percentiles $F_\alpha(m,n)$ se dan en el apéndice para valores seleccionados de α, m y n. Para valores pequeños de α se pueden obtener usando el hecho, que si $X \sim F(m,n)$, entonces $\frac{1}{X} \sim F(n,m)$.

Corolario. Sea $X_1, X_2, ..., X_m$ una muestra aleatoria de $X \sim N(\mu_1, \sigma^2)$. Sea $Y_1, Y_2, ..., Y_n$ una muestra de $Y \sim N(\mu_2, \sigma^2)$. Si las dos muestras son independientes, entonces

$$F = \frac{\Sigma_{i=1}^{m}(X_i - \bar{X})^2/(m-1)}{\Sigma_{i=1}^{n}(Y_i - \bar{Y})^2/(n-1)} \sim F(m-1, n-1). \qquad (7.20)$$

Este resultado se sigue del hecho que $(m-1)\frac{S_1^2}{\sigma^2}$ y $(n-1)\frac{S_2^2}{\sigma^2}$ tienen distribución chi-cuadrado con $m-1$ y $n-1$ grados de libertad respectivamente, donde

$$S_1^2 = \frac{\Sigma_{i=1}^{m}(X_i - \bar{X})^2}{m-1} \quad \text{y} \quad S_2^2 = \frac{\Sigma_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}.$$

7.5. Distribución de estadísticas de orden

Las estadísticas de orden juegan un papel importante en inferencia estadística y si $X_1, X_2, ..., X_n$ es una muestra aleatoria de una población que tiene función de probabilidad $f(\cdot)$, es común estar interesado en la distribución del "mínimo", del "máximo"y la mediana, por ejemplo. Estos estadísticos son algunos de los llamados *estadísticas de orden.*

Las estadísticas de orden no son independientes, pues si $Y_j \geq y$ entonces $Y_{j+1} \geq y$, luego es obligatorio conocer la distribución conjunta de ellas y las de algunas marginales.

Teorema 7.13. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población que tiene función de distribución $F(.)$ y sean $Y_1, Y_2, ..., Y_n$ las estadísticas de orden. Entonces la distribución marginal de Y_k, para $k = 1, 2, ..., n$, está dada por

$$F_{Y_k}(y) = \sum_{j=k}^{n} \binom{n}{k} [F(y)]^j [1 - F(y)]^{n-j}. \qquad (7.21)$$

Demostración. Sea y dado, y sea

$$Z_i = \begin{cases} 1, & \text{si } x_i \leq y, \\ 0, & \text{en otros casos} \end{cases}$$

Sea $Y = \sum\limits_{i=1}^{n} Z_i$ entonces, Y que es el número de $X_i \leq y$, tiene distribución binomial con parámetros n y $p = F(y)$.

Pero, $\{Y_k \leq y\}$ es equivalente a $\{\sum_{i=1}^{n} Z_i \geq k\}$ (verifíquelo), de manera que,

$$F_{Y_k}(y) = P(Y_k \leq y) = P(Y \geq k) = \sum_{j=k}^{n} \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j}$$

Corolario. La función de distribución del máximo y del mínimo son respectivamente

$$F_{Y_n}(y) = [F(y)]^n \text{ y } F_{Y_1}(y) = 1 - [1 - F(y)]^n.$$

Estos resultados se obtuvieron en el capítulo anterior (ver 6.12 y 6.13).

Teorema 7.14. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población con función de probabilidad $f(\cdot)$ y de distribución $F(\cdot)$. Sean $Y_1, Y_2, ..., Y_n$ las correspondientes estadísticas de orden entonces,

$$f_{Y_k}(y) = \frac{n!}{(k-1)!(n-k)!} [F(y)]^{k-1} [1 - F(y)]^{n-k} f(y) \tag{7.22}$$

$$f_{Y_k,Y_l}(x, y) = \tfrac{n!}{(k-1)!(l-k-1)!(n-l)!} [F(x)]^{k-1}$$

$$\times [F(y) - F(x)]^{l-k-1} [1 - F(y)]^{n-l} f(x) f(y) I_{(x,\infty)}(y) \tag{7.23}$$

y

$$f_{Y_1,Y_2,...,Y_n}(y_1, y_2, ..., y_n) = \begin{cases} n! f(y_1) f(y_2) ... f(y_n), & \text{para } y_1 < y_2 < ... < y_n, \\ 0, & \text{en otro caso} \end{cases} \tag{7.24}$$

Demostración (Ejercicio)
Sugerencia: Con base en F_{Y_k} use la definición de límite para obtener f_{Y_k}.

Además de la media, la varianza, el mínimo y el máximo muestral, hay otros estadísticos, como la mediana que son de interés para el investigador.

Definición 7.4. Mediana y rangos. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población con función de probabilidad $f(\cdot)$ y sean $Y_1, Y_2, ..., Y_n$ las correspondientes estadísticas de orden. Se definen:

-La mediana muestral ($\xi_{.5}$) como la estadística de orden del centro, cuando n es impar, y como el promedio de las dos estadísticas del centro cuando n es par.
-El rango muestral (R), por la diferencia $Y_n - Y_1$ y el semirango muestral por el promedio $\frac{Y_1+Y_n}{2}$.

Nota
Las estadísticas de orden, como se dijo en el capítulo anterior, también se denotan por $X_{(1)}, X_{(2)}, ..., X_{(n)}$, donde $X_{(1)} = Y_1$, $X_{(2)} = Y_2$,...,$X_{(n)} = Y_n$. Con esta notación se tiene que la mediana es

$$\xi_{.5} = \begin{cases} X_{(\frac{n+1}{2})}, & \text{si } n \text{ es impar,} \\ \dfrac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}, & \text{si } n \text{ es par.} \end{cases}$$

Ejemplo 7.6. Si $X_1, X_2, ..., X_n$ es una muestra aleatoria de $f(\cdot)$, halle la distribución de $\xi_{.5}$ si n es impar.

Solución. Si n es impar entonces $n = 2k+1$ para algún $k \in Z^+$, es decir, $\xi_{.5} = Y_{k+1}$. Luego de la fórmula 7.22 se tiene que,

$$\begin{aligned} f_{Y_{k+1}}(y) &= \frac{(2k+1)!}{k!(2k+1-[k+1])!}[F(y)]^k[1-F(y)]^{2k+1-(k+1)}f(y) \\ &= \frac{(2k+1)!}{k!k!}[F(y)]^k[1-F(y)]^k f(y). \end{aligned}$$

Un resultado importante relacionado con la mediana es el teorema que se enuncia a continuacón:

Teorema 7.15. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población con función de distribución F_X que es estrictamente monótona. Si x_p es el percentil del $100p\,\%$ de la población$\big[F_X(x_p) = p\big]$, entonces la estadística de orden $[np]+1$ tiene distribución asintóticamente normal con valor esperado x_p y varianza $\frac{p(1-p)}{n[f_x(x_p)]^2}$. En particular si $p = \frac{1}{2}$, $x_{0.5}$ es la mediana poblacional y si $X \sim N(\mu, \sigma^2)$, la mediana muestral tiene distribución normal con valor esperado μ y varianza $\dfrac{\pi\sigma^2}{2n}$.

7.6. Estadísticos suficientes

Un *estadístico suficiente* para un parámetro θ, es un estadístico que, en un cierto sentido, captura toda la información acerca de θ contenida en la muestra.

Definición 7.5. Estadístico suficiente. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población con función de probabilidad $f(.;\theta)$, donde θ puede ser un vector. Un estadístico $S = s(X_1, X_2, ..., X_n)$ se dice que es *suficiente* para θ si y solo si la distribución condicional de la muestra $X_1, X_2, ..., X_n$ dado $S = s$ no depende de θ, para cualquier valor s de S.

Un aspecto importante de estos estadísticos, está contenido en el *principio de suficiencia* que establece: *Si* $S = s(X_1, X_2, ..., X_n)$ *es un estadístico suficiente para* θ*, entonces cualquier inferencia acerca de* θ *dependerá de la muestra solo a través del valor* $s(x_2, x_2, ..., x_n)$ *de* S.

Ejemplo 7.7. Una moneda, que no se sabe si es legal, se lanza n veces, y el resultado de cada lanzamiento se anota. Se desea "estimar" p, la probabilidad de obtener cara.

Parece ser que el número total de caras, suministra mucha más información acerca de p que los resultados totales. En efecto, sea X_i el número de caras en el lanzamiento i, $i = 1, 2, ..., n$, entonces $X_1, X_2, ..., X_n$ es una muestra aleatoria de $f(x; p) = p^x(1-p)^{1-x}$, para $x = 0, 1$.

Sea S el número total de caras, entonces $S = \sum X_i$ tiene distribución binomial con parámetros n y p. Así que

$$f(s; p) = \binom{n}{s} p^s(1-p)^{n-s}, \text{para } s = 0, 1, ..., n$$

y

$$f(x_1, x_2, ..., x_n; p) = p^{\sum x_i}(1-p)^{n-\sum x_i}, \text{para } x_i = 0, 1.$$

Si $\sum x_i = s$, entonces los eventos $\{X_1 = x_1, X_2 = x_2, ..., X_n = x_n\}$ y $\{X_1 = x_1, X_2 = x_2, ..., X_n = x_n, S = s\}$ son equivalentes y,

$$\begin{aligned} f(x_1, x_2, ..., x_n \mid s) &= \frac{P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n, S = s\})}{P(S = s)} \\ &= \frac{f(x_1, x_2, ..., x_n; p)}{f(s; p)} \\ &= \frac{p^{\sum x_i}(1-p)^{n-\sum x_i}}{\binom{n}{s} p^s(1-p)^{n-s}} = \frac{1}{\binom{n}{s}} = \frac{1}{\dbinom{n}{\sum x_i}} \end{aligned}$$

y esta última expresión no depende de p, de manera que $S = \sum X_i$ es un estadístico suficiente para p.

Una definición que es particularmente útil para demostrar que un estadístico dado no es suficiente es la siguiente:

Definición 7.6. Estadístico suficiente. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población con función de probabilidad $f(.;\theta)$, donde θ puede ser un vector. Un estadístico $S = s(X_1, X_2, ..., X_n)$ se dice que es *suficiente* para θ si y solo si la distribución condicional de T dado S no depende de θ para cualquier estadístico $T = t(X_1, X_2, ..., X_n)$.

El siguiente teorema da un criterio para obtener estadísticos suficientes.

Teorema 7.16. Criterio de factorización. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población con función de probabilidad $f(.;\theta)$, donde θ puede ser un vector. Un estadístico $S = s(X_1, X_2, ..., X_n)$ es suficiente, si y solo si,

$$f(x_1, x_2, ..., x_n; \theta) = g[s(x_1, x_2, ..., x_n; \theta)]h(x_1, x_2, ..., x_n)$$

donde la función h es no negativa y no incluye el parámetro θ y $g(s;\theta)$ es no negativa y depende de $x_1, x_2, ..., x_n$ solo a través de la función $s(., ..., .)$.

Demostración. Se hará la demostración para el caso discreto, el caso continuo se deja como ejercicio.

Supongase que $S = s(X_1, X_2, ..., X_n)$ es un estadístico suficiente y sea $g(s;\theta) = P_\theta(S = s)$ y $h(x_1, x_2, ..., x_n) = P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n | S = s)$.

Puesto que S es suficiente, la probabilidad condicional que define h no depende de θ. De esta manera la escogencia de h y g es legítima y por tanto,

$$\begin{aligned}
f(x_1, x_2, ..., x_n; \theta) &= P_\theta(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) \\
&= P_\theta(X_1 = x_1, X_2 = x_2, ..., X_n = x_n; S = s) \\
&= P_\theta(S = s)P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n | S = s) \\
&= g(s;\theta)h(x_1, x_2, ..., x_n).
\end{aligned}$$

Suponga que $f(x_1, x_2, ..., x_n; \theta) = g(s;\theta)(h(x_1, x_2, ..., x_n)$, donde se asume que para s fijo $h(x_1, x_2, ..., x_n)$ no depende de θ. Note que esto implica que si la función de probabilidad de $X_1, X_2, ..., X_n$ es cero sobre alguna región de Ω, entonces debe ser posible identificar esta región en términos de s y θ, y en términos de $x_1, x_2, ..., x_n$ y $s(x_1, x_2, ..., x_n)$ sin involucrar de otra forma las $x_1, x_2, ..., x_n$ con θ.

Si esto no es posible, entonces la función de probabilidad no está completamente especificada en la forma dada. Básicamente, si la ecuación $f(x_1, x_2, ..., x_n; \theta) = g(s; \theta)(h(x_1, x_2, ..., x_n)$ se cumple para algunas funciones g y h, entonces la función de probabilidad marginal de S debe ser de la forma

$$f_S(s; \theta) = g(s; \theta)c(s)$$

porque para s fijo sumando sobre el resto de variables no se puede incluir θ en la función. Así que

$$\begin{aligned} f(x_1, x_2, ..., x_n | s) &= \frac{f(x_1, x_2, ..., x_n; s)}{f_S(s; \theta)} \\ &= \frac{g(s(x_1, x_2, ..., x_n); \theta)(h(x_1, x_2, ..., x_n))}{f_S(s(x_1, x_2, ..., x_n); \theta)} \\ &= \frac{h(x_1, x_2, ..., x_n)}{c(s)} \end{aligned}$$

la cual es independiente de θ.

Ejemplo 7.8. Si $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x; p) = p^x(1-p)^{1-x}$, para $x = 0, 1$, halle un estadístico suficiente para p.

Solución. La función de probabilidad conjunta es

$$\begin{aligned} f((x_1, x_2, ..., x_n; p) &= \prod p^{x_i}(1-p)^{1-x_i} I_{\{0,1\}}(x_i) \\ &= p^{\sum x_i}(1-p)^{n-\sum x_i} \Pi I_{\{0,1\}}(x_i) \\ &= g(s; p)h(x_1, x_2, ..., x_n) \end{aligned}$$

donde $s = \sum x_i$ y $h(x_1, x_2, ..., x_n) = \prod I_{\{0,1\}}(x_i)$. Luego $S = \sum X_i$ es un estadístico suficiente para p.

Ejemplo 7.9. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim U(0, \theta)$, donde $\theta > 0$ y desconocido. Halle un estadístico suficiente para θ.

Solución. Si $X \sim U(0, \theta)$ entonces, $f(x; \theta) = \frac{1}{\theta}$, para $0 < x < \theta$, es decir, $f(x; \theta) = \frac{1}{\theta} I_{(0,\theta)}(x)$. Luego

$$f(x_1, x_2, ..., x_n; \theta) = \prod f(x_i, \theta) = \prod \frac{1}{\theta} I_{(0,\theta)}(x_i) = \frac{1}{\theta^n} \prod I_{(0,\theta)}(x_i).$$

Pero,

$$\begin{aligned}\prod I_{(0,\theta)}(x_i) &= 1, \text{para todo } 0 < x_i < \theta \\ &= 1, \text{para todo } y_1 > 0 \ \text{ y } y_n < \theta \\ &= I_{(0,\infty)}(y_1) I_{(0,\theta)}(y_n).\end{aligned}$$

De manera que,

$$f(x_1, x_2, ..., x_n; \theta) = \frac{1}{\theta^n} I_{(0,\theta)}(y_n) I_{(0,\infty)}(y_1) = g(y_n; \theta) h(x_1, x_2, ..., x_n)$$

y por tanto, $S = Y_n = máx(X_1, X_2, ..., X_n)$ es un estadístico suficiente para θ.

Para algunas funciones de probabilidad con parámetro θ, no existe un estadístico suficiente, sin embargo, siempre existe un conjunto de estadísticos suficientes conjuntamente.

Definición 7.7. Estadísticos suficientes conjuntamente. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población con función de probabilidad $f(.;\theta)$. Los estadísticos $S_1, S_2, ..., S_k$ se dicen que son suficientes conjuntamente para θ si y solo si la distribución condicional de $X_1, X_2, ..., X_n$ dado $S_1 = s_1, S_2 = s_2, ..., S_k = s_k$ no depende de θ.

Ejemplo 7.10. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$ y sean $Y_1, Y_2, ..., Y_n$ las correspondientes estadísticas de orden. Entonces $Y_1, Y_2, ..., Y_n$ son estadísticos suficientes conjuntamente.

En efecto, los conjuntos $\{X_1 = x_1, X_2 = x_2, ..., X_n = x_n; Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n\}$ y $\{X_1 = x_1, X_2 = x_2, ..., X_n = x_n\}$ son equivalentes, por tanto

$$\begin{aligned}f(x_1, x_2, ..., x_n | y_1, y_2, ..., y_n) &= \frac{f(x_1, x_2, ..., x_n)}{f(y_1, y_2, ..., y_n)} \\ &= \frac{\prod f(x_i)}{n! \prod f(y_i)} = \frac{1}{n!}.\end{aligned}$$

Teorema 7.17.. Si $S_1, S_2, ..., S_k$ son estadísticos suficientes conjuntamente, entonces cualquier conjunto de funciones uno a uno o transformaciones, de $S_1, S_2, ..., S_k$ es también suficiente conjuntamente.

Por ejemplo, si $\sum X_i$ y $\sum X_i^2$ son estadísticos suficientes conjuntamente, entonces $\bar{X}$ y S^2 son también suficientes conjuntamente. La demostración de este teorema se deja como ejercicio.

El siguiente teorema, conocido como segundo criterio de factorización da un método para obtener estadísticos suficientes conjuntos.

Teorema 7.18. Criterio de factorización. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población con función de probabilidad $f(.;\theta)$, donde θ puede ser un vector. Un conjunto de estadísticos $\{S_1, S_2, ..., S_k\}$, donde $S_i = s_i(X_1, X_2, ..., X_n)$ es suficiente conjuntamente si y solo si la función de probabilidad conjunta, f, se puede escribir como

$$f(x_1, x_2, ..., x_n; \theta) = g(s_1, s_2, ..., s_k; \theta)h(x_1, x_2, ..., x_n)$$

donde h es no negativa y no incluye el parámetro θ y g es no negativa y depende de $x_1, x_2, ..., x_n$ solo a través de las funciones $s_1, s_2, ..., s_k$.

Ejemplo 7.11. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu, \sigma^2)$, donde μ y σ^2 son desconocidos. Halle estadísticos suficientes conjuntamente para $\theta = (\mu, \sigma^2)$.

Solución.

$$f(x_1, x_2, ..., x_n; \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{\left[-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\right]}$$

Pero

$$\sum(x_i - \mu)^2 = \sum x_i^2 - 2\mu\sum x_i + n\mu^2$$

Sea $s_1 = \sum x_i$ y $s_2 = \sum x_i^2$ luego,

$$\begin{aligned} f(x_1, x_2, ..., x_n; \mu, \sigma^2) &= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}s_2 + \frac{\mu}{\sigma^2}s_1} e^{-\frac{n}{2}(\frac{\mu}{\sigma})^2} \\ &= g(s_1, s_2; \mu, \sigma^2)h(x_1, x_2, ..., x_n) \end{aligned}$$

donde $h(x_1, x_2, ..., x_n) = 1$.

Por lo anterior, $S_1 = \sum X_i$ y $S_2 = \sum X_i^2$ son estadísticos suficientes conjuntamente para $\theta = (\mu, \sigma^2)$.

Ejemplo 7.12. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim U(\theta, \theta + 1)$, donde $\theta > 0$ y desconocido. Halle estadísticos suficientes para θ.

Solución. Si $X \sim U(\theta, \theta+1)$ entonces $f(x;\theta) = 1$, para $\theta < x < \theta + 1$. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$ entonces, $X_1, X_2, ..., X_n$ es una muestra de $f(x;\theta) = I_{(\theta,\theta+1)}(x)$. De manera que

$$f(x_1, x_2, ..., x_n; \theta) = \Pi I_{(\theta,\theta+1)}(x_i).$$

Esta función toma el valor 1 si y solo si $\theta < x_i$ y $x_i < \theta + 1$ para todo i. Por consiguiente

$$\begin{aligned} f(x_1, x_2, ..., x_n; \theta) &= \Pi I_{(\theta,\infty)}(x_i) \Pi I_{(-\infty,\theta+1)}(x_i) \\ &= I_{(\theta,\infty)}(y_1) I_{(-\infty,\theta+1)}(y_n) \\ &= g(y_1, y_n; \theta) h(x_1, x_2, ..., x_n). \end{aligned}$$

Por el criterio de factorización $S_1 = Y_1$ y $S_2 = Y_n$ son estadísticos suficientes conjuntamente para θ. Aquí $h(x_1, x_2, ..., x_n) = 1$.

En la busqueda de estadísticos suficientes el concepto de familia exponencial ofrece algunos criterios para obtenerlos.

Definición 7.8. Familia exponencial. Una familia de probabilidades (fmp o fdp), $\{f(.;\theta) : \theta \in \Theta\}$, donde θ es unidimensional, se dice que es de la familia exponencial si $f(\cdot;\theta)$ se puede expresar como

$$f(x;\theta) = a(\theta) b(x) e^{c(\theta)d(x)}$$

Para $-\infty < x < \infty$, para todo $\theta \in \Theta$ el espacio de parámetros, y para escogencias apropiadas de las funciones $a(.)$, $b(.)$, $c(.)$, y $d(.)$.

Ejemplo 7.13. La función de masa de probabilidad de la distribución de Bernoulli, $f(x;p) = p^x(1-p)^{1-x}$ para $x = 0, 1$, pertenece a la familia exponencial.

En efecto,

$$\begin{aligned} f(x;p) &= \frac{p^x}{(1-p)^x}(1-p) I_{\{0,1\}}(x) = \left(\frac{p}{1-p}\right)^x (1-p) I_{\{0,1\}}(x) \\ &= (1-p) I_{\{0,1\}}(x) e^{ln\left(\frac{p}{1-p}\right)^x} = (1-p) I_{\{0,1\}}(x) e^{\left[x \ln \frac{p}{1-p}\right]} \\ &= a(p) b(x) e^{c(p)d(x)} \end{aligned}$$

donde $a(p) = 1 - p$, $b(x) = I_{\{0,1\}}(x)$, $c(p) = \ln\left(\frac{p}{1-p}\right)$ y $d(x) = x$.

Ejemplo 7.14. La función de densidad de probabilidad de la distribución exponencial, $f(x;\lambda) = \lambda e^{-\lambda x}$ para $x > 0$ y $\lambda > 0$, pertenece a la familia exponencial.

En efecto,

$$f(x;\lambda) = \lambda e^{-\lambda x} I_{(0,\infty)}(x) = \lambda I_{(0,\infty)}(x) e^{[-\lambda x]} = a(\lambda)b(x)e^{c(\lambda)d(x)}$$

con $a(\lambda) = \lambda$, $b(x) = I_{(0,\infty)}(x)$, $c(\lambda) = -\lambda$ y $d(x) = x$.

Teorema 7.19. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población con función de probabilidad $f(.;\theta)$. Si f pertenece a la familia exponencial entonces, $\sum d(x_i)$ es un estadístico suficiente para θ.

Demostración. Si f pertenece a la familia exponencial, entonces

$$\begin{aligned} f(x_1, x_2, ..., x_n; \theta) = \prod f(x_i;\theta) = \prod a(\theta)b(x_i)e^{c(\theta)d(x_i)} &= a^n(\theta)\big[\Pi b(x_i)\big]e^{c(\theta)\sum d(x_i)} \\ = a^n(\theta)e^{c(\theta)\sum d(x_i)} \prod b(x_i) &= g\Big(\sum d(x_i);\theta\Big)h(x_1, x_2, ..., x_n) \end{aligned}$$

con $h(x_1, x_2, ..., x_n) = \Pi b(x_i)$. De esta manera, $\Sigma d(X_i)$ es un estadístico suficiente para θ.

En los ejemplos 7.13 y 7.14 se mostró que las funciones de probabilidad de la distribución de Bernoulli y de la distribución exponencial pertenecen a la familia exponencial, por tanto, $\sum X_i$ es un estadístico suficiente tanto para p como para λ.

Definición 7.9. Familia exponencial de k parámetros. Una familia de fmp o de fdp $\{f(.;\theta_1, \theta_2, ..., \theta_k) : \theta_1, \theta_2, ..., \theta_k \in \Theta\}$, se dice que es de la familia exponencial si f se puede expresar como

$$f(x;\theta_1, \theta_2, ..., \theta_k) = a(\theta_1, \theta_2, ..., \theta_k)b(x)e^{\sum c_j(\theta_1,\theta_2,...,\theta_k)d_j(x)}.$$

Para escogencias apropiadas de las funciones a, b, c_j y d_j y $j = 1, 2, ..., k$.

Ejemplo 7.15. Sea $f(x;\theta) = \phi_{\mu,\sigma^2}(x)$ la fdp de la distribución normal, entonces f pertenece a la familia exponencial.

En efecto,

$$\begin{aligned} f(x;\theta) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x^2-2x\mu+\mu^2)} \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}x^2+\frac{\mu}{\sigma^2}x-\frac{\mu^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\mu^2}{2\sigma^2}} e^{-\frac{1}{2\sigma^2}x^2+\frac{\mu}{\sigma^2}x} \\ &= a(\mu,\sigma^2)b(x)e^{\sum_{j=1}^{2} c_j(\mu,\sigma^2)d_j(x)} \end{aligned}$$

donde, $a(\mu,\sigma^2) = (1/\sqrt{2\pi}\sigma)e^{-\frac{\mu^2}{2\sigma^2}}$, $b(x) = 1$, $c_1(\mu,\sigma^2) = -\frac{1}{2\sigma^2}$, $c_2(\mu,\sigma^2) = \frac{\mu}{\sigma^2}$, $d_1(x) = x^2$, y, $d_2(x) = x$.

Teorema 7.20. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población con función de probabilidad $f(.;\theta_1,\theta_2,...,\theta_k))$. Si f pertenece a la familia exponencial, entonces $\sum d_1(x_i), \sum d_2(x_i), ..., \sum d_k(x_i)$ es un conjunto de estadísticos conjuntamente suficientes para $\theta_1, \theta_2, ..., \theta_k$.

Demostración. (Ejercicio)

En el ejemplo anterior y de acuerdo a este teorema $\sum X_i$ y $\sum X_i^2$ son estadísticos conjuntamente suficientes para los parámetros μ y σ^2 de la distribución normal.

7.7. Aproximación para grandes muestras

El teorema central del límite nos permite llegar al siguiente resultado:

Teorema 7.21. Si $Y_n \sim \chi^2_n$, entonces, $Z_n = \dfrac{Y_n - n}{\sqrt{2n}}$ tiene distribución aproximadamente normal (0,1), cuando $n \longrightarrow \infty$.

Demostración. Sean $X_1, X_2, ..., X_n$ variables aleatorias independientes tales que $X_i \sim \chi^2_1$. Entonces $Y_n = \Sigma_{i=1}^n X_i \sim \chi^2_n$. Esto significa que $E(Y_n) = n$, $V(Y_n) = 2n$ y por el teorema central del límite

$$Z_n = \frac{Y_n - n}{\sqrt{2n}} \longrightarrow^d Z \sim N(0,1).$$

De esto se sigue que percentiles chi-cuadrado se pueden aproximar en términos de percentiles normales estándar para n grande. Específicamente,

$$\alpha = P\big(Y_n \leq \chi^2_n(\alpha)\big) = \Phi\left(\frac{\chi^2_n(\alpha) - n}{\sqrt{2n}}\right).$$

De manera que, $z_\alpha \approxeq \dfrac{\chi^2_n(\alpha) - n}{\sqrt{2n}}$ y por tanto, $\chi^2_n(\alpha) = n + z_\alpha\sqrt{2n}$.

Por ejemplo para $n = 50$ y $\alpha = 0.9$ se tiene que $\chi^2_n(0.9) \approxeq 50 + 1.28(10) = 62.8$. Comparado con el valor exacto que es 63.17 se nota que es una buena aproximación.

Ejemplo 7.16. Si $X_1, X_2, ..., X_n$ es una muestra aleatoria de $X \sim N(\mu, \sigma^2)$, pruebe que $\dfrac{\sqrt{n-1}(S^2_n - \sigma^2)}{\sqrt{2}\sigma^2}$ es aproximadamente normal cuando n es grande.

Prueba. Se conoce que $V_n = \frac{(n-1)S^2}{\sigma^2}$ tiene distribución chi-cuadrado con $n-1$ grados de libertad. Así que por el teorema anterior,

$$\frac{V_n - (n-1)}{\sqrt{2(n-1)}} \longrightarrow^d Z \sim N(0,1)$$

pero,

$$\frac{V_n - (n-1)}{\sqrt{2(n-1)}} = \frac{\frac{(n-1)S^2}{\sigma^2} - (n-1)}{\sqrt{2(n-1)}} = \frac{(n-1)(S^2 - \sigma^2)}{\sqrt{2(n-1)}\sigma^2} = \frac{\sqrt{n-1}(S^2 - \sigma^2)}{\sqrt{2}\sigma^2}.$$

Es decir,

$$\frac{\sqrt{n-1}(S^2_n - \sigma^2)}{\sqrt{2}\sigma^2} \longrightarrow^d Z \sim N(0,1).$$

Esto también significa, por el teorema 4.14 que, aproximadamente,

$$S^2_n \sim N\left(\sigma^2, \frac{2\sigma^4}{n-1}\right).$$

Además, si $Y_n = S^2_n$, y $g(y) = \sqrt{y}$, entonces $g'(y) = \frac{1}{2}y^{-\frac{1}{2}}$, $g'(\sigma^2) = \frac{1}{2\sigma}$ y aproximadamente

$$S_n \sim N\left(\sigma, \frac{\sigma^2}{2(n-1)}\right).$$

Un resultado similar al teorema anterior se obtiene para cuando $Y_n \sim t_n$. En este caso $Z_n = \dfrac{Y_n}{\sqrt{n/(n-2)}} \longrightarrow^d Z \sim N(0,1)$ cuando $n \longrightarrow \infty$, y percentiles t pueden ser aproximados en términos de percentiles normales estándar para n grande. Específicamente,

$$\alpha = P\big(Y_n \leq t_n(\alpha)\big) = P\left(\frac{Y_n}{\sqrt{n/(n-2)}} \leq \frac{t_n(\alpha)}{\sqrt{n/(n-2)}}\right) = \Phi\left(\frac{t_n(\alpha)}{\sqrt{n/(n-2)}}\right).$$

De esta manera, $z_\alpha = \dfrac{t_n(\alpha)}{\sqrt{n/(n-2)}}$ y por tanto, $t_n(\alpha) = \sqrt{\frac{n}{n-2}} z_\alpha$.

Así, por ejemplo si $\alpha = 0.975$ y $n = 120$, entonces de la tabla de la distribución t, $t_{120}(0.975) = 1.980$. De la tabla de la normal estandar $z_{0.975} = 1.96$ y entonces $t_{120}(0.975) \approxeq 1.9765$.

7.8. Ejercicios

1. Sean X_1 y X_2 variables aleatorias normales cada una con media μ y varianza σ^2 , y sean $Y_1 = X_1 + X_2$ y $Y_2 = X_1 - X_2$. Probar que Y_1 y Y_2 son variables aleatorias independientes y normalmente distribuidas.

2. Sean $X \sim \chi^2_m$, $Y \sim \chi^2_n$ e independientes. Si $n > m$ ¿será $Y - X \sim \chi^2$?

3. Suponga que $X \sim \chi^2_m$ y que $W = X + Y \sim \chi^2_{m+n}$. Si X y Y son independientes, use la función generadora de momentos para probar que $Y \sim \chi^2_n$.

4. Una muestra aleatoria de tamaño 25 es tomada de $X \sim E(\lambda)$. Encuentre k tal que $P(k\bar{X} < \lambda) = 0.95$.

5. Sea $Z_1, Z_2, ..., Z_{16}$ una muestra aleatoria de $Z \sim N(0, 1)$. Encuentre,

   a) $P(\bar{Z} < \frac{1}{2})$.

   b) $P\big[\sum_{i=1}^{16} Z_i^2 < 32\big]$.

   c) $P\left[\sum_{i=1}^{16} (Z_i - \bar{Z})^2 < 36\right]$.

6. Si $T \sim t_n$, encuentre la distribución de T^2.

7. Sea $X_1, X_2, ..., X_{25}$ una muestra aleatoria de $X \sim N(9, 16)$. Encuentre,

   a) $P(4 < \bar{X} < 12)$.

   b) $P(2.54 < 5(\bar{X} - 9)/4)$

   c) $P(S^2 < 31.9375)$.

8. Suponga que $Z \sim N(0,1)$, $V_1 \sim \chi_5^2$, $V_2 \sim \chi_9^2$, y que son independientes. Encuentre
   a) $P(V_1 + V_2 < 8.6)$
   b) $P(Z > 0.611\sqrt{V_2})$
   c) $P(V_1/V_2 < 1.45)$
   d) El valor de k tal que $P(\frac{V_1}{V_1+V_2} < k) = 0.95$.

9. Demostrar que si $X \sim F(2, 2b)$, entonces $P(X > x) = (1 + \frac{x}{b})^{-b}$.

10. Sea $X_1, X_2, ..., X_{12}$ una muestra de $X \sim B(1,2)$, use el teorema central del límite para aproximar $P(\bar{X} \leq 0.5)$.

11. Sean X_1, X_2, X_3 variables aleatorias independientes tales que $X_i \sim N(i, i^2)$. Usando solo estas tres variables aleatorias, dar un ejemplo de un estadístico que tenga,

    a) Distribución χ_3^2.
    b) Distribución $F(1,2)$.
    c) distribución t_2.

12. Demuestre el teorema 7.5.

13. Sea U_1, U_2 una muestra aleatoria de $U \sim (0,1)$. Sean Y_1 y Y_2 las correspondientes estadísticos de orden.

    a) Para $0 < y_2 < 1$, ¿cuál es la distribución condicional de Y_1 dado $Y_2 = y_2$.
    b) ¿Cual es la distribución de $Y_2 - Y_1$?

14. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(\cdot)$, halle la distribución $\xi_{.5}$ si n es par.

15. Demuestre el teorema 7.14

16. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim P(\lambda)$. Verifique que $S = \sum X_i$ es suficiente para λ.

17. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim G(p)$. Pruebe que $S = \sum X_i$ es suficiente para p.

18. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim G(2, \theta)$. Pruebe que $S = \sum X_i$ es suficiente para θ.

19. Suponga que $X_1, X_2, ..., X_n$ son variables aleatorias independientes tales que $X_i \sim B(n_i, p)$. Pruebe que $S = \sum X_i$ es suficiente para p.

20. Suponga que $X_1, X_2, ..., X_n$ son variables aleatorias independientes con $X_i \sim Bn(r_i, p)$. Encuentre un estadístico suficiente para p.

21. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu, \sigma^2)$. Encuentre un estadístico suficiente para,

    a) μ si σ^2 es conocido.

    b) σ^2 si μ es conocido.

22. Considere una muestra aleatoria de $X \sim U(\theta_1, \theta_2)$. Demuestre que

    a) Y_1 es suficiente para θ_1 si θ_2 es conocido.

    b) Y_1 y Y_n son conjuntamente suficientes para θ_1 y θ_2.

23. Suponga que $X_1, X_2, ..., X_n$ es una muestra aleatoria de $X \sim B(\theta_1, \theta_2)$. Encuentre estadísticos suficientes conjuntamente para θ_1 y θ_2

24. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta) = \frac{ln\theta}{\theta-1}\theta^x I_{(0,1)}(x), \theta > 0$. Encuentre un estadístico suficiente para θ, si existe.

25. Si $X \sim W(a, b)$ demostrar que

    a) Si $a > 0$ pero fijo y $b \in \Theta = (0, \infty)$, entonces $f(x; b)$ pertenece a la familia exponencial.

    b) Si a y b son desconocidos, entonces $f(x; a, b)$ no pertenece a la familia esxponencial.

26. Demuestre el teorema 7.15

27. Demuestre el teorema 7.17

28. Demuestre el teorema 7.20

Capítulo 8

Estimación puntual de parámetros

8.1. Introducción

En el capítulo anterior se comentó que una muestra es útil para hacer inferencias acerca de una población. Dos problemas fundamentales de la inferencia estadística son la estimación de parámetros y las pruebas de hipótesis. La estimación puede hacerse de dos maneras: en forma *puntual*, que es el objeto de este capítulo, y por *intervalos*, el cual se tratará a continuación de este.

El problema es el siguiente: suponga que el resultado de un experimento es una variable aleatoria X y que $f(x;\theta)$ representa su fmp o su fdp. Con base en una muestra aleatoria $X_1, X_2, ..., X_n$ tomada de X (también se dice tomada de $f(x;\theta)$) se estimará o representará θ o una función de θ, $\tau(\theta)$. En general, una función de la muestra resuelve el problema.

Definición 8.1. Estimador puntual. Un estadístico, $T = t(X_1, X_2, ..., X_n)$ que se emplea para estimar o representar el valor de $\tau(\theta)$ se llama un estimador puntual de $\tau(\theta)$ y un valor observado del estadístico, $\mathbf{t} = t(x_1, x_2, ..., x_n)$ se llama una *estimación.*

Nótese que en la definición se han usado tres tipos de la letra "te" en la notación: T que representa el estadístico usado como estimador, t que representa la función y $\mathbf{t}$ que representa un valor de T. Otra notación, que se empleará en este texto, incluye un acento circunflejo para estas representaciones, así: $\hat{\Theta}$ representa el estimador y $\hat{\theta}$ la estimación.

Para aclarar el concepto, dada una muestra de una población, $\bar{X} = \frac{\Sigma X_i}{n}$ es un estimador puntual de μ y $\bar{x} = \frac{\Sigma x_i}{n}$ es una estimación. Aquí T es $\bar{X}$, t es $\bar{x}$ y $\mathbf{t}$ es la suma de los argumentos dividido por n.

De la definición se nota que hay dos problemas por resolver: dar métodos para obtener estimadores puntuales y obtener criterios para encontrar "buenos"estimadores.

8.2. Métodos de estimación

En algunos casos, pueden encontrarse estimadores razonables de los parámetros con base en la intuición, por ejemplo con su análogo muestral y pueden ser "buenos" estimadores; la media muestral es un "buen"estimador de la media poblacional μ y la varianza muestral es un "buen"estimador de la varianza poblacional σ^2. Sin embargo este método intuitivo no es lo más común y por eso métodos generales han sido desarrollados para obtener estimadores.

8.2.1. Método de momentos

Este es quizás el método más antiguo para obtener estimadores puntuales y tiene la virtud que es simple y casi siempre produce algún tipo de estimador.

Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población con función de probabilidad $f(x; \theta_1, \theta_2, ..., \theta_k)$. Los momentos alrededor del origen μ'_j, generalmente dependen de los parámetros, es decir, $E(X^j) = \mu'_j(\theta_1, \theta_2, ..., \theta_k)$, $j = 1, 2, ..., k$. Los momentos muestrales alrededor del origen, M'_j dependen de la muestra, pues $M'_j = \frac{\Sigma X_i^j}{n}$.

El principio del método establece que se escoja como estimadores de los parámetros $\theta_1, \theta_2, ..., \theta_k$ los valores $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k$ que son soluciones de las ecuaciones

$$M'_1 = \mu'_1,\ M'_2 = \mu'_2, \cdots,\ M'_k = \mu'_k. \tag{8.1}$$

Ejemplo 8.1. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x : p) = p^x(1-p)^{1-x}$, $x = 0, 1$. Como $E(X) = p$ y $M'_1 = \frac{\Sigma X_i}{n}$, al resolver $\mu'_1 = M'_1$ se tiene que $\hat{p} = \frac{\Sigma X_i}{n} = \bar{X}$, pues $\mu'_1 = E(X)$. Luego se dice que $\hat{P} = \hat{X}$ es el estimador por el método de momentos de p y $\hat{p} = \hat{x}$ es la estimación por este método.

Ejemplo 8.2. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x : \theta) = \theta e^{-\theta x} I_{(0,\infty)}$, $\theta > 0$. Como $E(X) = \frac{1}{\theta}$ y $M_1' = \frac{\Sigma X_i}{n}$, entonces de $E(X) = M_1'$ se tiene que $\hat{\Theta} = \frac{n}{\Sigma X_i} = \dfrac{1}{\bar{X}}$ es el estimador de θ por el método de momentos.

Ejemplo 8.3. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x : \mu, \sigma^2)$, donde μ y σ^2 son la media y la varianza de la población (desconocidas). Del capítulo 3 se sabe que $\mu_1' = E(X) = \mu$ y que $\mu_2' = E(X^2) = \sigma^2 + \mu^2$, así que por el método de momentos las soluciones de $M_1' = \mu_1'$ y $M_2' = \mu_2'$ son $\hat{\mu} = \bar{X}$ y $\hat{\sigma}^2 = \frac{\Sigma X_i}{n} - \bar{X}^2 = \frac{n-1}{n} S^2$ y son los estimadores por el método de momentos de μ y σ^2 respectivamente.

Ejemplo 8.4. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu, \sigma^2)$. Del capítulo 3 se conoce que $E(X) = \mu$ y $V(X) = \sigma^2$, así es que los estimadores por el método de momentos son los dados en el ejemplo anterior.

Ejemplo 8.5. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim G(r, \lambda)$. Como se conoce que $E(X) = \frac{r}{\lambda}$ y $V(X) = \frac{r}{\lambda^2}$, luego $E(X^2) = \frac{r}{\lambda^2} + \frac{r^2}{\lambda^2}$. De $M_1' = \mu_1'$ y $M_2' = \mu_2'$ se tiene que $\frac{r}{\lambda} = \frac{\Sigma X_i}{n}$ y $\frac{r+r^2}{\lambda^2} = \frac{\Sigma X_i^2}{n}$. De esta manera se concluye que $\hat{r} = \frac{n\bar{X}}{(n-1)S^2}$ y $\hat{\lambda} = \frac{n}{(n-1)} S^2$ son los estimadores de r y λ, respectivamente por el método de momentos.

Ejemplo 8.6. Sea $X_1, X_2, ..., X_m$ una muestra aleatoria de $X \sim B(n, p)$. Como $E(X) = np$ y $E(X^2) = np(1-p) + (np)^2$, entonces de $M_1' = \mu_1'$ y $M_2' = \mu_2'$ se tiene que, $np = \frac{\Sigma X_i}{m}$ y $np(1 - p + np) = \frac{\Sigma X_i^2}{m}$. Después de algunos cálculos fáciles se llega a que un estimador por el método de momentos de n es $\hat{N} = \frac{\bar{X}^2}{\bar{X} - \frac{m-1}{m} S^2}$ y el de p es $\hat{P} = \frac{\bar{X}}{\hat{N}} = \frac{\bar{X} - \frac{m-1}{m} S^2}{\bar{X}}$.

Este ejemplo muestra que el método de momentos no proporciona los mejores estimadores, pues es posible obtener estimaciones negativas de n y p cuando se sabe que por definición deben ser positivos.

8.2.2. Método de máxima verosimilitud

Se considera ahora un método que con mucha frecuencia conduce a estimadores que poseen propiedades deseables, particularmente propiedades para grandes muestras. La idea es usar un valor en el espacio de parámetros Θ como un estimador del parámetro desconocido, que corresponda a la más alta probabilidad de los valores observados. Para introducir el método observe el siguiente problema sencillo de estimación.

Suponga que una moneda es sesgada y que la probabilidad de obtener *cara* es uno de los tres valores 1/4, 2/5 o 3/4 . Un experimento consiste en lanzar la moneda n veces y obsevar el número de caras; la distribución del número de caras X es binomial con parámetro n y p, donde el espacio de parámetros es Θ={ 1/4, 2/5, 3/4}. Note que el estimador por el método de momentos de p, el cual es $\bar{X}$ no produce estimaciones razonables, pues $\bar{x} = 0$, 0.5, 1 y éstos no son valores de Θ.

Si la moneda se lanza dos veces y se intenta estimar el parámetro desconocido p, se analizan los valores de X y sus respectivas probabilidades. Aquí se tiene en cuenta que

$$f(x;p) = \binom{2}{x} p^x (1-p)^{2-x}, \text{ para } \ x = 0, 1, 2.$$

Los valores de f para $x = 0, 1, 2$ se dan en la tabla 8.1 y para los diferentes p.

	x		
p	0	1	2
$\frac{1}{4}$	$\frac{9}{16}$	$\frac{6}{16}$	$\frac{1}{16}$
$\frac{2}{5}$	$\frac{9}{25}$	$\frac{12}{25}$	$\frac{4}{25}$
$\frac{3}{4}$	$\frac{1}{16}$	$\frac{6}{16}$	$\frac{9}{16}$

Tabla 8.1: fmp conjunta del número de caras al lanzar una moneda sesgada dos veces

Si al realizar el experimento se observa que $x = 0$, parece más probable $p = \frac{1}{4}$ en vez de los otros dos valores. Similarmente, si $x = 1$ parece más probable $p = \frac{2}{5}$ y si $x = 2$ es más probable $p = \frac{3}{4}$. De manera que las estimaciones que maximizan la fmp conjunta es

$$\hat{p} = \begin{cases} \frac{1}{4}, & \text{si } x = 0 \\ \frac{2}{5}, & \text{si } x = 1 \\ \frac{3}{4}, & \text{si } x = 2. \end{cases}$$

En general, para las variables aleatorias discretas $X_1, X_2, ..., X_n$, la función de masa de probabilidad conjunta f evaluada en un conjunto particular de datos muestrales

$x_1, x_2, ..., x_n$, $f(x_1, x_2, ..., x_n; \theta)$ representa la probabilidad que el conjunto observado $x_1, x_2, ..., x_n$ ocurra. Para variables aleatorias continuas, $f(x_1, x_2, ..., x_n; \theta)$ no es una probabilidad, pero aún refleja la posibilidad relativa de que el conjunto de datos se produzca, y esta posibilidad depende del valor verdadero del parámetro.

Definición 8.2. Función de verosimilitud. La función de masa o de densidad de probabilidad conjunta de $X_1, X_2, ..., X_n$ evaluada en $x_1, x_2, ..., x_n$, $f(x_1, x_2, ..., x_n; \theta)$ se conoce como la función de verosimilitud (fdv). Para $x_1, x_2, ..., x_n$ fijos la fdv es una función de θ y a menudo se denota por $L(\theta)$. Si $X_1, X_2, ..., X_n$ representa una muestra aleatoria de $f(x; \theta)$, la fdv es

$$L(\theta) = f(x_1; \theta) f(x_2; \theta) ... f(x_n; \theta). \tag{8.2}$$

Definición 8.3. Estimador de máxima verosimilitud. Sea $L(\theta)$ la fdv de $X_1, X_2, ..., X_n$, donde $\theta \in \Theta$. Para un conjunto dado de observaciones $x_1, x_2, ..., x_n$, un valor $\hat{\theta}$ en Θ que maximiza a $L(\theta)$ se llama una estimación de máxima verosimilitud de θ. Es decir, una estimación de máxima verosimilitud $\hat{\theta}$, es un valor de θ que satisface

$$f(x_1, x_2, ..., x_n; \hat{\theta}) = \max_{\theta \in \Theta} f(x_1, x_2, ..., x_n; \theta). \tag{8.3}$$

Nótese que:

a) Si cada conjunto de observaciones $x_1, x_2, ..., x_n$ corresponde a un valor único $\hat{\theta}$, entonces este procedimiento define una función, $\hat{\theta} = t(x_1, x_2, ..., x_n)$, que es la **estimación de máxima verosimilitud**. Esta misma función cuando se aplica a la muestra produce $\hat{\Theta} = t(X_1, X_2, ..., X_n)$ y se llama el **estimador de máxima verosimilitud** (MLE) de θ.

b) El rango del MLE coincide con el del parámetro.

c) Si el espacio de parámetros Θ, es un intervalo, y si $L(\theta)$ es diferenciable y alcanza un máximo en Θ, entonces el MLE será una solución de la ecuación

$$\frac{d}{d\theta} L(\theta) = 0. \tag{8.4}$$

En sentido estricto, si una o más soluciones de esta ecuación existen, entonces debe verificarse cuál, si existe, maximiza a $L(\theta)$. Notese además que cualquier valor de θ que maximiza a $L(\theta)$ tambien maximizará al $\ln L(\theta)$, puesto que la función logaritmo natural es estrictamente creciente en $(0, \infty)$, lo que implica que los extremos de $L(\theta)$ y $\ln L(\theta)$ coinciden. Esta última es más fácil de derivar y por eso se obtendrá el MLE resolviendo,

$$\frac{d}{d\theta} \ln L(\theta) = 0 \tag{8.5}$$

en lugar de la primera ecuación.

Ejemplo 8.7. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim B(1,p)$, se desea obtener un MLE de p.

Solución. Como $f(x;p) = p^x(1-p)^{1-x}I_{\{0,1\}}(x)$, la función de verosimilitud de p es,

$$L(p) = \Pi_{i=1}^n f(x_i;p) = \Pi_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\Sigma_{i=1}^n x_i}.$$

De manera que,

$$\ln L(p) = (\Sigma_{i=1}^n x_i)\ln p + (n - \Sigma_{i=1}^n x_i)\ln(1-p).$$

De $\frac{d}{d\theta}\ln L(p) = 0$ se tiene que $\frac{\Sigma_{i=1}^n x_i}{p} - \frac{n-\Sigma_{i=1}^n x_i}{1-p} = 0$. Así que $\hat{p} = \frac{\Sigma_{i=1}^n x_i}{n} = \bar{x}$ es la estimación de máxima verosimilitud para p y el MLE es $\bar{X}$.

Observación En lo que sigue y mientras no se diga lo contrario, Σ denota una suma de 1 a n y Π denota un producto de 1 a n.

Ejemplo 8.8. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim P(\lambda)$, se desea hallar un MLE de λ.

Solución. Como $f(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}I_{\{0,1,\cdots\}}(x)$ entonces $L(\lambda) = \Pi\frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\Sigma x_i}}{\Pi x_i!}$ y $\ln L(\lambda) = -n\lambda + \Sigma x_i(\ln\lambda) - \ln(\Pi x_i!)$.

De $\frac{d}{d\lambda}\ln L(\lambda) = 0$ se tiene que $-n + \frac{\Sigma x_i}{\lambda} = 0$. O sea que $\hat{\lambda} = \frac{\Sigma x_i}{n} = \bar{x}$ es la estimación de máxima verosimilitud de λ. El MLE es $\bar{X}$.

Ejemplo 8.9. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim E(\theta)$, se desea obtener un MLE de θ.

Solución. Como $f(x;\theta) = \theta e^{-\theta x}I_{(0,\infty)}(x)$ entonces,

$$L(\theta) = \Pi\theta e^{-\theta x_i} = \theta^n e^{-\theta\Sigma x_i} \quad \text{y} \quad \ln L(\theta) = n\ln\theta - \theta\Sigma x_i.$$

De $\frac{d}{d\theta}\ln L(\theta) = 0$ se tiene que $\frac{n}{\theta} - \Sigma x_i = 0$, y por tanto $\hat{\theta} = \frac{n}{\Sigma x_i} = \frac{1}{\bar{x}}$ es la estimación de máxima verosimilitud para θ. El MLE es $\dfrac{1}{\bar{X}}$.

Supóngase que ahora se desea obtener el MLE de $\tau = \tau(\theta) = P(X > 1) = e^{-\theta}$. Entonces, $\theta = -\ln\tau$ y la fdv es

$$L^*(\tau) = (-\ln\tau)^n e^{\ln\tau\Sigma x_i} \text{ y } l^* = \ln L^* = n\ln(-\ln\tau) + (\ln\tau)\Sigma x_i.$$

De $\frac{d}{d\tau}l^*(\tau) = 0$ se tiene que $\frac{n}{\tau \ln \tau} + \frac{1}{\tau}\Sigma x_i = 0$, es decir, $\frac{n}{-\Sigma x_i} = \ln \tau$.

Así que $\hat{\tau} = e^{-1/\bar{x}}$ es la estimación MLE de $\tau(\theta) = e^{-\theta}$.

En este ejemplo se obseva que $\hat{\tau} = \hat{\tau}(\theta) = \tau(\hat{\theta})$. Esta es la propiedad de invarianza.

Teorema 8.1. Propiedad de invarianza Si $\hat{\theta}$ es el MLE de θ entonces, para cualquier función $\tau(\theta)$, el MLE de $\tau(\theta)$ es $\tau(\hat{\theta})$.

Demostración. Sea $\hat{\theta}$ el MLE de θ y sea $\eta = \tau(\theta)$

- Si τ es uno a uno, $\theta = \tau^{-1}(\eta)$ y la fdv de $\tau(\theta)$, inducida por L está dada por

$$L^*(\eta) = \Pi f(x_i; \tau^{-1}(\eta)) = L[\tau^{-1}(\eta)]$$

y

$$\sup_{\eta} L^*(\eta) = \sup_{\eta} L[\tau^{-1}(\eta)] = \sup_{\theta} L(\theta)$$

Luego el máximo de $L^*(\eta)$ se alcanza en el máximo de $L(\theta)$ y como $\hat{\theta}$ es el MLE de θ, entonces $\hat{\eta} = \tau(\hat{\theta})$ es el MLE de $\tau(\theta)$.

- Si τ no es uno a uno, para un valor dado de η puede haber más de un valor de θ que satisface $\tau(\theta) = \eta$

Sea L^* la función de verosimilitud inducida por $\tau(\theta)$ y definida por

$$L^*(\eta) = \sup_{\{\theta:\tau(\theta)=\eta\}} L(\theta)$$

El valor $\hat{\eta}$ que maximiza $L^*(\eta)$ es el MLE de $\eta = \tau(\theta)$, y puede verse que el máximo de L^* y L coinciden. Sea $\hat{\eta}$ el valor de η que maximiza L^*. Se debe demostrar que $L^*(\hat{\eta}) = L^*[\tau(\hat{\theta})]$.

En efecto,

$$L^*(\hat{\eta}) = \sup_{\eta}\{\sup_{\{\theta:\tau(\theta)=\eta\}} L(\theta)\} = \sup_{\theta} L(\theta) = L(\hat{\theta}).$$

Además

$$L(\hat{\theta}) = \sup_{\{\theta:\tau(\theta)=\tau(\hat{\theta})\}} L(\theta) = L^*[\tau(\hat{\theta})].$$

Estas dos expresiones muestran que $L^*(\hat{\eta}) = L^*[\tau(\hat{\theta})]$ y así $\tau(\hat{\theta})$ es el MLE de $\tau(\theta)$.

Por esta propiedad, del ejemplo 8.7 se sabe que $\bar{X}$ es el MLE de p cuando $X \sim B(1,p)$, luego $\bar{X}(1-\bar{X})$ es el MLE de $V(X) = p(1-p)$. En el ejemplo 8.8 se encontró que $\bar{X}$ es el MLE de λ, cuando $X \sim P(\lambda)$, luego el MLE de $P(X>0) = 1-e^{-\lambda}$ es $1-e^{-\bar{X}}$.

Hasta ahora se han considerado distribuciones caracterizadas por un solo parámetro, sin embargo, el espacio de parámetros Θ puede ser un sistema k-dimensional, o un producto cartesiano de k intervalos. Cuando Θ es de esta forma, las derivadas parciales de $L(\theta_1, \theta_2, ..., \theta_k)$ existen, y si los MLE no se producen en la frontera de Θ, entonces los MLE serán soluciones de las ecuaciones

$$\frac{\partial}{\partial \theta_i} L(\theta_1, \theta_2, ..., \theta_k) = 0, \ i = 1, 2, ..., k. \tag{8.6}$$

En este caso también puede ser más fácil trabajar con

$$\frac{\partial}{\partial \theta_i} \ln L(\theta_1, \theta_2, ..., \theta_k) = 0, \ i = 1, 2, ..., k. \tag{8.7}$$

Ambos sistemas se llaman las ecuaciones de máxima verosimilitud y las soluciones se denotan por $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k$.

Ejemplo 8.10. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu, \sigma^2)$, se desean obtener los MLE de μ y σ^2.

Solución. Como $f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ entonces,

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{\frac{1}{2\sigma^2}\sum(x_i-\mu)^2}$$

y

$$\ln L(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{\sum(x_i-\mu)^2}{2\sigma^2}$$

luego,

$$\frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2) = \frac{\sum(x_i-\mu)}{\sigma^2}$$

y

$$\frac{\partial}{\partial \sigma^2} \ln L(\mu; \sigma^2) = -\frac{n}{2\sigma^2} + \frac{\sum(x_i-\mu)^2}{2\sigma^4}.$$

Las ecuaciones de máxima verosimilitud son

$$\frac{\sum(x_i-\mu)}{\sigma^2} = 0 \ \text{ y } \ \frac{\sum(x_i-\mu)^2}{2\sigma^4} - \frac{n}{2\sigma^2} = 0.$$

De esta manera, $\sum x_i = n\mu$ y $\sum(x_i - \mu)^2 = n\sigma^2$ y las estimaciones de máxima verosimilitud de μ y σ son $\hat{\mu} = \bar{x}$ y $\hat{\sigma^2} = \frac{\sum(x_i-\bar{x})^2}{n}$.

Ejemplo 8.11. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x; \lambda, \theta) = \lambda e^{-\lambda(x-\theta)}$ para $x \geq \theta, \theta > 0$ y $\lambda > 0$ (distribución exponencial con tiempo de garantía θ). Se desean obtener los MLE de λ y θ.

Solución. La fdv es

$$L(\lambda, \theta) = \lambda^n e^{-\lambda \sum(x_i-\theta)}, \text{ para todo } \ x_i \geq \theta$$

y

$$\ln L(\lambda, \theta) = n \ln \lambda - \lambda\Sigma(x_i - \theta), \text{ para todo } \ x_i \geq \theta.$$

Luego

$$\frac{\partial}{\partial\lambda} \ln L(\lambda, \theta) = \frac{n}{\lambda} - \Sigma(x_i - \theta)$$

y

$$\frac{\partial}{\partial\theta} \ln L(\lambda, \theta) = n\lambda$$

Al resolver las ecuaciones de verosimilitud se encuentra que $\frac{n}{\lambda} = \Sigma x_i - n\theta$ y $n\lambda = 0$. Pero $n\lambda$ no puede ser 0 porque $n > 0$ y $\lambda > 0$. Sin embargo, $L(\lambda; \theta)$ es máxima respecto a θ si $\hat{\theta} = x_{(1)}$, el mínimo de las $x_i, i = 1, 2, ..., n$. Así es que de $\frac{n}{\lambda} = \Sigma x_i - nx_{(1)}$ resulta $\hat{\lambda} = \frac{n}{\sum x_i - nx_{(1)}}$. Por tanto, el MLE de θ es $\hat{\Theta} = X_{(1)}$ y el de λ es $\hat{\Lambda} = \dfrac{1}{(\bar{X} - X_{(1)})}$.

Ejemplo 8.12. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim G(r, \lambda)$. Obtenga los MLE de r y λ.

Solución. Si $X \sim G(r, \lambda)$ entonces,

$$f(x; r, \lambda) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} I_{(0,\infty)}(x)$$

y

$$L(r, \lambda) = \frac{\lambda^{nr}}{[\Gamma(r)]^n} (\Pi x_i)^{r-1} e^{-\lambda \Sigma x_i}.$$

Luego,

$$\ln L(r, \lambda) = nr \ln \lambda - n \ln \Gamma(r) + (r - 1) \ln \Pi x_i - \lambda \Sigma x_i.$$

De las ecuaciones de verosimilitud,

$$\frac{nr}{\lambda} - \Sigma x_i = 0 \text{ y , } n \ln \lambda - \frac{n}{\Gamma(r)} \Gamma'(r) + \ln \Pi x_i = 0.$$

Así que,

$$\hat{\lambda} = \frac{\hat{r}}{\bar{x}} \text{ y , } \frac{\Gamma'(r)}{\Gamma(r)} - \ln \frac{\hat{r}}{\bar{x}} - \frac{1}{n} \ln \Pi x_i = 0.$$

Como $\ln \frac{\hat{r}}{\bar{x}} = \ln \hat{r} - \ln \bar{x}$ y, $\frac{1}{n} \ln \Pi x_i = \ln(\Pi x_i)^{\frac{1}{n}}$, sea $\widehat{x} = (\Pi x_i)^{\frac{1}{n}}$ que es la media geométrica entonces,

$$\hat{\lambda} = \frac{\hat{r}}{\bar{x}} \text{ y , } \frac{\Gamma'(r)}{\Gamma(r)} - \ln \hat{r} + \ln \bar{x} - \ln \widehat{x} = 0.$$

Esto significa que,

$$\hat{\lambda} = \frac{\hat{r}}{\bar{x}} \text{ y , } \frac{\Gamma'(r)}{\Gamma(r)} - \ln \hat{r} + \ln \frac{\bar{x}}{\widehat{x}} = 0.$$

Sea $\psi(t) = \frac{\Gamma'(t)}{\Gamma(t)}$ entonces,

$$\hat{\lambda} = \frac{\hat{r}}{\bar{x}}.$$

$$\psi(r) - \ln \hat{r} + \ln \frac{\bar{x}}{\widehat{x}} = 0.$$

Esta última ecuación no tiene solución algebraica sin embargo una solución numérica puede obtenerse. En 1960 Greenwood y Duran [9] encontraron para $y = \ln \frac{\bar{x}}{\widehat{x}}$ que

$$\hat{r} = \frac{0.5000876 + 0.1648852y - 0.0544274y^2}{y}, \quad 0 < y \leq 0.5772$$

$$\hat{r} = \frac{8.898919 + 9.059950y + 0.9775373y^2}{(17.79728 + 11.968477y + y^2)y}, \quad 0.5772 < y \leq 17$$

$$\hat{r} = \frac{1}{y}, \qquad y > 17.$$

Teorema 8.2. Propiedad de invarianza Si $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k)$ denota el MLE de $\theta = (\theta_1, \theta_2, ..., \theta_k)$, entonces el MLE de $\tau = (\tau_1(\theta), ..., \tau_r(\theta))$ es $\hat{\tau} = (\tau_1(\hat{\theta}), ..., \tau_r(\hat{\theta}))$.

Para probar el teorema se procede como en el caso de un parámetro. Si τ representa una transformación *uno a uno*, entonces una reparametrizacion de la función de verosimilitud se puede definir, y el MLE de τ se obtiene como la transformación del MLE de θ. En el caso de una transformación que no es *uno a uno*, la función de verosimilitud relativa a τ debe extenderse de manera similar al caso de un solo parámetro.

En el ejemplo 8.10 si $\tau = \mu + z_\alpha \sigma$, donde $0 < \alpha < 1$, entonces el MLE de τ es $\hat{\tau} = \bar{X} + z_\alpha \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{n}}$.

Teorema 8.3. Un MLE (o un conjunto de MLEs) de $\tau(\theta)$ depende de la muestra solo a través de un (o un conjunto de) estadístico(s) suficiente(s).

Demostración. Sea S un estadístico suficiente para θ, luego

$$L(\theta) = f(x_1, x_2, ..., x_n; \theta) = g(s; \theta) h(x_1, x_2, ..., x_n).$$

De $\frac{d}{d\theta} L(\theta) = 0$ se tiene que $\frac{d}{d\theta} g(s; \theta) = 0$ y si T es el MLE de $\tau(\theta)$ entonces T depende de la muestra solo a través de S.

Los ejemplos anteriores respaldan el teorema.

8.3. Criterios para evaluar estimadores

Los métodos presentados anteriormente son técnicas razonables para encontrar estimadores puntuales. Sin embargo, algunas veces métodos diferentes producen estimadores distintos y es una preocupación discernir acerca de cuales de estos métodos, producen *mejores* estimadores. En esta sección se estudian varias propiedades deseables.

8.3.1. Cercanía y error cuadrático medio

Se considera inicialmente la calidad de un estimador, para muestras finitas, con el error cuadrático medio.

Definición 8.4. Error cuadrático medio Sea $X_1, X_2, ..., X_n$ una muestra de $f(x; \theta)$ y sea $T = t(X_1, X_2, ..., X_n)$ un estimador de $\tau(\theta)$. El *error cuadrático medio*(MSE) de T se define por

$$MSE_T(\theta) = E[T - \tau(\theta)]^2. \tag{8.8}$$

El error cuadrático medio mide el promedio de las diferencias al cuadrado entre el estimador y $\tau(\theta)$, y tiene al menos dos ventajas sobre otras medidas de distancia como son: es manejable asintóticamente y tiene la interpretación que

$$MSE_T(\theta) = V(T) + [b(T)]^2, \tag{8.9}$$

donde $b(T) = E(T) - \tau(\theta)$.

En efecto,

$$MSE_T(\theta) = E\big\{[T - E(T)] + [E(T) - \tau(\theta)]\big\}^2,$$

es decir que,

$$MSE_T(\theta) = E[T - E(T)]^2 + 2E\{[T - E(T)][E(T) - \tau(\theta)]\} + E[E(T) - \tau(\theta)]^2.$$

Luego

$$MSE_T(\theta) = V(T) + [E(T) - \tau(\theta)]^2 = V(T) + [b(T)]^2,$$

donde $b(T) = E(T) - \tau(\theta)$ se llama el *sesgo* del estimador.

Definición 8.5. Sesgo y Estimador insesgado El *sesgo*, $b(T)$ de un estimador $T = t(X_1, X_2, ..., X_n)$ de $\tau(\theta)$, es la diferencia entre el valor esperado de T y $\tau(\theta)$. Un estimador cuyo sesgo es identicamente 0 se dice es *insesgado*, y satisface

$$E(T) = \tau(\theta), \text{para todo } \theta \in \Theta.$$

El error cuadrático medio incorpora dos componentes importantes, una que mide la variabilidad (precisión) y la otra mide el sesgo (exactitud). Un estimador que tiene buenas propiedades en MSE tiene varianza y sesgo pequeños. Para encontrar un estimador con estas características, se necesita hallar estimadores que controlen ambos, varianza y sesgo. Los estimadores insesgados hacen un buen trabajo de controlar el sesgo pues su error cuadrático medio es su varianza; es decir, si T es un estimador insesgado de $\tau(\theta)$ entonces,

$$MSE_t(\theta) = V(T). \tag{8.10}$$

Si un estimador insesgado T se usa para asignar un valor de $\tau(\theta)$, entonces el valor real de $\tau(\theta)$ es posible que no se alcance para cualquier estimación t, pero el promedio de estas estimaciones será $\tau(\theta)$.

Ejemplo 8.13. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x; \mu, \sigma^2)$, donde μ y σ^2 son la media y la varianza, respectivamente, de la población. Se demostró en el capítulo anterior que $E(\bar{X}) = \mu$ y $E(S^2) = \sigma^2$, luego $\bar{X}$ es un estimador insesgado de μ y S^2 lo es de σ^2.

Si μ y σ^2 son desconocidos $\Theta = \{(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\} \subset R^2$, es el espacio de parámetros. Si solo uno de los parámetros es conocido, entonces $\Theta \subset R$. Por ejemplo si la población es normal con media $\mu = 4$, $\Theta = \{\sigma^2 : \sigma^2 > 0\}$.

Ejemplo 8.14. Si $X_1, X_2, ..., X_n$ es una muestra aleatoria de $X \sim N(\mu, \sigma^2)$ entonces,
-$\bar{X}$ es un estimador insegado de μ y $MSE_{\bar{X}}(\mu) = V(\bar{X}) = \frac{\sigma^2}{n}$.
-S^2 es un estimador insesgado de σ^2 y $MSE_{S^2}(\sigma^2) = V(S^2) = \frac{2\sigma^4}{n-1}$.

El MSE de $\bar{X}$ sigue siendo $\frac{\sigma^2}{n}$ aunque el supuesto de normalidad se cambie, sin embargo, el de S^2 puede no ser el mismo.

Ejemplo 8.15. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim P(\lambda)$. Se conoce que la media y la varianza de X es λ, y como $E(\bar{X}) = \lambda$ y $E(S^2) = \lambda$, entonces $\bar{X}$ y S^2 son estimadores insesgados de λ. Luego,

$$MSE_{\bar{X}}(\lambda) = V(\bar{X}) = \frac{\lambda}{n}$$

y

$$MSE_{S^2}(\lambda) = V(S^2) = \frac{1}{n}\big(\mu_4 - \frac{n-3}{n-1}\sigma^4\big) = \frac{\lambda}{n} + \frac{2\lambda^2}{n-1}.$$

Estos resultados sugieren que $\bar{X}$ es *mejor* estimador de λ que S^2 bajo el criterio del error cuadrático medio.

Nota
Aunque muchos estimadores insesgados son razonables desde el punto de vista del error cuadrático medio, se debe tener en cuenta que controlando el sesgo no garantiza que el MSE esté controlado. En particular, a veces es el caso que al producir un equilibrio entre la varianza y el sesgo resulte una mejora en el MSE, de manera que un pequeño aumento en el sesgo pueda ser objeto de una mayor disminución en la varianza.

Ejemplo 8.16. Se encontró que el MLE de σ^2 cuando $X_1, X_2, ..., X_n$ es una muestra aleatoria de $X \sim N(\mu, \sigma^2)$ es $\hat{\sigma}^2 = \frac{1}{n}\Sigma(X_i - \bar{X})^2$. Este estimador es tal que,

$$E(\hat{\sigma}^2) = E\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}\sigma^2 \ , \ b(\hat{\sigma}^2) = -\frac{\sigma^2}{n}$$

y

$$V(\hat{\sigma}^2) = \left(\frac{n-1}{n}\right)^2 V(S^2) = \left(\frac{n-1}{n}\right)^2 \frac{2\sigma^4}{n-1} = \frac{2(n-1)}{n^2}\sigma^4.$$

Luego,

$$MSE_{\hat{\sigma}^2}(\sigma^2) = V(\hat{\sigma}^2) + [b(\hat{\sigma}^2)]^2 = \frac{2n-1}{n^2}\sigma^4.$$

Obsérvese ahora que

$$\frac{2n-1}{n^2} - \frac{2}{n-1} = -\frac{3n-1}{n^2(n-1)} < 0, \text{ para } \ n \geq 2.$$

De manera que,

$$\frac{2n-1}{n}\sigma^4 < \frac{2}{n-1}\sigma^4.$$

De esta expresión y del ejemplo 8.14 se concluye que

$$MSE_{\hat{\sigma}^2}(\sigma^2) < MSE_{S^2}(\sigma^2).$$

Así que $\hat{\sigma}^2$ es un estimador sesgado que es mejor que el estimador insesgado S^2 bajo el criterio del error cuadrático medio.

Otra idea que se emplea para determinar cuál es mejor, entre dos estimadores de un parámetro, es seleccionar aquel estimador que tienda a estar *más cerca* o *más concentrado* alrededor del verdadero valor del parámetro.

Definición 8.6. Estimador más concentrado y el más concentrado. Sean T_1 y T_2 dos estimadores de $\tau(\theta)$. Se dice que T_1 es un estimador más concentrado que T_2 de $\tau(\theta)$ si

$$P[\tau(\theta) - \epsilon < T_1 < \tau(\theta) + \epsilon] \geq P[\tau(\theta) - \epsilon < T_2 < \tau(\theta) + \epsilon] \tag{8.11}$$

para todo $\epsilon > 0$.

Un estimador T^* de $\tau(\theta)$ se dice *el más concentrado* si es más concentrado que cualquier otro estimador.

No es claro cómo obtener un estimador que sea **el más concentrado**, pero algunos criterios pueden ayudar a tomar decisiones. Por ejemplo, si T es un estimador insesgado de $\tau(\theta)$, por la desigualdad de Chebyshev se sigue que,

$$P[\tau(\theta) - \epsilon < T < \tau(\theta) + \epsilon] \geq 1 - \frac{V(T)}{\epsilon^2}, \text{ para todo } \epsilon > 0$$

Esto sugiere que para estimadores insesgados, aquel que tenga la varianza más pequeña tenderá a ser más concentrado y por tanto puede ser el preferido. En el ejemplo 8.15 , como $V(\bar{X}) = \frac{\lambda}{n}$ y $V(S^2) = \frac{\lambda}{n} + \frac{2\lambda^2}{n-1}$, entonces $\bar{X}$ es mejor estimador que S^2 bajo este criterio.

8.3.2. Estimadores insesgados de varianza mínima

En algunos casos un estimador T puede tener una varianza más pequeña que T^* para algunos valores de θ, pero también una varianza mayor que T^* para otros valores de θ. En tales casos ningún estimador puede decirse que es mejor estimador que el otro. En ciertos casos es posible demostrar que un estimador insesgado tiene la varianza más pequeña entre todos los estimadores posibles para todos los valores posibles de θ. Este puede ser un buen estimador.

Definición 8.7. Estimador insesgado de varianza mínima uniforme. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$. Un estimador T^* de $\tau(\theta)$ se dice que es un *estimador insesgado de varianza mínima uniforme*(UMVUE) si:
1. T^* es un estimador insesgado de $\tau(\theta)$.
2. Para cualquier otro estimador insesgado T de $\tau(\theta)$, $V(T^*) \leq V(T)$, para todo $\theta \in \Theta$.

En algunos casos, un acotamiento inferior puede obtenerse para la varianza de los estimadores insesgados. Si la varianza de un estimador insesgado alcanza tal acotamiento inferior, entonces ese estimador es un UMVUE.

Teorema 8.4. Desigualdad de Cramer-Rao. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$ y sea $T = t(X_1, X_2, ..., X_n)$ un estimador insesgado de $\tau(\theta)$. Si se satisfacen:

- $\frac{\partial}{\partial\theta} \ln f(x;\theta)$ existe para todo x y todo θ.
- $\frac{\partial}{\partial\theta} E(T) = \int \cdots \int \frac{\partial}{\partial\theta} t(x_1, x_2, ..., x_n) f(x_1, x_2, ..., x_n; \theta) dx_1 ... dx_n$.
- $\frac{\partial}{\partial\theta} \int ... \int f(x_1, x_2, ..., x_n; \theta) dx_1 \cdots dx_n = \int ... \int \frac{\partial}{\partial\theta} f(x_1, x_2, ..., x_n; \theta) dx_1 \cdots dx_n$.

Entonces,

$$V(T) \geq \frac{\left[\tau'(\theta)\right]^2}{nE\left[\frac{\partial}{\partial\theta} \ln f(x;\theta)\right]^2}. \tag{8.12}$$

La igualdad se da si y solo si existe una función, $k(\theta, n)$ tal que

$$\sum \frac{\partial}{\partial\theta} \ln f(x_i;\theta) = k(\theta,n)[t(x_1,x_2,...,x_n) - \tau(\theta)].$$

En este caso $T = t(X_1, X_2, ..., X_n)$ es un UMVUE.

Demostración. Sea $h(x_1, x_2, ..., x_n; \theta) = \frac{\partial}{\partial\theta} \ln f(x_1, x_2, ..., x_n; \theta)$, entonces

$$h(x_1,x_2,...,x_n;\theta) = \frac{1}{f(x_1,x_2,...,x_n;\theta)} \frac{\partial}{\partial\theta} f(x_1,x_2,...,x_n;\theta).$$

Al definir $H = h(X_1, X_2, ..., X_n; \theta)$ entonces,

$$\begin{aligned} E(H) &= \int ... \int h(x_1,x_2,...,x_n;\theta) f(x_1,x_2,...,x_n;\theta) dx_1...dx_n \\ &= \int ... \int \frac{\partial}{\partial\theta} f(x_1,x_2,...,x_n;\theta) dx_1...dx_n \\ &= \frac{\partial}{\partial\theta} \int ... \int f(x_1,x_2,...,x_n;\theta) dx_1...dx_n = \frac{\partial}{\partial\theta} 1 = 0. \end{aligned}$$

Como T es un estimador insesgado de $\tau(\theta)$ entonces,

$$\tau(\theta) = E(T) = \int ... \int t(x_1,x_2,...,x_n) f(x_1,x_2,...,x_n;\theta) dx_1...dx_n,$$

luego,

$$\begin{aligned} \tau'(\theta) = \frac{\partial}{\partial\theta}\tau(\theta) &= \int ... \int t(x_1,x_2,...,x_n) \frac{\partial}{\partial(\theta)} f(x_1,x_2,...,x_n;\theta) \Pi dx_i \\ &= \int ... \int t(x_n,x_2,...,x_n) h(x_1,x_2,...,x_n;\theta) f(x_1,x_2,...,x_n;\theta) \Pi dx_i \\ &= E(TH). \end{aligned}$$

De esta manera, $V(H) = E(H^2)$ y $cov(T, H) = E(TH)$. Además, como el coeficiente de correlación ρ, satisface $|\rho| \leq 1$, entonces

$$\frac{[Cov(T,H)]^2}{V(T)V(H)} \leq 1.$$

y en consecuencia,

$$V(T) \geq \frac{[E(TH)]^2}{V(H)} = \frac{[\tau'(\theta)]^2}{E\big[\frac{\partial}{\partial\theta} \ln f(X_1,X_2,...,X_n;\theta)\big]^2}. \tag{8.13}$$

Como $X_1, X_2, ..., X_n$ es una muestra aleatoria

$$f(x_1, x_2, ..., x_n; \theta) = f(x_1; \theta) f(x_2; \theta) ... f(x_n; \theta),$$

luego

$$E\left[\frac{\partial}{\partial\theta} \ln f(X_1, X_2, ..., X_n; \theta)\right]^2 = E\left[\sum \frac{\partial}{\partial\theta} \ln f(X_i; \theta)\right]^2 = nE\left[\frac{\partial}{\partial\theta} \ln f(X; \theta)\right]^2.$$

Sustituyendo este resultado en 8.13 se tiene que

$$V(T) \geq \frac{\left[\tau'(\theta)\right]^2}{nE\left[\frac{\partial}{\partial\theta} \ln f(X; \theta)\right]^2}.$$

Esta desigualdad se conoce como la desigualdad de Cramer-Rao y el miembro de la derecha como la cota de Cramer-Rao .

Este teorema tiene dos usos: da un acotamiento inferior para la varianza de los estimadores insesgados (un investigador que emplea un estimador insesgado cuya varianza está cerca a la cota de Cramer-Rao, sabe que está usando un buen estimador), y si se puede encontrar un estimador insesgado cuya varianza coincide con la cota de Cramer-Rao, este es un UMVUE.

Ejemplo 8.17. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim B(1, p)$, se desea hallar un UMVUE para p.

Solución. Por lo que $f(x; p) = p^x (1-p)^{1-x} I_{\{0,1\}}(x)$ entonces $\ln f(x; p) = x \ln p + (1-x)\ln(1-p)$, luego

$$\frac{\partial}{\partial p} \ln f(x; p) = \frac{x}{p} - \frac{1-x}{1-p} = \frac{x-p}{p(1-p)}$$

y

$$\sum \frac{\partial}{\partial p} \ln f(x_i; p) = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p} = \frac{n}{p(1-p)}(\bar{x} - p) = k(p, n)(\bar{x} - p).$$

Puesto que $\bar{X}$ es un estimador insesgado de p, entonces $\bar{X}$ es un UMVUE de $\tau(p) = p$.

Para hallar la cota de Cramer-Rao de los estimadores insesgados de $\tau(p) = p$ observe que

$$\tau'(p) = 1, \quad \text{y} \quad E\big[\frac{\partial}{\partial p}\ln f(X,p)\big]^2 = E\left[\frac{X-p}{p(1-p)}\right]^2 = \frac{1}{p(1-p)}.$$

Si T es cualquier estimador insesgado de p, por 8.12 se tiene que $V(T) \geq \frac{p(1-p)}{n}$. El miembro de la derecha (la cota de Cramer-Rao) es precisamente $V(\bar{X})$.

Ejemplo 8.18. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim P(\lambda)$, se desea hallar un UMVUE de λ.

Solución. Cómo $\ln f(x;\lambda) = x\ln\lambda - \lambda - \ln x!$, entonces $\frac{\partial}{\partial\lambda}\ln f(x;\lambda) = \frac{x}{\lambda} - 1 = \frac{x-\lambda}{\lambda}$.

Luego

$$E\big[\frac{\partial}{\partial\lambda}\ln f(x;\lambda)\big]^2 = \frac{1}{\lambda^2}E(X-\lambda)^2 = \frac{V(X)}{\lambda^2} = \frac{1}{\lambda}$$

Si T es un estimador insesgado de λ, entonces $V(T) \geq \frac{1}{n\frac{1}{\lambda}} = \frac{\lambda}{n}$.

Ahora, $\sum \frac{\partial}{\partial\lambda}\ln f(x_i;\lambda) = \frac{\sum x_i}{\lambda} - n = \frac{n}{\lambda}(\bar{x}-\lambda)$. Por lo que $E(\bar{X}) = \lambda$ y $V(\bar{X}) = \frac{\lambda}{n}$ coincide con la cota de Cramer-Rao entonces, $\bar{X}$ es un UMVUE de λ.

Observaciones

1. Bajo ciertos supuestos que incluyen la existencia de la segunda derivada y la validez del intercambio de ciertas diferenciaciones e integraciones

$$E\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right]^2 = -E\left[\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta)\right].$$

   Esta igualdad es útil cuando la esperanza de la izquierda es más difícil de obtener que la de la derecha.

   Ejemplo 8.19. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu;\sigma^2)$. Si μ es conocido, halle un UMVUE para alguna función $\tau(\sigma^2)$.

   Solución. Como μ es conocido, la fdp es $f(x;\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ y

$$\ln f(x;\sigma^2) = -\ln\sqrt{2\pi} - \frac{1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(x-\mu)^2,$$

por tanto

$$\frac{\partial}{\partial \sigma^2} \ln f(x;\sigma^2) = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(x-\mu)^2 = \frac{1}{2\sigma^4}[(x-\mu)^2 - \sigma^2](*)$$

Para hallar el acotamiento inferior de la cota de Cramer-Rao, el valor esperado de (*) al cuadrado no es fácil de obtener, luego se encuentra ahora la segunda derivada:

$$\frac{\partial^2}{\partial (\sigma^2)^2} \ln f(x;\sigma^2) = \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(x-\mu)^2 = -\frac{1}{\sigma^4}\left[\frac{1}{\sigma^2}(x-\mu)^2 - \frac{1}{2}\right].$$

De esta manera, $-E\left[\frac{\partial^2}{\partial(\sigma^2)^2} \ln f(X;\sigma^2)\right] = \frac{1}{\sigma^4}(1-\frac{1}{2}) = \frac{1}{2\sigma^4}$, y si T es un estimador insesgado de σ^2, entonces $V(T) \geq \frac{2\sigma^4}{n}$.

Cualquier estimador insesgado cuya varianza sea $\frac{2\sigma^4}{n}$ es un UMVUE de σ^2. Puesto que $V(S^2) = \frac{2\sigma^4}{n-1}$, entonces S^2 no es un UMVUE de σ^2. Sin embargo, de (*) se tiene que

$$\sum \frac{\partial}{\partial \sigma^2} \ln f(x_i;\sigma^2) = \frac{1}{2\sigma^4}\left[\sum (x_i-\mu)^2 - n\sigma^2\right] = \frac{n}{2\sigma^4}\left[\frac{\sum (x_i-\mu)^2}{n} - \sigma^2\right].$$

Si $T = \frac{\sum (X_i-\mu)^2}{n}$, entonces $E(T) = \sigma^2$, es decir, es un estimador insesgado de σ^2 y por tanto $\frac{\sum (X_i-\mu)^2}{n}$ es un UMVUE de σ^2.

2. Sea $\hat{\theta}$ el MLE de θ, el cual se obtiene por la solución de

$$\frac{\partial}{\partial \theta} \ln L(\theta) = 0.$$

Si T^* es un estimador insesgado de $\tau^*(\theta)$ cuya varianza coincide con el acotamiento inferior de Cramer-Rao, entonces $t^*(x_1, x_2, ..., x_n) = \tau^*(\hat{\theta})$ es un UMVUE.

En el ejemplo 8.7 se encontró que $\bar{X}$ es el MLE de p y se obtuvo por la solución de $\frac{\partial}{\partial \theta} \ln L(\theta) = 0$. Sea $\tau(p) = P(X=0) = 1-p$ y sea $T = 1-\bar{X}$. Como $E(T) = 1-p$, entonces $1-\bar{X}$ es un UMVUE de $1-p$.

Ejemplo 8.20. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim E(\theta)$. Se desea hallar un UMVUE para alguna función de $\tau(\theta)$.

Por lo que $\ln f(x;\theta) = \ln\theta - \theta x$, entonces $\frac{\partial}{\partial\theta}\ln f(x;\theta) = \frac{1}{\theta} - x$, y $E[\frac{\partial}{\partial\theta}\ln f(X;\theta)]^2 = V(X) = \frac{1}{\theta^2}$. Luego para un estimador insesgado T de $\tau(\theta) = \frac{1}{\theta}$ se tiene que,

$$V(T) \geq \frac{(-1/\theta^2)^2}{n\frac{1}{\theta^2}} = \frac{1}{n\theta^2}.$$

Se sabe además que el MLE de θ obtenido por $\frac{\partial}{\partial\theta}\ln f(x;\theta) = 0$ es $\dfrac{1}{\bar{X}}$. Si $T = \bar{X}$, entonces $E(T) = \frac{1}{\theta}$ y $V(T) = \frac{1}{n\theta^2}$. Por tanto, por la observación anterior $\bar{X}$ es un UMVUE de $\frac{1}{\theta}$.

3. Si T es un estimador insesgado de algún $\tau(\theta)$ cuya varianza coincide con la cota de Cramer-Rao, entonces $f(x;\theta)$ es de la *clase exponencial*; reciprocamente, si $f(x;\theta)$ es de la clase exponencial, entonces existe un estimador insesgado de alguna función $\tau(\theta)$ cuya varianza coincide con la cota de Cramer-Rao. El ejemplo anterior ilustra también esta observación.

Las comparaciones que incluyen varianzas de los estimadores se usa a menudo para decidir cuál método hace uso más **eficiente** de los datos.

Definición 8.8. Eficiencia. Sean T y T^* dos estimadores insesgados de $\tau(\theta)$, se define la *eficiencia relativa* de T respecto a T^* por

$$re(T, T^*) = \frac{V(T^*)}{V(T)}$$

Un estimador insesgado T^* de $\tau(\theta)$ se dice que es *eficiente* si $re(T, T^*) \leq 1$ para todo estimador insesgado T de $\tau(\theta)$, y todo $\theta \in \Theta$. La eficiencia de un estimador insesgado T de $\tau(\theta)$ está dada por

$$e(T) = re(T, T^*)$$

si T^* es un estimador insesgado de $\tau(\theta)$.

Ejemplo 8.21. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu, \sigma^2)$. Se desea determinar la eficiencia relativa de la mediana muestral respecto a $\bar{X}$.

Solución. Sean $\hat{\theta}_1 = \bar{X}$ y $\hat{\theta}_2$ la mediana muestral. Entonces la eficiencia relativa de la mediana muestral relativa a la media muestral es

$$re(\hat{\theta}_2, \hat{\theta}_1) = \frac{V(\hat{\theta}_1)}{V(\hat{\theta}_2)} = \frac{\frac{\sigma^2}{n}}{\frac{\pi\sigma^2}{2n}} = 0.6366.$$

De aquí que se debería preferir el uso de la media muestral antes que la mediana muestral.

8.3.3. Suficiencia e insesgamiento

El concepto de suficiencia se emplea ahora para encontrar estimadores insesgados. El teorema principal de esta sección relaciona los estadísticos suficientes con los estimadores insesgados y hace uso de los resultados,

$$E(X) = E[E(X|Y)]$$

y

$$V(X) = V[E(X|Y)] + E[V(X|Y)].$$

Teorema 8.5. (Rao-Blackwell). Sea T un estimador insesgado y S un estadístico suficiente para $\tau(\theta)$. Si $T^* = E(T|S)$ entonces,

1. T^* es un estadístico que es función de S.
2. T^* es un estimador insesgado de $\tau(\theta)$.
3. T^* es un UMVUE de $\tau(\theta)$.

Demostración.

1. Si S es un estadístico suficiente, la distribución de T dado S es independiente de θ, luego $T^* = E(T|S)$ es independiente de θ y solo es función de S.
2. $E(T^*) = E[E(T|S)] = E(T) = \tau(\theta)$, por hipótesis.
3. $V(T) = V\big[E(T|S)\big] + E\big[V(T|S)\big] = V(T^*) + E\big[V(T|S)\big] \geq V(T^*)$. Esto es válido para cualquier estadístico T.

Luego $T^* = E(T|S)$ es un UMVUE de $\tau(\theta)$.

Ejemplo 8.22. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim B(1,p)$. Se desea construir explicitamente un estimador $T^* = E(T|S)$.

Solución. Como $E(X_1) = p$ y ΣX_i es un estadístico suficiente para p (ver ejemplo 7.8) entonces, $T^* = E(X_1|\Sigma X_i)$ es un UMVUE de p. Pero, ¿como es T^*?

Para ver esto se encuentra primero la distribución condicional de X_1 dado $\Sigma X_i = s$. Así,

$$\begin{aligned}
P(X_1 = 0|\Sigma X_i = s) &= \frac{P(X_1 = 0; \Sigma X_i = s)}{P(\Sigma X_i = s)} = \frac{P(X_1 = 0; \Sigma_{i=2}^n X_i = s)}{P(\Sigma X_i = s)} \\
&= \frac{P(X_1 = 0)P(\Sigma_{i=2}^n X_i = s)}{P(\Sigma X_i = s)} \\
&= \frac{(1-p)\binom{n-1}{s}p^s(1-p)^{n-1-s}}{\binom{n}{s}p^s(1-p)^{n-s}} = \frac{n-s}{n}.
\end{aligned}$$

$$\begin{aligned}
p(X_1 = 1|\Sigma X_i = s) &= \frac{P(X_1 = 1; \Sigma X_i = s)}{P(\Sigma X_i = s)} = \frac{P(X_1 = 1; \Sigma_{i=2}^n X_i = s-1)}{P(\Sigma X_i = s)} \\
&= \frac{P(X_1 = 1)P(\Sigma_{i=2}^n X_i = s-1)}{P(\Sigma X_i = s)} \\
&= \frac{p\binom{n-1}{s-1}p^{s-1}(1-p)^{n-s}}{\binom{n}{s}p^s(1-p)^{n-s}} = \frac{s}{n}.
\end{aligned}$$

Se observa que la distribución condicional de X_1 dado $\Sigma X_i = s$ es independiente de p, como debe ser.

Luego

$$T^* = E(X_1|\Sigma X_i) = 0.\frac{n-s}{n} + 1.\frac{s}{n} = \frac{s}{n} = \bar{X}.$$

Observe que $V(X_1) = p(1-p) > V(\bar{X}) = \frac{p(1-p)}{n}$ para $n > 1$.

Teorema 8.6. Si T^* es un UMVUE de $\tau(\theta)$, entonces es único.
Prueba. (Ejercicio)

El teorema de Rao-Blackwell nos dice cómo encontrar un estimador insesgado por condicionamiento de estadísticos suficientes y que para algunos problemas de estimación será un UMVUE. Para identificar estos problemas por el cual se obtiene un UMVUE, se da el concepto de completez.

Definición 8.9. Familia de densidades completa. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$ con $\theta \in \Theta$ y sea $T = t(X_1, X_2, ..., X_n)$ un estadístico. La familia de densidades de T se dice completa si y solo si

$$E[\phi(T)] = 0 \text{ para todo } \theta \in \Theta, \text{ implica que } P[\phi(T) = 0] = 1 \text{ para todo } \theta \in \Theta,$$

donde $\phi(T)$ es un estadístico.

Además, T se dice completo si y solo si su familia de densidades es completa.

Teorema 8.7. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$ con $\theta \in \Theta$, y Θ es un intervalo. Si $f(x;\theta)$ pertenece a la familia exponencial, entonces $\Sigma d(X_i)$ es un estadístico suficiente completo.
Prueba.(Ejercicio)

Teorema 8.8. Lehmann- Scheffé. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$. Si S es un estadístico suficiente completo y si $T^* = t^*(S)$ es un estimador insesgado de $\tau(\theta)$, entonces T^* es un UMVUE de $\tau(\theta)$.

Demostración. Sea T' otro estimador insesgado de $\tau(\theta)$ tal que $T' = t'(S)$, donde t' es una función. Entonces $E[T^* - T'] = 0$, para todo $\theta \in \Theta$, y $T^* - T'$ es una función de S.

Por la completez de S, $P[t^*(S) - t'(S)] = 1$, para todo $\theta \in \Theta$. Luego, hay un solo estimador insesgado de $\tau(\theta)$ que es función de S.

Sea T otro estimador insesgado de $\tau(\theta)$. Entonces, T^* debe ser igual $E(T|S)$ porque $E(T|S)$ es un estimador insesgado de $\tau(\theta)$ que depende de S. Por el teorema 8.5, $V(T)^* \leq V(T)$ para todo $\theta \in \Theta$, así que T^* es un UMVUE.

Ejemplo 8.23. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim P(\lambda)$. Como $f(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} I_{\{0,1,2,...\}}(x)$ pertenece a la familia exponencial con $d(x) = x$, entonces $\sum X_i$ es un estadístico suficiente completo. Una función de $\sum X_i$ con esperanza λ es $\bar{X}$, de manera que este es el UMVUE de λ.

8.4. Propiedades para muestras grandes

La calidad que un estimador sea insesgado es también útil al considerar sus propiedades asintóticas. Un estimador T de $\tau(\theta)$ puede no ser *bueno* para un tamaño de muestra n pequeño, pero puede ser razonable si tiene buenas propiedades asintóticas a medida que n aumenta. Con mucha frecuencia es posible evaluar las propiedades asintóticas de un estimador cuando estas son difíciles de determinar para muestras pequeñas. A continuación se definen algunas propiedades deseables para muestras grandes.

Definición 8.10. Consistencia simple. Una sucesión $\{T_n\}_{n\geq 1}$, de estimadores de $\tau(\theta)$ se dice que es de consistencia simple, si para todo $\epsilon > 0$

$$\lim_{n\to\infty} P\Big[|T_n - \tau(\theta)| < \epsilon\Big] = 1 \text{ para todo } \theta \in \Theta. \tag{8.14}$$

Es decir, si T_n converge estocásticamente a $\tau(\theta)$.

Definición 8.11. Consistencia en MSE. Una sucesión $\{T_n\}_{n\geq 1}$, de estimadores de $\tau(\theta)$ se dice que es de consistencia en error cuadrático medio , si para todo $\epsilon > 0$:

$$\lim_{n\to\infty} MSE_{T_n}(\theta) = \lim_{n\to\infty} E[T_n - \tau(\theta)]^2 = 0 \text{ para todo } \theta \in \Theta. \tag{8.15}$$

La consistencia en MSE implica que tanto el sesgo como la varianza del estimador se aproximan a 0.

En efecto, como $MSE_{T_n}(\theta) = V(T_n) + [b(T_n)]^2$ y si T_n es consistente en MSE, entonces $V(T_n) \to 0$ y $b(T_n) \to 0$ cuando $n \to \infty$.

Ejemplo 8.24. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$ tales que $\mu = E(X_i)$ y $\sigma^2 = V(X_i)$. Las sucesiones de estimadores $\bar{X}_n = \dfrac{1}{n}\sum X_i$ de μ y $S_n^2 = \dfrac{\sum (X_i - \bar{X}_n)^2}{n-1}$ de σ^2 son consistentes en MSE para μ y σ^2 respectivamente.

En efecto,

$$\lim_{n\to\infty} MSE_{\bar{X}_n}(\mu) = \lim_{n\to\infty} V(\bar{X}_n) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$$

y

$$\lim_{n\to\infty} MSE_{S_n^2}(\sigma^2) = \lim_{n\to\infty} V(S_n^2) = \lim_{n\to\infty} \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right) = 0.$$

Teorema 8.9. Si una sucesión $\{T_n\}_{n\geq 1}$ de estimadores de $\tau(\theta)$ es consistente en MSE, entonces también lo es de consistencia simple.

Demostración. Sea T_n una sucesión de estimadores de $\tau(\theta)$ consistente en MSE, entonces

$$\begin{aligned}\lim_{n\to\infty} P\big[|T_n-\tau(\theta)|<\epsilon\big] &= \lim_{n\to\infty} P\big[(T_n-\tau(\theta))^2<\epsilon^2\big] \\ &\geq \lim_{n\to\infty}\left[1-\frac{E(T_n-\tau(\theta))^2}{\epsilon^2}\right]=1.\end{aligned}$$

Ninguna probabilidad es mayor que 1, así que $\lim_{n\to\infty} P\big[|T_n-\tau(\theta)|<\epsilon\big]=1$.

El recíproco del teorema no es necesariamente cierto.

Definición 8.12. Insesgamiento asintótico. Una sucesión $\{T_n\}_{n\geq 1}$ de estimadores se dice asintóticamente insesgada para $\tau(\theta)$ si,

$$\lim_{n\to\infty} E(T_n)=\tau(\theta) \text{ para todo } \theta\in\Theta. \tag{8.16}$$

Es posible, también, formular una versión asintótica de la eficiencia.

Definición 8.13. Eficiencia asintótica. Sean $\{T_n\}$ y $\{T_n^*\}$ dos sucesiones de estimadores asintóticamente insesgados para $\tau(\theta)$. Se define la *eficiencia asintótica* de T_n relativa a T_n^* por

$$are(T_n,T_n^*)=\lim_{n\to\infty}\frac{V(T_n^*)}{V(T_n)}.$$

La sucesión $\{T_n^*\}$ se dice que es *eficiente asintóticamente* si $are(T_n,T_n^*)\leq 1$ para cualquier otra sucesión asintóticamente insesgada $\{T_n\}$ y todo $\theta\in\Theta$.

La eficiencia asintótica de una sucesión asintóticamente insesgada $\{T_n\}$, está dada por

$$ae\{T_n\}=are(T_n,T_n^*), \quad \text{si } \; T_n \text{ es eficiente asintóticamente.}$$

La cota de Cramer-Rao no siempre es alcanzable para un n fijo, pero para grandes muestras ésta es, a menudo, asintóticamente alcanzable.

Ejemplo 8.25. Sea $X_1,X_2,...,X_n$ una muestra aleatoria de $f(x;\theta)=\theta e^{-\theta x}I_{(0,\infty)}(x)$. Entonces $\bar{X}$ es un estimador eficiente para $\dfrac{1}{\theta}$.

En efecto, del ejemplo **8.9**, se tiene que el estimador de máxima verosimilitud de θ es $\dfrac{1}{\bar{X}}$. Por la propiedad de invarianza $T_n = \bar{X}$ es un estimador de $\dfrac{1}{\theta}$ y puesto que $E(\bar{X})$ es $\dfrac{1}{\theta}$, entonces $\bar{X}$ es un estimador insesgado de $\dfrac{1}{\theta}$. La cota de Cramer-Rao para un estimador insesgado T de $\dfrac{1}{\theta}$ es

$$CCR(T) = \frac{\left(-\dfrac{1}{\theta^2}\right)^2}{nE\left[\dfrac{\partial}{\partial\theta}\ln f(x;\theta)\right]^2} = \frac{\dfrac{1}{\theta^4}}{nE\left(\dfrac{1}{\theta} - X\right)^2} = \frac{1}{n\theta^4 \cdot \dfrac{1}{\theta^2}} = \frac{1}{n\theta^2}.$$

La varianza de $\bar{X}$ es $V(\bar{X}) = \frac{1}{n\theta^2}$, así que $\lim\limits_{x\to\infty} \frac{CCR}{V(\bar{X})} = \lim\limits_{x\to\infty} 1 = 1$. Luego, $\bar{X}_n$ es una sucesión asintóticamente eficiente para $\dfrac{1}{\theta}$.

Ejemplo 8.26. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta) = e^{-(x-\theta)} I_{(\theta,\infty)}(x)$. Demuestre que $X_{(1)}$, el mínimo de las X_i es un estimador eficiente para θ.

Prueba. Del ejemplo 8.11 se infiere para $\lambda = 1$ que $L(\theta) = e^{-\sum(x_i-\theta)}$ es máxima para $\hat{\theta} = x_{(1)}$.

De 6.12, $f_{X_{(1)}}(x) = n[1 - F_X(x)]^{n-1} f_X(x)$ y como $F_X(x) = 1 - e^{-(x-\theta)}$ para $x > \theta$, entonces $f_{X_{(1)}}(x) = ne^{-n(x-\theta)}$ para $x > \theta$.

De esta manera,

$$E[X_{(1)}] = \int_\theta^\infty nxe^{-n(x-\theta)}dx = \theta + \frac{1}{n} \;\text{ y }\; \lim_{n\to\infty} E(X_{(1)}) = \theta.$$

Es decir, $X_{(1)}$ es un estimador asintóticamente insesgado para θ.

Ahora,

$$E[X_{(1)}^2] = \int_\theta^\infty nx^2 e^{-n(x-\theta)}dx = \theta^2 + \frac{2\theta}{n} + \frac{2}{n^2}.$$

Luego $V[X_{(1)}] = E[X_{(1)}^2] - [E(X_{(1)})]^2 = \dfrac{1}{n^2}$.

Sea $\hat{\Theta}_1 = X_{(1)} - \dfrac{1}{n}$, entonces $\lim\limits_{x\to\infty} E(\hat{\Theta}_1) = \theta$. Por tanto,

$$are(\hat{\Theta}_1, X_{(1)}) = \lim_{n\to\infty} \frac{V[X_{(1)}]}{V(\hat{\Theta}_1)} = 1.$$

Se puede demostrar que $\hat{\Theta}_1 = X_{(1)} - \dfrac{1}{n}$ es un UMVUE para θ. Por esta razón, $X_{(1)}$ es un estimador eficiente para θ.

Un estimador con varianza de orden $\dfrac{1}{n^2}$ se conoce como un estimador *super-eficiente.*

A menudo es difícil obtener una expresión exacta para la varianza de un estimador propuesto. Una aproximación, que también se usa, restringe la atención solo a estimadores que son asintóticamente normales y reemplaza la varianza exacta por varianzas asintóticas en la definición de eficiencia asintótica relativa.

Definición 8.14. Mejores estimadores asintóticamente normales. Una sucesión de $\{T_n^*\}$ de estimadores de $\tau(\theta)$ se dice que es mejor asintóticamente normal (BAN o CANE) si y solo si,

a) La distribución de $\sqrt{n}[T_n^* - \tau(\theta)]$ es $N[0, {\sigma^*}^2(\theta)]$, cuando $n \to \infty$.

b) Para todo $\epsilon > 0$, $\lim\limits_{n\to\infty} P\big[|T_n^* - \tau(\theta)| > \epsilon\big] = 0$ para todo $\theta \in \Theta$.

c) Para cualquier otra sucesión $\{T_n\}$ tal que $\sqrt{n}[T_n - \tau(\theta)]$ es $N(0, \sigma^2(\theta))$ entonces, ${\sigma^*}^2(\theta)) \leq \sigma^2(\theta))$.

Es claro, de la definición, que las dos sucesiones $\{T_n^*\}$ y $\{T_n\}$ son estimadores de consistencia simple.

Ejemplo 8.27. Sean $X_1, X_2, ..., X_n, ...$ variables aleatorias independientes normales (μ, σ^2) y sea $T_n^* = \bar{X}_n$. Entonces $E(T_n^*) = \mu$ y $V(T_n^*) = \dfrac{\sigma^2}{n}$ y,

$$\frac{\sqrt{n}(T_n^* - \mu)}{\sigma} = \frac{T_n^* - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1).$$

Luego, $\sqrt{n}(T_n^* - \mu) \sim N(0, \sigma^2)$.

Ningún otro estimador tiene varianza más pequeña que $\bar{X}_n$ en el límite. Por ejemplo, si T_n es la mediana muestral, se deduce del teorema 7.15 que $\sqrt{n}(T_n - \mu) \sim N\left(0, \dfrac{\pi}{2}\sigma^2\right)$. Como $\sigma^2 < \dfrac{\pi}{2}\sigma^2$ entonces, $\bar{X}_n$ es mejor estimador que T_n, la mediana muestral.

8.5. Estimadores minimax y de Bayes

Cuando una estimación difiere del valor verdadero del parámetro a estimar, se puede considerar que hay una pérdida, una funcion de la diferencia entre la estimación y el parámetro. En la decisión tomada (escogencia del estimador) se puede cometer un error. Si es así, se necesita alguna medida de la gravedad del error . El MSE es un caso especial de una función llamada *función de pérdida*

8.5.1. Funciones de pérdida y estimadores minimax

Definición 8.15. Función de pérdida. Sea T un estimador de $\tau(\theta)$. Una función de pérdida L, es cualquier función de valor real tal que

$$L[t;\tau(\theta)] \geq 0 \text{ para todo } t, \text{ y } L[t;\tau(\theta)] = 0 \text{ cuando } t = \tau(\theta) \tag{8.17}$$

La función de pérdida en un problema de estimación puntual refleja el hecho que será mayor para un error grande que para un error pequeño. Como es natural, se desean pérdidas pequeñas.

Ejemplo 8.28. Son funciones de pérdida para un estimador T de $\tau(\theta)$

$$L_1(t;\theta) = |t - \tau(\theta)|, \text{ pérdida del error absoluto}$$

$$L_2(t;\theta) = [t - \tau(\theta)]^2, \text{ pérdida del error al cuadrado.}$$

L_1 y L_2 son funciones crecientes con el crecimiento de la distancia, con un valor mínimo cuando $t = \tau(\theta)$, es decir, la pérdida es mínima cuando la acción es correcta.

L_2 penaliza más que L_1 para grandes discrepancias y L_1 penaliza menos que L_2 para pequeñas discrepancias.

Una variación de L_2 es

$$L_3(t;\theta) = \begin{cases} [t - \tau(\theta)]^2 \text{ si } t < \tau(\theta) \\ a[(t - \tau(\theta)]^2 \text{ si } t \geq \tau(\theta) \text{ y } a > 1. \end{cases}$$

L_3 penaliza más la sobrestimación que la subestimación.

En general, un experimentador debe considerar las consecuencias de varios errores en la estimación para diferentes valores de θ y especificar una pérdida que refleja estas consecuencias.

En un análisis de decisión, la calidad de un estimador está cuantificada por la *función de riesgo*.

Definición 8.16. Función de riesgo. Sea T un estimador de $\tau(\theta)$, se define la función de riesgo de T como el valor esperado de la pérdida; es decir,

$$R_T(\theta) = E[L(T;\theta)]. \tag{8.18}$$

Si la función de pérdida es L_2, el riesgo es el MSE del estimador.

Cuando se comparan dos estimadores T_1 y T_2 de $\tau(\theta)$, es posible que se pueda tener un riesgo más pequeño que el otro. Esto no siempre es posible si $R_T(\theta) = MSE_T(\theta)$.

Puesto que el valor verdadero de $\tau(\theta)$ es desconocido, se debe usar un estimador T que tenga un valor pequeño de $R_T(\theta)$ para todos los valores de θ.

Definición 8.17. Estimador admisible. Un estimador T_1 de $\tau(\theta)$ es *mejor estimador* que T_2 si y solo si,

$$R_{T_1}(\theta) \leq R_{T_2}(\theta), \text{ para todo } \theta \in \Theta, \text{ y } R_{T_1}(\theta) < R_{T_2}(\theta), \text{ para al menos un } \theta \in \Theta. \tag{8.19}$$

Un estimador T se dice *admisible* si y solo si no hay otro mejor estimador.

Ejemplo 8.29. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu, \sigma^2)$. Se desea estimar μ.

Solución. Sean $T_1 = \bar{X}$ y T_2 la mediana muestral. Si $L(t;\mu) = (t-\mu)^2$ entonces, $R_{T_1}(\mu) = MSE_{T_1}(\mu) = \dfrac{\sigma^2}{n}$ y, $R_{T_2}(\theta) = MSE_{T_2}(\mu) = \dfrac{\pi}{2}\dfrac{\sigma^2}{n}$.

Luego, $R_{T_1}(\mu) < R_{T_2}(\mu)$ y por tanto $\bar{X}$ es mejor estimador que la mediana muestral.

Otro criterio que algunas veces se emplea para seleccionar un estimador de la clase de los estimadores admisibles es el *criterio minimax*.

Definición 8.18. Estimador minimax. Un estimador T^* de $\tau(\theta)$ se dice que es un *estimador minimax* si

$$\max_\theta R_{T^*}(\theta) \leq \max_\theta R_T(\theta), \text{para todo estimador } T \text{ de } \tau(\theta) \tag{8.20}$$

Es decir, T^* es un estimador tal que,

$$\max_{\theta} R_{T^*}(\theta) = \min_{T} \max_{\theta} R_T(\theta), \text{ para todo estimador } T.$$

8.5.2. Estimador de Bayes

En el problema de obtener un estimador puntual como se ha tratado, problema clásico, se ha supuesto que la muestra viene de alguna densidad conocida $f(x;\theta)$, y θ es un valor fijo pero desconocido. En la teoría bayesiana θ se considera como un valor de una variable aleatoria Θ que se puede describir por una distribución de probabilidad (llamada distribución a priori). Esta es una distribución subjetiva, basada en la experiencia del investigador y se formula antes de tomar la muestra. Después de seleccionar la muestra de una población indexada por θ, la distribución *a priori* se combina con la información muestral. Esta distribución así obtenida se llama distribución *a posteriori*.

Si se denota la fdp o la fmp de la distribución a priori de Θ por $g(\theta)$ y la distribución muestral por $f(x|\theta)$, entonces la *distribución posterior* (distribución condicional de θ dada la muestra) está dada por

$$f_{\Theta|X_1,X_2,...,X_n}(\theta|x_1,x_2,...,x_n) = \frac{f(x_1,x_2,...,x_n|\theta)g(\theta)}{\int f(x_1,x_2,...,x_n|\theta)g(\theta)d\theta}. \tag{8.21}$$

Definición 8.19. Estimador posterior de Bayes. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x|\theta)$ y sea θ un valor de la variable aleatoria Θ con función de densidad conocida $g(\theta)$. Se define el estimador posterior de Bayes de $\tau(\theta)$ con respecto a la densidad a priori g por

$$E\big[\tau(\Theta)|X_1,X_2,...,X_n\big] = \int \tau(\theta) f(\theta|x_1,x_2,...,x_n)d\theta.$$

Es decir,

$$E\big[\tau(\Theta)|X_1,X_2,...,X_n\big] = \frac{\int \tau(\theta)\big[\Pi f(x_i|\theta)\big]g(\theta)d\theta}{\int \big[\Pi f(x_i|\theta)\big]g(\theta)d\theta}. \tag{8.22}$$

Ejemplo 8.30. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x|\lambda) = \dfrac{\lambda^x e^{-\lambda}}{x!}$, para $x = 0, 1, \cdots$ y $\lambda \in \Lambda$. Si Λ tiene densidad $g(\lambda) = \dfrac{\lambda^{\alpha-1}e^{-\lambda}}{\Gamma(\alpha)}$, para $\lambda > 0$ y $\alpha > 0$, se desea hallar el estimador a posteriori de Bayes.

Solución. La fmp a posteriori de Bayes es

$$
\begin{aligned}
f(\lambda|x_1,x_2,...,x_n) &= \frac{\big[\Pi f(x_1,x_2,...,x_n|\lambda)\big]g(\lambda)}{\int_0^\infty \big[\Pi f(x_1,x_2,...,x_n|\lambda)\big]g(\lambda)d\lambda} \\
&= \frac{\frac{\lambda^{\sum x_i+\alpha-1}e^{-(n+1)\lambda}}{\Gamma(\alpha)\Pi x_i!}}{\int_0^\infty \frac{\lambda^{\sum x_i+\alpha-1}e^{-(n+1)\lambda}}{\Gamma(\alpha)\Pi x_i!}d\lambda} \\
&= \frac{\lambda^{\sum x_i+\alpha-1}e^{-(n+1)\lambda}}{\int_0^\infty \lambda^{\sum x_i+\alpha-1}e^{-(n+1)\lambda}d\lambda}.
\end{aligned}
$$

Si $u=(n+1)\lambda$ entonces,

$$
\begin{aligned}
\int_0^\infty \lambda^{\Sigma x_i+\alpha-1}e^{-(n+1)\lambda}d\lambda &= \frac{1}{n+1}\int_0^\infty \left(\frac{u}{n+1}\right)^{\Sigma x_i+\alpha-1} e^{-u}du \\
&= \frac{1}{(n+1)^{\Sigma x_i+\alpha}}\Gamma(\Sigma x_i+\alpha).
\end{aligned}
$$

Luego

$$
f(\lambda|x_1,x_2,...,x_n) = \frac{(n+1)^{\Sigma x_i+\alpha}\lambda^{\Sigma x_i+\alpha-1}e^{-(n+1)\lambda}}{\Gamma\big(\Sigma x_i+\alpha\big)}.
$$

Por tanto, la estimación posterior de Bayes es

$$
\begin{aligned}
E(\Lambda|x_1,x_2,...,x_n) &= \frac{(n+1)^{\Sigma x_i+\alpha}}{\Gamma(\Sigma x_i+\alpha)}\int_0^\infty \lambda^{\Sigma x_i+\alpha}e^{-(n+1)\lambda}d\lambda \\
&= \frac{(n+1)^{\Sigma x_i+\alpha}}{\Gamma(\Sigma x_i+\alpha)}\frac{\Gamma(\Sigma x_i+\alpha+1)}{(n+1)^{\Sigma x_i+\alpha}+1} \\
&= \frac{\Sigma x_i+\alpha}{n+1}.
\end{aligned}
$$

Es claro que el estimador posterior de Bayes es

$$
E(\Lambda|X_1,X_2,...,X_n) = \frac{\Sigma X_i+\alpha}{n+1}.
$$

Finalmente, es de anotar, que las funciones de riesgo de dos estimadores T_1 y T_2, como funciones de θ se pueden cruzar, siendo $R_{T_1}(\theta) < R_{T_2}(\theta)$ para algunos valores de θ y $R_{T_1}(\theta) > R_{T_2}(\theta)$ para otros valores de θ, y como θ es desconocido, es difícil hacer una escogencia entre T_1 y T_2. La dificultad radica en la dependencia del riesgo como función de θ. La integral

$$\int_{\Theta} R_T(\theta)g(\theta)d\theta$$

donde Θ es una variable aleatoria y g(.) es su función de probabilidad, no depende de θ, se conoce como el *riesgo de Bayes* y se denota por $r(t)$. Es decir, el **riesgo de Bayes** es

$$r(t) = E_\theta[R_T(\theta)] = \int_{\Theta} R_T(\theta)g(\theta)d\theta. \tag{8.23}$$

De esta manera, al comparar dos estimadores se debe preferir al que tenga riesgo de Bayes menor. Esto conduce al concepto de *estimador de Bayes*.

Definición 8.20. Estimador de Bayes. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x|\theta)$, el estimador de Bayes T^* de $\tau(\theta)$, relativo a la función de riesgo $R_{T^*}(\theta)$ y a la función de densidad g(.) es el estimador con riesgo esperado mínimo, es decir,

$$E[R_{T^*}(\theta)] \leq E[R_T(\theta)], \text{ para todo estimador } T \text{ de } \tau(\theta). \tag{8.24}$$

Los siguientes teoremas dan estimadores de Bayes para funciones de pérdida especiales.

Teorema 8.10. El estimador de Bayes T de $\tau(\theta)$ bajo la función de pérdida $L(T;\theta) = [t - \tau(\theta)]^2$ es

$$T = \int_{\theta} \tau(\theta) f_{\Theta|X}(\theta|x)d\theta, \tag{8.25}$$

es decir, el estimador posterior de Bayes.

Teorema 8.11. El estimador de Bayes T de θ bajo la función de pérdida $L(T;\theta) = |t - \theta|$ es la mediana de la distribución posterior.

Teorema 8.12. Si T^* es un estimador de Bayes con riesgo constante $R_{T^*}(\theta) = c$, entonces T^* es un estimador minimax.

Ejemplo 8.31. En el ejemplo 8.30 obtenga el estimador de Bayes, T, de $\tau(\lambda) = \frac{1}{\lambda}$ si $L(T, \lambda) = (t - \lambda)^2$.

Solución. En el ejemplo 8.30 se encontró que:

$$f_{\Theta|X}(\theta|x) = \frac{(n+1)^{\sum x_i + \alpha} \lambda^{\sum x_i + \alpha - 1} e^{-(n+1)\lambda}}{\Gamma\left(\sum x_i + \alpha\right)}.$$

Por el teorema 8.10, el estimador de Bayes T de $\frac{1}{\lambda}$ bajo la función de pérdida error cuadrático es

$$T = \int_\lambda \frac{1}{\lambda} f_{\Lambda|X}(\lambda|x)d\lambda = \frac{(n+1)^{\Sigma x_i+\alpha}}{\Gamma(\Sigma x_i+\alpha)} \int_0^\infty \lambda^{\Sigma x_i+\alpha-2} e^{-(n+1)\lambda}.$$

Al hacer $u = (n+1)\lambda$, como en el ejemplo anterior se tiene que,

$$\begin{aligned} T &= \frac{(n+1)^{\sum x_i+\alpha}}{\Gamma(\Sigma x_i+\alpha)} \cdot \frac{1}{(n+1)^{\sum x_i+\alpha-1}} \int_0^\infty u^{\Sigma x_i+\alpha-2} e^{-u} du \\ &= \frac{n+1}{\Gamma(\Sigma x_i+\alpha)} \Gamma(\Sigma x_i+\alpha-1) = \frac{n+1}{\Sigma x_i+\alpha-1}. \end{aligned}$$

8.6. Ejercicios

1. Una observación se toma de una variable aleatoria discreta X con fmp $f(x;\theta)$, donde $\theta \in \{1,2,3\}$(ver tabla). Encuentre el estimador de máxima verosimilitud de θ.

x	$f(x;1)$	$f(x;2)$	$f(x;3)$
0	$\frac{1}{3}$	$\frac{1}{4}$	0
1	$\frac{1}{3}$	$\frac{1}{4}$	0
2	0	$\frac{1}{4}$	$\frac{1}{4}$
3	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{2}$
4	$\frac{1}{6}$	0	$\frac{1}{4}$

2. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$, encuentre los estimadores por los métodos de momentos y de máxima verosimilitud cuando,

   a) $f(x;\theta) = \theta x^{\theta-1}; 0 < x < 1$ y , $\theta > 0$.
   b) $f(x;\theta) = (\theta+1)x^{-\theta-2}; x > 1$ y , $\theta > 0$.
   c) $f(x;\theta) = \theta^2 x e^{-\theta x}; x > 0$ y $\theta > 0$.

3. Encuentre los estimadores por el método de momentos y de máxima verosimilitud, cuando $X_1, X_2, ..., X_n$ es una muestra aleatoria de cada una de las siguientes distribuciones,

   a) $X \sim G(p)$.

b) $X \sim Bn(5, p)$.

c) $X \sim N(0, \theta)$.

d) $X \sim W(\theta, 1)$.

e) X con función de densidad $f(x, \lambda) = \frac{1}{2}e^{-|x-\lambda|}, -\infty < x < \infty$.

4. Sea $X_1, X_2, \cdots, X_n$ una muestra aleatoria de una distribución geométrica con parámetro p. Encuentre los estimadores de máxima verosimilitud de las siguientes cantidades,

a) $E(X) = 1/p$.

b) $V(X) = (1-p)/p^2$.

c) $P[X > k] = (1-p)^k$, para cualquier $k = 1, 2, ...$

5. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de

$$F(x; \alpha, \beta) = \left(\frac{x}{\beta}\right)^{\alpha} I_{[0,\beta]}(x), \text{para } \alpha > 0 \; y \; \beta > 0$$

a) Encuentre los estimadores por el método de momentos de α y β.

b) Encuentre los estimadores por el método de máxima verosimilitud de α y β.

c) La longitud(en centímetros) de huevos de codorniz encontrados en los nidos pueden ser modelados por esta distribucón. Las siguientes longitudes de 14 huevos de codorniz se encontraron: 2.2, 2.4, 2.1, 2.4, 2.5, 2.4, 2.2, 2.4, 2.3, 2.3, 2.4, 2.3, 2.1, 2.2, encuentre las estimaciones por el método de momentos y por máxima verosimilitud de α y β.

6. Sea $X_1, X_2, \cdots, X_n$ una muestra aleatoria de $f(x, \lambda) = (1+\lambda)x^{\lambda}I_{(0,1)(x)}$. Probar que el MLE de λ es

$$\hat{\Lambda} = -\left(1 + \frac{n}{\sum lnX_i}\right).$$

7. Basado en una muestra aleatoria de tamaño n de una distribución normal $X \sim N(\mu, \sigma^2)$, encuentre los estimadores de máxima verosimilitud de

a) $P[X > c]$ para cualquier c.

b) El percentil del 95 % de X.

8. Sea $Y_1, Y_2, \cdots, Y_n$ una muestra aleatoria de una distribución lognormal con parámetros μ y σ^2.

a) Encuentre los estimadores de máxima verosimilitud de μ y σ^2.

b) Encuentre el estimador de máxima verosimilitud de $E(Y)$.

Indicación: Recuerde que $Y \sim \text{logn}(\mu, \sigma^2)$ si $\ln(Y) \sim N(\mu, \sigma^2)$, o de forma equivalente, $Y = e^X$, donde $X \sim N(\mu, \sigma^2)$.

9. Si $X \sim B(n, p)$ y $\hat{p} = X/n$.

   a) Encuentre una constante c tal que $E[c\hat{p}(1 - \hat{p})] = p(1 - p)$.

   b) Encuentre un estimador insesgado de $V(X)$.

   c) Considere una muestra aleatoria de tamaño N de $B(n, p)$. Encuentre estimadores insesgados de p y $V(X)$ basados en la muestra aleatoria.

10. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de $X \sim U(\theta - 1, \theta + 1)$.

    a) Demuestre que la media muestral $\bar{X}$, es un estimador insesgado de θ.

    b) Demuestre que el "semirango", $\frac{X_{(n)} + X_{(1)}}{2}$ es un estimador insesgado de θ.

11. Si T es un estimador insesgado de θ, probar que T^2 es un estimador sesgado de θ^2, excepto para el caso trivial cuando $V(T) = 0$.

12. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución $N(0, \theta)$.

    a) ¿Es el estimador de máxima verosimilitud, $\widehat{\theta}$ un estimador insesgado de θ?

    b) ¿Es $\widehat{\theta}$ un $UMVUE$ de θ?

13. Sea $X \sim P(\mu)$, y sea $\theta = P[X = 0] = e^{-\mu}$.

    a) ¿Es $\widehat{\theta} = e^{-X}$ un estimador insesgado de θ?

    b) Muestre que $\widetilde{\theta} = u(X)$ es un estimador insesgado de θ, donde $u(0) = 1$ y $u(x) = 0$ si $x = 1, 2, \ldots$.

    c) Compare los estimadores de máxima verosimilitud de $\widehat{\theta}$ y $\widetilde{\theta}$ para estimar $\theta = e^{-\mu}$ cuando $\mu = 1$ y $\mu = 2$.

14. Considere una muestra aleatoria de tamaño n de una distribución con fdp $f(x; \theta) = \frac{1}{\theta} I_{(0,\theta]}(x), x > 0$.

    a) Encuentre el estimador de máxima verosimilitud $\widetilde{\theta}$.

    b) Encuentre el estimador de momentos $\widehat{\theta}$.

    c) ¿Es $\widehat{\theta}$ insesgado?

d) ¿Es $\widetilde{\theta}$ insesgado?

e) Compare los MSE de $\widehat{\theta}$ y $\widetilde{\theta}$.

15. Suponga que dos muestras aleatorias independientes de tamaños n_1 y n_2 de poblaciones normales con varianza común σ^2, son seleccionadas. Sean

$$S_i^2 = \frac{\Sigma_{j=1}^{n_i}(X_{ji} - \bar{X}_i)^2}{n_i - 1} \quad \text{para i=1,2}$$

a) Demuestre que $S^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ es un estimador insesgado de σ^2.

b) Encuentre $V(S^2)$.

16. Considere una muestra aleatoria de tamaño $n = 2$ de una distribución normal, $X_i \sim N(\theta, 1)$, donde $\Theta = \{\theta | 0 \leqslant \theta \leqslant 1\}$. Defina los estimadores como sigue: $\widehat{\theta}_1 = (1/2)X_1 + (1/2)X_2$, $\widehat{\theta}_2 = (1/4)X_1 + (3/4)X_2$, $\widehat{\theta}_3 = (2/3)X_1$ y $\widehat{\theta}_4 = (2/3)\widehat{\theta}_1$. Considere la pérdida de error cuadrado $L(t;\theta) = (t - \theta)^2$.

a) Compare las funciones de riesgo para estos estimadores.

b) Compare los estimadores por el principio minimax.

c) Encuentre el riesgo de Bayes de los estimadores, usando $\Theta \sim U(0, 1)$.

d) Encuentre el riesgo de Bayes de los estimadores, usando $\Theta \sim Be(2, 1)$.

17. Considere una muestra aleatoria de tamaño n de una distribución de Bernoulli, $X_i \sim B(1, p)$. Para la distribución a-priori $P \sim U(0, 1)$ y la función de pérdida el error cuadrático, encuentre:

a) El estimador de Bayes de p.

b) El estimador de Bayes de $p(1 - p)$.

c) Riesgo de Bayes para el estimador en a).

18. Sea X una variable aleatoria discreta con $P(X = -1) = 2p(1 - p)$ y $P(X = k) = p^k(1 - p)^{3-k}$, $k = 0, 1, 2, 3$, y $0 < p < 1$.

a) Determine si existe un UMVUE de p.

b) Determine si existe un UMVUE de $p(1 - p)$.

19. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de $X \sim B(1, p)$. Demuestre que

a) El UMVUE de $p(1 - p)$ es $T_n = \frac{n\bar{x}(1-\bar{x})}{n-1}$.

b) La varianza de T_n no alcanza la cota de Cramer Rao.

20. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de $X \sim N(\mu, \sigma^2)$ con σ^2 conocido.

a) Encuentre el UMVUE de $\tau(\mu) = e^{t\mu}$ con $t \neq 0$ fijo.

b) Determine si la varianza del estimador anterior alcanza la cota de Cramer Rao.

21. Considere una muestra aleatoria de tamaño n de una distribución de Poisson, $X_i \sim P(\mu)$.

a) Encuentre la cota de Cramer-Rao para las varianzas de los estimadores insesgados de μ.

b) Encuentre la cota de Cramer-Rao para las varianzas de los estimadores insesgados de $\theta = e^{-\mu}$.

c) Encuentra un UMVUE de μ.

d) Encuentre el estimador de máxima verosimilitud $\widehat{\theta}$ de θ.

e) ¿Es $\widehat{\theta}$ un estimador insesgado de θ?

f) ¿Es $\widehat{\theta}$ asintóticamente insesgado?

g) Muestre que $\widetilde{\theta} = [(n-1)/n]^{\sum X_i}$, es un estimador insesgado de θ.

h) Encuentre $V(\widetilde{\theta})$ y compare con la cota de Cramer-Rao de b).

Indicación: Note que $Y = \sum X_i \sim P(n\mu)$, y que $E(\widetilde{\theta})$ y $V(\widetilde{\theta})$ están relacionadas por la fgm de Y.

22. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución exponencial con media $1/\theta$, y asuma que la densidad a-priori de θ también es exponencial con media $1/\beta$, donde β es conocida.

a) Muestre que la distribución posterior es $G[(\beta + \sum x_i)^{-1}, n+1]$.

b) Usando el error cuadrático como pérdida, encuentre el estimador de Bayes de θ.

c) Usando el error cuadrático como pérdida, encuentre el estimador de Bayes de $\mu = 1/\theta$.

d) Usando el error absoluto como pérdida, encuentre el estimador de Bayes de θ. Use la notación de chi-cuadrado para expresar la solución.

e) Usando el error absoluto como pérdida, encuentre el estimador de Bayes de $\mu = 1/\theta$.

23. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de $f(x, \theta) = \frac{1}{\theta}e^{-\frac{1}{\theta}x}, x > 0, \theta > 0$. Defina $\hat{\Theta}_1 = \bar{X}$ y $\hat{\Theta}_2 = \frac{n\bar{X}}{n+1}$.

    a) Compare las varianzas de $\hat{\Theta}_1$ y $\hat{\Theta}_2$ para $n = 2$.

    b) Compare los MSE de $\hat{\Theta}_1$ y $\hat{\Theta}_2$ para $n = 2$.

    c) Encuentre el riesgo de Bayes de $\hat{\Theta}_1$ suponiendo que θ tiene fdp $f(\theta) = \frac{1}{2}e^{-\frac{1}{2}\theta}, \theta > 0$.

24. Suponga que $X \sim P(\lambda)$ y considere la función de pérdida $L(\hat{\lambda}, \lambda) = \frac{(\hat{\lambda}-\lambda)^2}{\lambda}$. Sea $\Lambda \sim G(\theta, \kappa)$, donde θ y κ son conocidos.

    a) Encuentre el estimador de Bayes de λ.

    b) Probar que $\hat{\Lambda} = X$ es el estimador minimax.

25. Pruebe el teorema 8.6.

26. Pruebe el teorema 8.7.

Capítulo 9

Estimación por intervalo

9.1. Introducción

En la estimación puntual, además del valor verdadero de un parámetro se quiere tener alguna certeza sobre que tan cerca está la estimación del parámetro. La varianza y el MSE del estimador dan alguna información sobre este punto. Sin embargo un enfoque que considera este problema es la *estimación por intervalo*. En general, en esta última se emplea un estimador puntual más o menos una cantidad constante, con la *confianza* que el valor verdadero del parámetro se encuentre dentro de este con cierta probabilidad.

Como en la estimación puntual el problema en la estimación por intervalo es doble: primero hay que determinar métodos para obtenerlos y luego verificar que los intervalos calculados son buenos u óptimos.

Definición 9.1. Intervalo de confianza. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$ y sean $L = l(X_1, X_2, ..., X_n)$ y $U = u(X_1, X_2, ..., X_n)$ dos estadísticos tales que $L \leq U$. El intervalo aleatorio (L, U) se dice que es un intervalo de confianza del $100\delta\,\%$ para $\tau(\theta)$ si,

$$P(L < \tau(\theta) < U) = \delta, \tag{9.1}$$

donde δ no depende de θ y $0 < \delta < 1$. A δ se le llama coeficiente de confianza.

Un valor observado del intervalo aleatorio es $(l(x_1, x_2, ..., x_n), u(x_1, x_2, ..., x_n))$ y se dice también que es un intervalo de confianza del $100\delta\,\%$ para $\tau(\theta)$. Este intervalo no es aleatorio y por tanto con esta convención se comete un abuso de lenguaje.

Ejemplo 9.1. Se desea obtener un intervalo de confianza del 95 % para la media μ de una población normal con varianza $\sigma^2 = 9$. Supóngase que para esto se toma una muestra de tamaño 16 y se obtuvo una media de $\bar{x} = 3.2$.

Para muestras de tamaño 16,

$$Z = \frac{\bar{X} - \mu}{\frac{3}{4}} \sim N(0,1) \text{ y } \ P(-1.96 < Z < 1.96) = 0.95.$$

Luego $P(-1.96 < \frac{\bar{X}-\mu}{\frac{3}{4}} < 1.96) = 0.95$. De manera que $P(-1.47 < \bar{X} - \mu < 1.47) = 0.95$.

Por tanto $P(\bar{X} - 1.47 < \mu < \bar{X} + 1.47) = 0.95$. Es decir, que $(\bar{X} - 1.47, \bar{X} + 1.47)$ (*) es un estimador por intervalo de μ con un 95 % de confianza.

Para $\bar{x} = 3.2$ se tiene que una estimación por intervalo del 95 % para μ es (1.73, 4.67). Esto significa que con una probabilidad de 0.95, este intervalo contendrá el parámetro μ.

Si $l(X_1, X_2, ..., X_n)$ y $u(X_1, X_2, ..., X_n)$ son estadísticos, los intervalos de confianza del 100δ % pueden ser de la forma,

$$\big(l(X_1, X_2, ..., X_n),\ \infty\big) \text{ si } \ P[l(X_1, X_2, ..., X_n) < \tau(\theta)] = \delta,$$

y

$$\big(-\infty,\ u(X_1, X_2, ..., X_n)\big) \text{ si } \ P[\tau(\theta) < u(X_1, X_2, ..., X_n)] = \delta$$

Estos intervalos se conocen como unilaterales.

9.2. Métodos para obtener intervalos

Entre los métodos discutidos en esta sección se consideran, el método de la cantidad pivotal, el método general y el Bayesiano.

9.2.1. Método de la cantidad pivotal

Definición 9.2. Cantidad pivotal. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$ y sea $Q = q(X_1, X_2, ..., X_n; \theta)$ una función de la muestra y de θ. Si Q tiene una distribución que no depende de θ se dice que es una cantidad pivotal.

Ejemplo 9.2. Si $X_1, X_2, ..., X_n$ es una muestra aleatoria de $X \sim N(\mu, 9)$ entonces, $Q_1 = \bar{X} - \mu$ y $Q_2 = \frac{\bar{X}-\mu}{\frac{3}{\sqrt{n}}}$ son cantidades pivotales porque $Q_1 \sim N(0, \frac{9}{n})$ y $Q_2 \sim N(0,1)$, pero $Q_3 = \frac{\bar{X}}{\mu}$ no es una cantidad pivotal porque $Q_3 \sim N(1, \frac{9}{n\mu^2})$.

Si se tiene una cantidad pivotal para $\tau(\theta)$, teóricamente se puede obtener una estimación por intervalo para esa función de θ.

Descripción del método de la cantidad pivotal.
Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$ y sea $Q = q(X_1, X_2, ..., X_n; \theta)$ una cantidad pivotal que tiene función de probabilidad f_Q, entonces existen q_1 y q_2 para cualquier δ, con $0 < \delta < 1$ tal que $P(q_1 < Q < q_2) = \delta$. Ahora, si para cada valor muestral posible $(x_1, x_2, ..., x_n)$, $\{q_1 < q(x_1, x_2, ..., x_n; \theta) < q_2\}$, si y solo si, $\{l(x_1, x_2, ..., x_n) < \tau(\theta) < u(x_1, x_2, ..., x_n)\}$, donde l y u son funciones no dependientes de θ; entonces (L, U) es un intervalo de confianza del $100\delta\,\%$ para $\tau(\theta)$. Aquí $L = l(X_1, X_2, ..., X_n)$ y $U = u(X_1, X_2, ..., X_n)$.

Comentarios

1. q_1 y q_2 son independientes de θ puesto que la distribución de Q lo es.

2. Para cualquier δ fijo existen muchos pares de valores q_1 y q_2 que pueden seleccionarse de manera que $P(q_1 < Q < q_2) = \delta$. De esta forma, diferentes pares de q_1 y q_2 producirán diferentes pares de l y u. Dada esta situación se debe seleccionar aquel par (q_1, q_2) que hagan a l y u muy cercanos. Por ejemplo si la longitud $L = u(x_1, x_2, ..., x_n) - l(x_1, x_2, ..., x_n)$ no es aleatorio, entonces se puede seleccionar aquel par que haga la longitud mínima; si la longitud del intervalo es aleatoria se puede seleccionar q_1 y q_2 que hagan la longitud promedio del intervalo más pequeña.

   En el ejemplo 9.1 se tiene también que $P(-1.75 < Z < 2.326) = 0.95$. Esto conduce a que
   $$\left(\bar{X} - 1.312, \bar{X} + 1.745\right)$$
   es también un intervalo de confianza del 95 % para μ. Pero la longitud de este intervalo es mayor que la del intervalo $(*)$. Luego se prefiere aquel.

3. El hecho esencial del método de la cantidad pivotal es que la desigualdad $\{q_1 < q(x_1, x_2, ..., x_n; \theta) < q_2)\}$ puede ser invertida o pivoteada en la de-

sigualdad $\{l(x_1, x_2, ..., x_n) < \tau(\theta) < u(x_1, x_2, ..., x_n)\}$ para cualquier valor muestral posible $(x_1, x_2, ..., x_n)$.

Una preocupación del estadístico está en cómo obtener cantidades pivotales para determinar intervalos de confianza. La siguiente conceptualización permite obtener algunas cantidades pivotales.

Definición 9.3. Parámetros de localización y de escala. Sea $\{f(x;\theta)|\theta \in \Theta\}$ una familia de densidades indexadas por θ, donde $-\infty < \Theta < \infty$.

El parámetro θ es un parámetro de localización para $f(x;\theta)$ si $f(x;\theta) = h(x-\theta)$ para alguna función de probabilidad $h(\cdot)$. Equivalentemente, θ es un parámetro de localización para $f(x;\theta)$ si la distribución de $X-\theta$ no depende de θ.

El parámetro θ es un parámetro de escala para $f(x;\theta)$ si $f(x;\theta) = \frac{1}{\theta}h(\frac{x}{\theta})$ para alguna función de probabilidad $h(\cdot)$. Equivalentemente, θ es un parámetro de escala para $f(x;\theta)$ si la distribución de $\frac{X}{\theta}$ es independiente de θ.

Los parámetros θ_1 y θ_2 se llaman de localización-escala si $f(x;\theta_1,\theta_2) = \frac{1}{\theta_2}h(\frac{x-\theta_1}{\theta_2})$.

Ejemplo 9.3. Si X es una variable aleatoria con con fdp $f_X(x) = e^{-(x-\eta)}, x > \eta$, entonces η es un parámetro de localización.

En efecto, sea $Y = X - \eta$ luego,

$$F_Y(y) = P(Y \leq y) = P(X - \eta \leq y) = P(X \leq y + \eta) = F_X(y + \eta).$$

Pero,

$$F_X(x) = \int_\eta^x e^{-(t-\eta)}dt = \int_0^{x-\eta} e^{-u}du = 1 - e^{-(x-\eta)},\ x > \eta.$$

Luego, $F_Y(y) = 1 - e^{-y}, y > 0$. Es decir que, $f_Y(y) = e^{-y}, y > 0$, que es la fdp de una distribución exponencial con parámetro 1 (no depende de η).

Ejemplo 9.4. Si X tiene una distribución exponencial con parámetro $\frac{1}{\lambda}$, entonces λ es un parámetro de escala.

En efecto, si $X \sim E\big(\frac{1}{\lambda}\big)$, entonces $F_X(x) = 1 - e^{-\frac{1}{\lambda}x}, x > 0$. Sea $Y = \frac{X}{\lambda}$ luego,

$$F_Y(y) = P(Y \leq y) = P\left(\frac{X}{\lambda} \leq y\right) = P(X \leq \lambda y) = F_X(\lambda y) = 1 - e^{-y}, y > 0$$

Nuevamente, $f_Y(y) = e^{-y}, y > 0$ que no depende de λ.

Ejemplo 9.5. Si $X \sim N(\mu, \sigma^2)$, entonces μ y σ son parámetros de localización-escala.

En efecto

$$f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{\sigma}\left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{2}\right)^2}\right] = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$$

donde $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$. Así que, μ y σ^2 son parámetros de localización-escala.

Teorema 9.1. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$ y suponga que $\hat{\theta}$ el estimador de máxima verosimilitud de θ existe, entonces

1. Si θ es un parámetro de localización, $Q = \hat{\theta} - \theta$ es una cantidad pivotal.
2. Si θ es un parámetro de escala, $Q = \frac{\hat{\theta}}{\theta}$ es una cantidad pivotal.

Teorema 9.2. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta_1,\theta_2)$, donde θ_1 y θ_2 son parámetros de localización-escala. Si los estimadores de máxima verosimilitud $\hat{\theta}_1$ de θ_1 y $\hat{\theta}_2$ de θ_2 existen, entonces $\frac{\hat{\theta}_1 - \theta_1}{\hat{\theta}_2}$ y $\frac{\hat{\theta}_2}{\theta_2}$ son cantidades pivotales.

Ejemplo 9.6. En el ejemplo 9.4, se probó que λ es un parámetro de escala en la fdp de la distribución exponencial con parámetro $\frac{1}{\lambda}$. Si $X_1, X_2, ..., X_n$ es una muestra aleatoria, entonces el MLE de λ es $\bar{X}$(verifíquelo) y por el teorema 9.1, $\frac{\bar{X}}{\lambda}$ es una cantidad pivotal. Así que dado δ, $0 < \delta < 0$, existen q_1 y q_2 con $q_1 < q_2$ tales que

$$\delta = P\left(q_1 < \frac{\bar{X}}{\lambda} < q_2\right) = P\left(\frac{1}{q_2} < \frac{\lambda}{\bar{X}} < \frac{1}{q_1}\right) = P\left(\frac{\bar{X}}{q_2} < \lambda < \frac{\bar{X}}{q_1}\right).$$

Esto es $\left(\frac{\bar{X}}{q_2}, \frac{\bar{X}}{q_1}\right)$ es un intervalo de confianza del $100\delta\,\%$ para λ.

Para escoger q_1 y q_2 se debe conocer la distribución de $\frac{\bar{X}}{\lambda}$. Sin embargo más adelante se tratará de resolver este problema.

Ejemplo 9.7. En el ejemplo 9.5 se demostró que μ y σ^2 son parámetros de localizacion-escala para la fdp de la distribución normal. Si $X_1, X_2, ..., X_n$ es una muestra aleatoria de esta distribución, se encontró en el capítulo anterior que $\bar{X}$

cantidades pivotales $\frac{n-1}{n\sigma^2}S^2$ y $\frac{\bar{X}-\mu}{\sqrt{\frac{n-1}{n}}S}$.

9.2.2. Método general

Si no se tiene una cantidad pivotal para determinar un intervalo para un parámetro θ, se puede obtener un estadístico (preferiblemente un estadístico suficiente o un estimador de máxima verosimilitud), con función de distribución que dependa de θ, pero no de otros parámetros desconocidos. Especificamente, sea $X_1, X_2, ..., X_n$ una muestra de $f(x;\theta)$ donde $\theta \in \Theta$ y Θ es un intervalo. Se desea estimar θ.

Sea $T = t(X_1, X_2, ..., X_n)$ un estadístico con fdp o fmp $g(t;\theta)$ y función de distribución $G(t;\theta)$. Supóngase además que para cada valor posible de θ se pueden encontrar valores $h_1(\theta)$ y $h_2(\theta)$ tales que,

$$P\big[h_1(\theta) < T < h_2(\theta)\big] = \delta. \tag{9.2}$$

Si T una variable aleatoria continua, defínanse las funciones $h_1(\theta)$ y $h_2(t)$ por

$$\int_{-\infty}^{h_1(\theta)} g(t;\theta)dt = G[h_1(\theta)] = \alpha_1 \quad \text{y} \quad \int_{h_2(\theta)}^{\infty} g(t;\theta)dt = G[h_2(\theta)] = \alpha_2$$

donde α_1 y α_2 son números fijos con $\alpha_1 > 0$, $\alpha_2 > 0$ y $\alpha_1 + \alpha_2 < 1$.

Sean h_1 y h_2 extrictamente monótonas(crecientes) como se observa en la figura 9.1 con $h_1(\theta) < h_2(\theta)$.(Las funciones h_1 y h_2 no son necesariamente lineales).

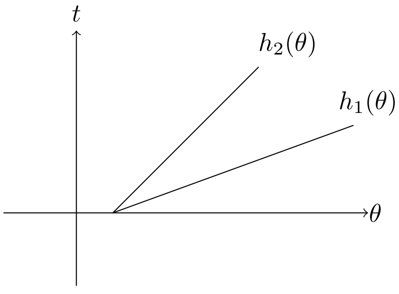


Figura 9.1: Funciones h_1 y h_2 para el método general

Entonces para cualquier valor posible t_0 de T le corresponden dos valores θ_L y θ_U que son funciones de t_0. Sean estos $\theta_L = \theta_L(t_0)$ y $\theta_U = \theta_U(t_0)$. Luego el intervalo (Θ_L, Θ_U) es un intervalo de confianza del $100(1-\alpha_1-\alpha_2)\,\%$ para θ. Ver figura 9.2.

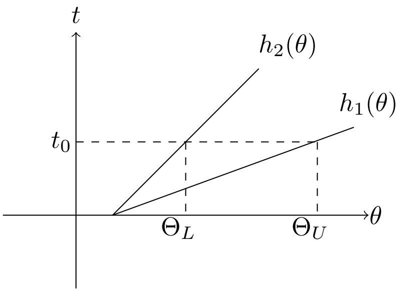


Figura 9.2: Intervalo para θ, método general

Al tomar un θ_0 en (Θ_L, Θ_U) se determina un intervalo $\big(h_1(\theta_0), h_2(\theta_0)\big)$ que contiene a t_0. Ver figura 9.3.

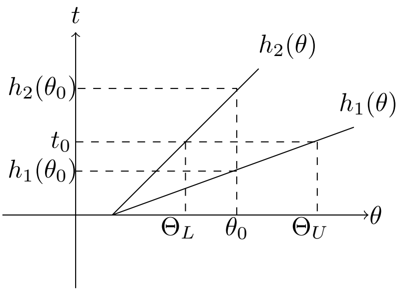


Figura 9.3: Intervalos para t y para θ, método general

Luego,

$$h_1(\theta_0) < t_0 < h_2(\theta_0) \text{ si y solo si } \theta_L(t_0) < \theta_0 < \theta_U(t_0).$$

Pero,

$$P_{\theta_0}[h_1(\theta_0) < t_0(X_1, X_2, \cdots .X_n) < h_2(\theta_0)] = 1 - \alpha_1 - \alpha_2.$$

Así que,

$$P_{\theta_0}[\theta_L(t_0)(X_1, X_2, \cdots .X_n) < \theta_0 < \theta_U(t_0)(X_1, X_2, \cdots .X_n)] = 1 - \alpha_1 - \alpha_2$$

Por tanto, $\big(\Theta_L(t_0), \Theta_U(t_0)\big)$ es un intervalo de confianza del $100(1-\alpha_1-\alpha_2)\,\%$ para θ. Los siguientes teoremas confirman el procedimiento descrito.

Teorema 9.3. Sea $X_1, X_2, ..., X_n$ una muestra de $f(x;\theta)$ y sea $T = t(X_1, X_2, ..., X_n)$ un estadístico, con función de distribución continua $G(t;\theta)$. Suponga además que, $h_1(\theta)$ y $h_2(\theta)$ son funciones que satisfacen

$$G_T[h_1(\theta)] = \alpha_1 \text{ y } G_T[h_2(\theta)] = 1-\alpha_2. \tag{9.3}$$

para cada $\theta \in \Theta$, Θ es un intervalo, donde $\alpha_1 > 0$, $\alpha_2 > 0$ y $\alpha_1 + \alpha_2 < 1$.

Sea t un valor observado de T. Si $h_1(\theta)$ y $h_2(\theta)$ son funciones crecientes de θ, entonces (θ_L, θ_U) es un intervalo para θ del $100\delta\,\%, \delta = 1-\alpha_1-\alpha_2$. Donde θ_L y θ_U son soluciones de

$$h_2(\theta_L) = t \text{ y } h_1(\theta_U) = t. \tag{9.4}$$

Ejemplo 9.8. Si $X_1, X_2, ..., X_n$ una muestra de $f(x;\theta) = \frac{1}{\theta}e^{-\frac{1}{\theta}x}, \theta > 0, 0 < x < \infty$ (distribución exponencial con $\lambda = \frac{1}{\theta}$), obtener un intervalo de confianza para θ.

Solución. Se busca un estadístico que tenga una distribucón conocida o que se pueda encontrar. Para esto se analiza la función de verosimilitud.

Como $f(x;\theta) = \frac{1}{\theta}e^{-\frac{1}{\theta}x}I_{(0,\infty}(x)$ entonces,

$$L(\theta) = \frac{1}{\theta^n}e^{-\frac{1}{\theta}\sum x_i}\Pi I_{(0,1)}(x_i).$$

Se observa que,

$$L(\theta) = g(\Sigma x_i;\theta)h(x_1, x_2, ..., x_n) \text{ con } h(x_1, x_2, ..., x_n) = \Pi I_{(0,1)}(x_i).$$

Es decir, que $T = \Sigma X_i$ es un estadístico suficiente para θ. Del ejemplo 6.5 se conoce que $\Sigma X_i \sim G\big(n, \frac{1}{\theta}\big)$, o sea que

$$f_T(t) = \frac{t^{n-1}e^{-\frac{1}{\theta}t}}{\theta^n\Gamma(n)}I_{(0,\infty)}(t).$$

Considérese el estadístico $S = \frac{2T}{\theta}$ entonces,

$$
\begin{aligned}
P(S \leq s) &= P\left(\frac{2T}{\theta} \leq s\right) = P\left(T \leq \frac{\theta}{2}s\right) \\
&= \int_0^{\frac{\theta}{2}s} \frac{t^{n-1}e^{-\frac{1}{\theta}t}}{\theta^n\Gamma(n)}dt = \int_0^s \frac{\left(\frac{\theta}{2}u\right)^{n-1}e^{-\frac{1}{2}u}\left(\frac{\theta}{2}\right)}{\theta^n\Gamma(n)}du \\
&= \int_0^s \frac{u^{n-1}e^{-\frac{1}{2}u}}{2^n\Gamma(n)}du.
\end{aligned}
$$

Esta es la función de distribución de una variable aleatoria chi-cuadrado con $2n$ grados de libertad.

Sean $\alpha_1 > 0$ y $\alpha_2 > 0$ tal que $\alpha_1 + \alpha_2 < 1$ entonces por 9.3, $\alpha_1 = G_T[h_1(\theta)]$, es decir, $\alpha_1 = P[T \leq h_1(\theta)] = P[\frac{2T}{\theta} \leq \frac{2h_1(\theta)}{\theta}]$, esto significa que $\frac{2h_1(\theta)}{\theta} = \chi^2_{2n}(\alpha_1)$. O sea que, $h_1(\theta) = \frac{\theta}{2}\chi^2_{2n}(\alpha_1)$ que es una función creciente de θ. Sea esta $h_1(\theta_U)$.

También, $1 - \alpha_2 = G_T[h_2(\theta)]$, o sea que, $1 - \alpha_2 = P[\frac{2T}{\theta} \leq \frac{2h_2(\theta)}{\theta}]$, lo que implica que $\frac{2h_2(\theta)}{\theta} = \chi^2_{2n}(1-\alpha_2)$, o $h_2(\theta) = \frac{\theta}{2}\chi^2_{2n}(1-\alpha_2)$, que es una función creciente de θ. Sea esta $h_2(\theta_L)$.

Por 9.4 $h_1(\theta_U) = \sum x_i$, luego $\theta_U = \frac{2\sum x_i}{\chi^2_{2n}(\alpha_1)}$, y también $h_2(\theta_L) = \sum x_i$, luego $\theta_L = \frac{2\sum x_i}{\chi^2_{2n}(1-\alpha_2)}$. Por lo tanto, un intervalo del 100 % de confianza para θ es

$$\left(\frac{2\sum x_i}{\chi^2_{2n}(1-\alpha_2)}, \frac{2\sum x_i}{\chi^2_{2n}(\alpha_1)}\right).$$

Observe que se ha resuelto completamenta el ejemplo 9.6.

Teorema 9.4. Sea $X_1, X_2, ..., X_n$ una muestra de $f(x;\theta)$ y sea $T = t(X_1, X_2, ..., X_n)$ un estadístico con función de distribución continua $G(t;\theta)$. Sea $\alpha_2 + \alpha_2 = 1 - \delta$ con $\alpha_1 > 0$, $\alpha_2 > 0$ valores fijos y $0 < 1 - \delta < 1$. Supóngase además que, para cada t las funciones θ_L y θ_U se pueden definir por

1. Si $G(t;\theta)$ es una función creciente de θ para cada t, se define θ_L y θ_U por

$$1 - G(\theta_U) = \alpha_2, \quad \text{y} \quad G(\theta_L) = \alpha_1.$$

2. Si $G(t;\theta)$ es una función decreciente de θ para cada t, se define θ_L y θ_U por

$$G(\theta_U) = \alpha_1, \quad \text{y} \quad 1 - G(\theta_L) = \alpha_2.$$

Entonces el intervalo ($\theta_L(T), \theta_U(T)$) es un intervalo del $100\delta\,\%$ de confianza para θ.

Teorema 9.5. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población discreta con fdp $f(x;\theta)$, y sea $T = t(X_1, X_2, ..., X_n)$ un estadístico, con función de distribución $G(t;\theta)$. Sea $\alpha_2 + \alpha_2 = 1 - \delta$ con $\alpha_1 > 0$, $\alpha_2 > 0$ valores fijos y $0 < 1 - \delta < 1$. Supóngase, además que para cada t las funciones θ_L y θ_U se pueden definir por

1. Si $G(t;\theta)$ es una función creciente de θ para cada t, se define θ_L y θ_U por
$$1 - G(\theta_U) = \alpha_2, \quad \text{y} \quad G(\theta_L) = \alpha_1.$$
2. Si $G(t;\theta)$ es una función decreciente de θ para cada t, se define θ_L y θ_U por
$$G(\theta_U) = \alpha_1, \quad \text{y} \quad 1 - G(\theta_L) = \alpha_2.$$

Entonces el intervalo ($\theta_L(T), \theta_U(T)$) es un intervalo del $100\delta\,\%$ de confianza para θ.

Es de anotar que en ausencia de información acerca de α_1 y α_2, es común escoger $\alpha_1 = \alpha_2 = \frac{1-\delta}{2}$, aunque esto no produce siempre intervalos óptimos.

Ejemplo 9.9. Sea $X_1, X_2, ..., X_n$ una muestra de $f(x;\theta) = \frac{2x}{\theta^2} I_{(0,\theta)}(x),\ \theta > 0$. Obtener un intervalo para θ.

Solución. Como en el ejemplo anterior se encuentra un estadístico que tenga función de distribución conocida o se pueda conocer. Para esto se considera $L(\theta)$. De esta manera,

$$L(\theta) = \Pi\left(\frac{2x_i}{\theta^2}\right) I_{(0,\theta)}(x_i) = \frac{2^n}{\theta^{2n}} \Pi x_i I_{(0,\theta)}(x_i) = \frac{2^n}{\theta^{2n}} (\Pi x_i) \Pi I_{(0,\theta)}(x_i).$$

Pero,

$$\begin{aligned} \Pi I_{(0,\theta)}(x_i) &= 1 \ \text{ si } 0 < x_i < \theta, \ \text{ para toda } i \\ &= 1 \text{ si } 0 < x_{(1)} \ \text{ y } \ x_{(n)} < \theta \\ &= 1 \text{ si } 0 < x_{(1)} < x_{(n)} \ \text{ y } \ 0 < x_{(n)} < \theta \\ &= I_{(0,y_n)}(y_1) I_{(0,\theta)}(y_n), \text{ donde } y_1 = x_{(1)} \text{ y } y_n = x_{(n)}. \end{aligned}$$

Luego,

$$L(\theta) = \frac{2^n}{\theta^{2n}} (\Pi x_i) I_{(0,y_n)}(y_1) I_{(0,\theta)}(y_n) = g(x_{(n)};\theta) h(x_1, x_2, ..., x_n).$$

Por tanto $T = X_{(n)} = \text{máx}(X_1, X_2, ..., X_n)$ es un estadístico suficiente para θ.

Ahora,

$$F_X(x) = \int_0^x \frac{2t}{\theta^2} dt = \frac{x^2}{\theta^2},\ 0 < x < \theta.$$

Luego, por el teorema 6.5 se tiene que,

$$G_{X_{(n)}}(y) = [F_X(y)]^n = \left[\frac{y^2}{\theta^2}\right]^n,\ 0 < y < \theta.$$

Puede comprobarse que $G_{X_{(n)}}$ es una función decreciente de θ. Entonces por el teorema 9.4, para $\alpha_1 > 0$ y $\alpha_2 > 0$ con $\alpha_1 + \alpha_2 < 1$ se tiene que,

1. De $G[h_2(\theta)] = \alpha_1$ implica que $\left[\frac{h_2(\theta)}{\theta}\right]^{2n} = \alpha_1$, si y solo si, $h_2(\theta) = \theta \sqrt[2n]{\alpha}_1$, que es una función creciente de θ. Pero $h_2(\theta) = x_{(n)}$, luego $x_{(n)} = \theta \sqrt[2n]{\alpha}_1$ o $\theta_U = \frac{x_{(n)}}{\sqrt[2n]{\alpha}_1}$.

2. De $1 - G[h_1(\theta)] = \alpha_2$ implica que $\left[\frac{h_1(\theta)}{\theta}\right]^{2n} = 1 - \alpha_2$, luego $h_1(\theta) = \theta \sqrt[2n]{1-\alpha_2}$, que es creciente en θ. De $h_1(\theta) = x_{(n)}$ y de la expresión anterior, $x_{(n)} = \theta \sqrt[2n]{1-\alpha_2}$ o $\theta_L = \frac{x_{(n)}}{\sqrt[2n]{1-\alpha_2}}$.

Así que un intervalo de confianza para θ del $100\delta\,\%, \delta = 1 - \alpha_1 - \alpha_2$ es

$$\left(\frac{X_{(n)}}{\sqrt[2n]{1-\alpha_2}}, \frac{X_{(n)}}{\sqrt[2n]{\alpha_1}}\right).$$

9.2.3. Método bayesiano

Sea $X_1, X_2, ..., X_n$ una muestra de $f(x;\theta)$ y suponga que $\theta \in \Theta$, donde Θ es una variable aleatoria. Del capítulo anterior se conoce que la distribución posterior de Θ dada la muestra es

$$f_{\Theta|X_1,X_2,...,X_n}(\theta|x_1, x_2, ..., x_n) = \frac{f(x_1, x_2, ..., x_n)g(\theta)}{\int f(x_1, x_2, ..., x_n)g(\theta)d\theta},$$

donde g es la densidad a priori de θ. Luego para δ dado, $0 < \delta < 1$, un intervalo de confianza Bayesiano para θ es (T_1, T_2) donde los valores observados t_1 y t_2 satisfacen,

$$\int_{t_1}^{t_2} f_{\Theta|X_1,X_2,...,X_n}(\theta|x_1, x_2, ..., x_n)d\theta = \delta. \tag{9.5}$$

En la práctica se escogen T_1 y T_2 de manera que $t_2 - t_1$ sea la más pequeña.

Ejemplo 9.10. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu, \sigma^2)$. Suponga que σ es conocido y que $\mu \sim N(\eta, \xi^2)$ con η y ξ conocidos. Se desea obtener un intervalo de confianza del $100\delta\,\%$ para μ.

Solución. Con los datos del problema,

$$f_X(x;\mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, x \in R, \mu \in R, \sigma > 0$$

y

$$g(\mu) = \frac{1}{\sqrt{2\pi}\xi} e^{-\frac{1}{2\xi^2}(\mu-\eta)^2}, \eta \in R, \xi > 0, \text{ pero fijos.}$$

Entonces

$$\begin{aligned}
\Pi f(x_i;\mu) g(\mu) &= \left(\frac{1}{\sqrt{2\pi}}\right)^{n+1} \frac{1}{\xi\sigma^n} e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2 - \frac{1}{2\xi^2}(\mu-\eta)^2} \\
&= k_1 e^{-\frac{1}{2\xi^2\sigma^2}[\xi^2\sum(x_i-\mu)^2 + \sigma^2(\mu-\eta)^2]} = k_1 e^{-\frac{1}{2\xi^2\sigma^2} h(\mu)}
\end{aligned}$$

donde

$$\begin{aligned}
h(\mu) &= \xi^2 \Sigma(x_i - \mu)^2 + \sigma^2(\mu - \eta)^2 \\
&= \xi^2 \left[\Sigma x_i^2 - 2\mu\Sigma x_i + n\mu^2\right] + \sigma^2[\mu^2 - 2\mu\eta + \eta^2] \\
&= \xi^2 \Sigma x_i^2 + \sigma^2\eta^2 + (n\xi^2 + \sigma^2)\mu^2 - 2\big(\xi^2\Sigma x_i + \sigma^2\eta\big)\mu \\
&= \xi^2 \Sigma x_i^2 + \sigma^2\eta^2 + (n\xi^2 + \sigma^2)\left[\mu^2 - 2\frac{\xi^2\Sigma x_i + \sigma^2\eta}{n\xi^2 + \sigma^2}\mu\right] \\
&= \xi^2 \Sigma x_i^2 + \sigma^2\eta^2 + (n\xi^2 + \sigma^2)\left[\mu - \frac{\xi^2\Sigma x_i + \sigma^2\eta}{n\xi^2 + \sigma^2}\right]^2 - \frac{(\xi^2\Sigma x_i + \sigma^2\eta)^2}{n\xi^2 + \sigma^2} \\
&= \xi^2 \Sigma x_i^2 + \sigma^2\eta^2 - \frac{(\xi^2\Sigma x_i + \sigma^2\eta)^2}{n\xi^2 + \sigma^2} + (n\xi^2 + \sigma^2)\left[\mu - \frac{\xi^2\Sigma x_i + \sigma^2\eta}{n\xi^2 + \sigma^2}\right]^2.
\end{aligned}$$

La suma de los tres primeros términos de $h(\mu)$ no dependen de μ y es un número, sea este ω, entonces

$$h(\mu) = \omega + (n\xi^2 + \sigma^2)\left[\mu - \frac{\xi^2\Sigma x_i + \sigma^2\eta}{n\xi^2 + \sigma^2}\right]^2.$$

Luego,

$$\begin{aligned}\Pi f(x_i;\mu)g(\mu) &= k_1 e^{-\frac{1}{2\xi^2\sigma^2}\left[\omega+(n\xi^2+\sigma^2)\left(\mu-\frac{\xi^2\sum x_i+\sigma^2\eta}{n\xi^2+\sigma^2}\right)^2\right]}\\ &= k_1 e^{-\frac{\omega}{2\xi^2\sigma^2}} e^{-\frac{n\xi^2+\sigma^2}{2\xi^2\sigma^2}\left(\mu-\frac{\xi^2\sum x_i+\sigma^2\eta}{n\xi^2+\sigma^2}\right)^2}\\ &= k_2 e^{-\frac{1}{2\frac{\xi^2\sigma^2}{n\xi^2+\sigma^2}}\left(\mu-\frac{\xi^2\sum x_i+\sigma^2\eta}{n\xi^2+\sigma^2}\right)^2}.\end{aligned}$$

Sea $a=\frac{\xi^2\sum x_i+\sigma^2\eta}{n\xi^2+\sigma^2}$ y $b^2=\frac{\xi^2\sigma^2}{n\xi^2+\sigma^2}$, entonces $\Pi f(x_i;\mu)g(\mu)=k_2e^{-\frac{1}{2b^2}(\mu-a)^2}$.

Por lo tanto,

$$f(\mu|x)=\frac{k_2e^{-\frac{1}{2b^2}(\mu-a)^2}}{\int k_2e^{-\frac{1}{2b^2}(\mu-a)^2}d\mu}=\frac{e^{-\frac{1}{2b^2}(\mu-a)^2}}{\int e^{-\frac{1}{2b^2}(\mu-a)^2}d\mu}.$$

En esta expresión al multiplicar numerador y denominador por $\frac{1}{\sqrt{2\pi}b}$, la integral del denominador es 1 (¿por que?), luego

$$f(\mu|x)=\frac{1}{\sqrt{2\pi}b}e^{-\frac{1}{2b^2}(\mu-a)^2}.$$

que es la función de densidad de una distribución normal con media a y varianza b^2.

Si δ es dado, $0<\delta<1$ al resolver $\int_{q_1}^{q_2} f(\mu|x)d\mu=\delta$ se obtiene un intervalo bayasiano del $100\delta\,\%$ para μ. Esto es equivalente a resolver $P(q_1<\mu<q_2)=\delta$, o, $P(\frac{q_1-a}{b}<\frac{\mu-a}{b}<\frac{q_2-a}{b})=\delta$.

Luego, $\Phi(\frac{q_2-a}{b})-\Phi(\frac{q_1-a}{b})=\delta$.

En la próxima sección se prueba que un buen intervalo se obtiene cuando

$$\frac{q_2-a}{b}=z_{\frac{1+\delta}{2}} \quad \text{y} \quad \frac{q_1-a}{b}=-z_{\frac{1+\delta}{2}}$$

Es decir, cuando $q_2=a+bz_{\frac{1+\delta}{2}}$ y $q_1=a-bz_{\frac{1-\delta}{2}}$. De manera que un intervalo del $100\delta\,\%$ de confianza para μ es

$$\left(\frac{\xi^2\Sigma x_i+\sigma^2\eta}{n\xi^2+\sigma^2}-z_{\frac{1+\delta}{2}}\sqrt{\frac{\xi^2\sigma^2}{n\xi^2+\sigma^2}},\frac{\xi^2\Sigma x_i+\sigma^2\eta}{n\xi^2+\sigma^2}+z_{\frac{1+\delta}{2}}\sqrt{\frac{\xi^2\sigma^2}{n\xi^2+\sigma^2}}\right).$$

9.3. Intervalos de confianza para muestras de la distribución normal

En esta sección se consideran intervalos de confianza para μ y para σ^2, además para la diferencia de dos medias y el cociente de dos varianzas de muestras de la distribución normal.

a) **Intervalos para la media μ.**

Sea $X_1, X_2, ..., X_n$ una muestra de una distribución normal $X \sim N(\mu, \sigma^2)$. Se presentan los casos cuando σ es conocido y cuando es desconocido.

-Si σ^2 es conocido, sea esta σ_0^2, entonces $Q = \dfrac{\bar{X} - \mu}{\frac{\sigma_0}{\sqrt{n}}}$ tiene distribución normal estándar, es decir Q es una cantidad pivotal y $f_Q(q) = \phi(q) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}q^2}$. Luego, dado δ fijo con $0 < \delta < 1$, existen q_1 y q_2 con $q_1 < q_2$ tales que

$$P(q_1 < Q < q_2) = \delta.$$

Observe que

$$\begin{aligned}\{q_1 < Q < q_2\} \ \text{ si y solo si } &\left\{q_1 < \frac{\bar{X} - \mu}{\frac{\sigma_0}{\sqrt{n}}} < q_2\right\}, \\ \text{si y solo si } &\left\{\bar{X} - q_2\frac{\sigma_0}{\sqrt{n}} < \mu < \bar{X} - q_1\frac{\sigma_0}{\sqrt{n}}\right\}.\end{aligned}$$

Por tanto, $P\left(\bar{X} - q_2\dfrac{\sigma_0}{\sqrt{n}} < \mu < \bar{X} - q_1\dfrac{\sigma_0}{\sqrt{n}}\right) = \delta$ y esto significa que $\left(\bar{X} - q_2\dfrac{\sigma_0}{\sqrt{n}}, \bar{X} - q_1\dfrac{\sigma_0}{\sqrt{n}}\right)$ es un intervalo de confianza del $100\delta\,\%$ para μ.

La longitud del intervalo es $L = \dfrac{\sigma_0}{\sqrt{n}}(q_2 - q_1)$ la cual será mínima al seleccionar q_1 y q_2 bajo la restricción,

$$\delta = P(q_1 < Q < q_2) = \Phi(q_2) - \Phi(q_1) = \int_{q_1}^{q_2} \phi(t)dt.$$

Derivando L y esta última expresión con respecto a q_1 se tiene que $\phi(q_2) = \phi(q_1)$ y esto significa que $q_1 = q_2$ o $q_1 = -q_2$, con $q_2 > 0$. Pero q_1 no puede ser igual a q_2 (¿por qué?), luego $q_1 = -q_2$.

Así es que, $\delta = P(-q_2 < Q < q_2) = \Phi(q_2) - \Phi(-q_2) = 2\Phi(q_2) - 1$, o sea que $\Phi(q_2) = \dfrac{1+\delta}{2}$ lo que implica que $q_2 = z_{\frac{1+\delta}{2}}$.

Por tanto, un intervalo del $100\delta\,\%$ para μ con σ conocido es

$$\left(\bar{X} - \frac{\sigma_0}{\sqrt{n}} z_{\frac{1+\delta}{2}},\ \bar{X} + \frac{\sigma_0}{\sqrt{n}} z_{\frac{1+\delta}{2}}\right). \tag{9.6}$$

-Si σ^2 es desconocido, se sabe de 7.18 que $T = \dfrac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ tiene distribución t con $n-1$ grados de libertad. Entonces para δ dado, con $0 < \delta < 1$, existe t [1] tal que $P\big(-t < \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} < t\big) = \delta$ y esto significa que $P\big(\bar{X} - t\frac{S}{\sqrt{n}} < \mu < \bar{X} + t\frac{S}{\sqrt{n}}\big) = \delta$. Pero t es el $t_{\frac{1+\delta}{2}}$ de la distribución t con $n-1$ grados de libertad, es decir que un intervalo de confianza del $100\delta\,\%$ para μ es

$$\left(\bar{X} - \frac{S}{\sqrt{n}} t_{\frac{1+\delta}{2}}(n-1),\ \bar{X} + \frac{S}{\sqrt{n}} t_{\frac{1+\delta}{2}}(n-1)\right). \tag{9.7}$$

Es de anotar que los intervalos obtenidos en 9.6 y en 9.7 no son únicos, sin embargo son los de menor longitud entre todos los intervalos posibles.

b) **Intervalos para la varianza σ^2.**

Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu, \sigma^2)$. Se obtendrán intervalos para σ^2 cuando μ es conocida y cuando es desconocida.

-Si $\mu = \mu_0$, con μ_0 conocido, se conoce de 7.15 que $U = \displaystyle\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2$ tiene una distribución chi-cuadrado con n grados de libertad ($U \sim \chi^2_n$), es decir, U es una cantidad pivotal. Luego, dado δ fijo con $0 < \delta < 1$, existen q_1 y q_2, con $q_1 < q_2$ tales que

$$\delta = P\left(q_1 < \sum_{1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 < q_2\right) = P\left(\frac{1}{q_2} < \frac{\sigma^2}{\sum_1^n (X_i - \mu)^2} < \frac{1}{q_1}\right).$$

[1] En realidad existen muchos t_1 y t_2 tales que $P(t_1 < T < t_2) = \delta$. Pero por tener T una distribución simétrica alrededor de 0, los intervalos óptimos se obtienen para $t_1 = -t_2$ y $t_2 > 0$.

De aquí se obtiene que un intervalo de confianza del 100δ % para la varianza es

$$\left(\frac{\Sigma_1^n(x_i-\mu_0)^2}{q_2},\ \frac{\Sigma_1^n(x_i-\mu_0)^2}{q_1}\right). \tag{9.8}$$

Se pueden escoger q_1 y q_2 de manera que

$$P(U<q_1)=P(U>q_1)=\frac{1-\delta}{2}.$$

Por lo tanto

$$q_1=\chi^2_{n(\frac{1-\delta}{2})} \quad \text{y} \quad q_2=\chi^2_{n(\frac{1+\delta}{2})}.$$

Luego, un intervalo del 100δ % para σ^2 cuando μ es conocida es,

$$\left(\frac{\Sigma_1^n(x_i-\mu_0)^2}{\chi^2_{n(\frac{1+\delta}{2})}},\ \frac{\Sigma_1^n(x_i-\mu_0)^2}{\chi^2_{n(\frac{1-\delta}{2})}}\right). \tag{9.9}$$

-Si μ es desconocido, del teorema 7.10 se tiene que $U=\dfrac{(n-1)S^2}{\sigma^2}$ tiene una distribución χ^2 con $n-1$ grados de libertad. Así que dado δ, $0<\delta<1$, existen q_1 y q_2 tales que $P(q_1<U<q_2)=\delta$; es decir, $P\left(q_1<\frac{(n-1)S^2}{\sigma^2}<q_2\right)=\delta$. Así que

$$\delta=P\left(\frac{1}{q_2}<\frac{\sigma^2}{(n-1)S^2}<\frac{1}{q_1}\right)=P\left(\frac{(n-1)S^2}{q_2}<\sigma^2<\frac{(n-1)S^2}{q_1}\right).$$

Luego, un intervalo de confianza del 100δ % para la varianza es

$$\left(\frac{(n-1)S^2}{q_2},\ \frac{(n-1)S^2}{q_1}\right). \tag{9.10}$$

Donde q_1 y q_2 se obtienen como se hizo anteriormente. De modo que

$$q_1=\chi^2_{(n-1)(\frac{1-\delta}{2})} \quad \text{y} \quad q_2=\chi^2_{(n-1)(\frac{1+\delta}{2})}.$$

Es decir, un intervalo del 100δ % para σ^2 cuando μ es desconocida es,

$$\left(\frac{(n-1)S^2}{\chi^2_{(n-1)(\frac{1+\delta}{2})}},\ \frac{(n-1)S^2}{\chi^2_{(n-1)(\frac{1-\delta}{2})}}\right). \tag{9.11}$$

Ejemplo 9.11. Del índice de variación diaria de las 38 acciones de mayor rendimiento en Colombia (El Tiempo 28/07/12), se encontró que $\bar{x}=1.215$ y $s^2=5.209$. Si se supone que los datos siguen una distribución normal, a.) un intervalo de confianza del 95 % para la media es $(1.215\text{-}\frac{2.28}{\sqrt{38}}{*}2.03, 1.215+\frac{2.28}{\sqrt{38}} * 2.03)$. Esto es (0.464, 1.966).

b.) un intervalo de confianza del 95 % para la varianza es $(\frac{37*5.209}{47}, \frac{37*5.209}{16.8})$. Esto es (4.10, 11.47).

c) **Intervalos para la diferencia de dos medias**.

En este libro se consideran dos muestras de dos distribuciones normales que son independientes.

Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu_1, \sigma^2)$ y $Y_1, Y_2, ..., Y_m$ una muestra aleatoria de $Y \sim N(\mu_2, \sigma^2)$. Si las muestras son independientes se desea obtener un intervalo de confianza del $100\delta\,\%$ para $\mu_2 - \mu_1$.

Del segundo corolario del teorema 7.8 se sabe que $\bar{Y}-\bar{X} \sim N\left(\mu_2 - \mu_1, \frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right)$.

Por otro lado del teorema 7.10 se infiere que,

$$\frac{\Sigma_1^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1} \quad \text{y} \quad \frac{\Sigma_1^m (Y_i - \bar{Y})^2}{\sigma^2} \sim \chi^2_{m-1}.$$

Así que del teorema 7.9,

$$\frac{\Sigma_1^n (X_i - \bar{X})^2}{\sigma^2} + \frac{\Sigma_1^m (Y_i - \bar{Y})^2}{\sigma^2} \sim \chi^2_{n+m-2}.$$

Por tanto del teorema 7.11 se tiene que,

$$T = \frac{\frac{\bar{Y}-\bar{X}-(\mu_2-\mu_1)}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}}}{\frac{1}{\sigma}\sqrt{\frac{\sum\limits_1^n (X_i-\bar{X})^2 + \sum\limits_1^m (Y_i-\bar{Y})^2}{m+n-2}}} = \frac{\bar{Y} - \bar{X} - (\mu_2 - \mu_1)}{\sqrt{\frac{1}{m} + \frac{1}{n}}\sqrt{\frac{\sum\limits_1^n (X_i-\bar{X})^2 + \sum\limits_1^m (Y_i-\bar{Y})^2}{m+n-2}}}$$

tiene una distribución t con $m + n - 2$ grados de libertad.

Para δ dado, con $0 < \delta < 1$, existe t tal que $P(-t < T < t) = \delta$, donde $t = t_{(m+n-2, \frac{1+\delta}{2})}$.

Es decir, $P\left(-t < \frac{\bar{Y}-\bar{X}-(\mu_2-\mu_1)}{\sqrt{(\frac{1}{m}+\frac{1}{n})S_p^2}} < t\right) = \delta$ y por tanto,

$$P\left(\bar{Y}-\bar{X}-t\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)S_p^2} < \mu_2-\mu_1 < \bar{Y}-\bar{X}+t\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)S_p^2}\right) = \delta$$

donde $S_p^2 = \dfrac{\Sigma_1^n(X_i-\bar{X})^2+\Sigma_1^m(Y_i-\bar{Y})^2}{m+n-2}$.

De esta manera, un intervalo del $100\delta\,\%$ para $\mu_2-\mu_1$ es

$$\left(\bar{Y}-\bar{X}-t\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)S_p^2},\ \bar{Y}-\bar{X}+t\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)S_p^2}\right). \tag{9.12}$$

Cabe recordar que existen muchos pares (t_1,t_2) tales que $P(t_1 < T < t_2) = \delta$.

Ejemplo 9.12. Teniendo en cuenta los datos de la variación diaria del ejercicio 5 del capítulo 1 y de la variación del ejemplo 9.8, obtenga un intervalo de confianza del 90 % para la diferencia de las medias.

Solución. Para los datos del ejercicio 5 del primer capítulo, n=39, $\bar{x}$=-0.487 y s_x^2=1.406. Del ejemplo 9.8, m=38,$\bar{y}$=1.215 y s_y^2=5.209. Luego $\bar{y}-\bar{x}$=1.702, $\sqrt{(\frac{1}{m}+\frac{1}{n})s_p^2}$= 0.413 y para δ=0.90, $t_{(n+m-2,\frac{1+\delta}{2})} = t_{(75,\ 0.95)}$=1.66.

De manera que un intervalo del 90 % de confianza para la diferencia de medias es (1.016, 2.388).

d) **Intervalos para el cociente de dos varianzas**.

Sea $X_1,X_2,...,X_n$ una muestra aleatoria de $X \sim N(\mu_1,\sigma_1^2)$ y $Y_1,Y_2,...,Y_m$ una muestra aleatoria de $Y \sim N(\mu_2,\sigma_2^2)$. Si las muestras son independientes se desea obtener un intervalo de confianza del $100\delta\,\%$ para $\frac{\sigma_1^2}{\sigma_2^2}$.

Del teorema 7.12 se sabe que,

$$U = \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} = \frac{\sigma_1^2 S_2^2}{\sigma_2^2 S_1^2} \sim F(m-1,n-1).$$

Dado δ, $0 < \delta < 1$, existen u_1 y u_2 tales que $P(u_1 < U < u_2) = \delta$. Luego $P\left(u_1 < \frac{\sigma_1^2 S_2^2}{\sigma_2^2 S_1^2} < u_2\right) = \delta$ y esto significa que $P\left(\frac{S_1^2}{S_2^2}u_1 < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}u_2\right) = \delta$.

Si se escogen u_1 y u_2 con la condición $P(U \le u_1) = P(U > u_2) = \frac{1-\delta}{2}$, resulta que $u_1 = F_{(m-1,n-1,\ \frac{1-\delta}{2})}$ y $u_2 = F_{(m-1,n-1,\ \frac{1+\delta}{2})}$.

De manera que un intervalo de confianza del $100\delta\,\%$ para $\frac{\sigma_1^2}{\sigma_2^2}$ es

$$\left(F_{(m-1,n-1,\ \frac{1-\delta}{2})}\frac{S_1^2}{S_2^2}, F_{(m-1,n-1,\ \frac{1+\delta}{2})}\frac{S_1^2}{S_2^2}\right). \tag{9.13}$$

En el ejemplo anterior un intervalo de confianza para σ_1^2/σ_2^2 es (0.138, 0.526).

9.4. Intervalos para muestras de distribuciones continuas

No siempre es posible encontrar una cantidad pivotal basada en el estimador de máxima verosimilitud, pero para una muestra de una distribución continua con un parámetro desconocido, al menos una cantidad pivotal se puede obtener usando el teorema de la transformación de la probabilidad integral.

Teorema 9.6. Transformación de probabilidad integral. Si X es una variable aleatoria con función de distribución $F_X(x)$ continua, entonces $U = F_X(X)$ tiene distribución uniforme en (0,1) y por tanto $-\ln F_X(X;\theta)$ tiene distribución exponencial con parámetro 1.

Demostración. Se sabe que si Y tiene tiene una distribución uniforme en el intervalo (0,1) entonces $F_Y(y) = y$, para $0 < y < 1$. De esta manera,

$F_U(u) = P(U \le u) = P(F_X(X) \le u) = P(X \le F_X^{-1}(u)) = F_X(F_X^{-1}(u)) = u$. Es decir, U tiene distribución uniforme en $(0,1)$.

Por otro lado, $P(-\ln U \ge u) = P(\ln U \le -u) = P(U \le e^{-u}) = e^{-u}$. Esto significa que $P(-\ln U < u) = 1 - e^{-u}$ que es la función de distribución de una distribución exponencial con parámetro 1.

Además, para una muestra $X_1, X_2, ..., X_n$ de una distribución exponencial con parámetro θ se puede demostrar que $-2\sum_{i=1}^{n} \ln F_i(X;\theta)$ tiene una distribución chi-cuadrado con $2n$ grados de libertad y equivalentemente $-\sum_{i=1}^{n} \ln F_i(X;\theta)$ tiene una distribución gamma con parámetros n y 1. Luego para δ dado con $0 < \delta < 1$ se tiene que

$$P\left(\chi^2_{2n(\frac{1-\delta}{2})} < -2\sum_{i=1}^{n} \ln F_i(X;\theta < \chi^2_{2n(\frac{1+\delta}{2})}\right) = \delta. \tag{9.14}$$

Pivoteando esta desigualdad se genera un intervalo de confianza para θ. Si $F(x,\theta)$ es una función monótona creciente se puede verificar también que $1-F(X,\theta)$ tiene una distribución uniforme en (0,1) y $-2\sum_{i=1}^{n}[1-\ln F_i(X;\theta)] \sim \chi^2_{2n}$.

Ejemplo 9.13. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de X, la cuál tiene distribución de Pareto con parámetros 1 y k. Obtener un intervalo de confianza del $100\delta\,\%$ para k.

Solución. La función de distribución de X es $F_X(x;k) = 1-(1+x)^{-k}$, $x > 0$. Entonces $-2\sum_{i=1}^{n}[1-\ln F_i(X;k)] = -2\sum_{i=1}^{n}\ln(1+X_i)^{-k} = 2k\sum_{i=1}^{n}\ln(1+X_i)$. Así que para δ dado $0 < \delta < 1$,

$$P\left(\chi^2_{2n(\frac{1-\delta}{2})} < 2k\sum_{i=1}^{n} \ln(1+X_i) < \chi^2_{2n(\frac{1+\delta}{2})}\right) = \delta.$$

Es decir, que un intervalo de confianza del $100\delta\,\%$ para k es

$$\left(\frac{\chi^2_{2n(\frac{1-\delta}{2})}}{2\sum_{i=1}^{n}\ln(1+X_i)}, \frac{\chi^2_{2n(\frac{1+\delta}{2})}}{2\sum_{i=1}^{n}\ln(1+X_i)}\right).$$

Por ejemplo, si el tamaño de muestra es 5, un intervalo de confianza del 90 % para k es $\left(\frac{\chi^2_{10}(0.05)}{2\sum_{i=1}^{5}\ln(1+X_i)}, \frac{\chi^2_{10}(0.95)}{2\sum_{i=1}^{5}\ln(1+X_i)}\right)$. Es decir, que una estimación del 90 % para k es

$$\left(\frac{1.97}{\sum_{i=1}^{5}\ln(1+x_i)}, \frac{9.15}{\sum_{i=1}^{5}\ln(1+x_i)}\right).$$

9.5. Intervalos de confianza aproximados

A veces es posible encontrar una sucesión de estimadores $T_n = t_n(X_1, X_2, ..., X_n)$ de θ en $f(X;\theta)$ los cuales están distribuidos asintóticamente normal alrededor de θ; es decir, T_n está distribuido aproximadamente de forma normal con media θ y varianza $\sigma_n^2(\theta)$. En particular, para muestras grandes el estimador de máxima verosimilitud de θ, T_n es aproximadamente normal con media θ y varianza

$$\sigma_n^2(\theta) = \frac{1}{nE[\frac{\partial}{\partial\theta}\ln f(X;\theta)]^2} = \frac{-1}{nE[\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta)]}.$$

De esta forma, $\frac{T_n-\theta}{\sigma_n(\theta)}$ se puede tratar como una cantidad pivotal.

Ejemplo 9.14. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una distribución exponencial con parámetro λ. Obtenga un intervalo de confianza del $100\delta\,\%$ para λ.

Solución. Del ejemplo 8.9, se sabe que el estimador de máxima verosimilitud de λ es $\dfrac{1}{\bar{X}_n}$ y como $\sigma_n^2(\lambda) = \frac{\lambda^2}{n}$, luego $\dfrac{1/\bar{X}_n - \lambda}{\frac{\lambda}{\sqrt{n}}} \sim N(0,1)$. Así que dado δ con $0 < \delta < 1$, existe $z = z_{\frac{1+\delta}{2}}$ tal que $P\big(-z < (1/\bar{X}_n - \lambda)/\frac{\lambda}{\sqrt{n}} < z\big) = \delta$.

La desigualdad $-z < \dfrac{\frac{1}{\bar{X}_n} - \lambda}{\frac{\lambda}{\sqrt{n}}} < z$ es equivalente a

$$-z < \frac{\frac{1}{\bar{X}_n} - \lambda}{\frac{\lambda}{\sqrt{n}}} \quad \text{y} \quad \frac{\frac{1}{\bar{X}_n} - \lambda}{\frac{\lambda}{\sqrt{n}}} < z.$$

De la primera desigualdad, $\lambda < \dfrac{\frac{1}{\bar{X}_n}}{1 - \frac{z}{\sqrt{n}}}$ y de la segunda $\lambda > \dfrac{\frac{1}{\bar{X}_n}}{1 + \frac{z}{\sqrt{n}}}$.

Por tanto un intervalo de confianza del $100\delta\,\%$ aproximado para λ es

$$\left(\frac{\frac{1}{\bar{X}_n}}{1 + \frac{z}{\sqrt{n}}}, \frac{\frac{1}{\bar{X}_n}}{1 - \frac{z}{\sqrt{n}}} \right).$$

.

9.5.1. Intervalos de confianza para proporciones

En esta sección se trata de obtener intervalos de una proporción p, de una distribución de Bernouli y para la diferencia de dos proporciones $p_1 - p_2$ de dos poblaciones de Bernoulli.

Sea $X_1, X_2, ..., X_n$ una muestra de $f(x;p) = p^x(1-p)^{1-x}I_{\{0,1\}}(x)$. Se desea obtener un intervalo de confianza para p.

Se conoce del capítulo anterior que $\bar{X}$ es un estimador insesgado y eficiente para p, y para n grande $\bar{X}$ es aproximadamente normal con media p y varianza $\frac{p(1-p)}{n}$.

Sea $Z = \frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}}$, entonces Z es aproximadamente normal con media 0 y varianza 1. Para δ dado, $0 < \delta < 1$, existe $z = z_{\frac{1+\delta}{2}}$ tal que $P(-z < Z < z) = \delta$.

Esto significa que $P\left(-z < \frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} < z\right) = \delta$.

De la desigualdad dentro del paréntesis, se concluye que $\frac{|\bar{x}-p|}{\sqrt{\frac{p(1-p)}{n}}} < z$.

Luego, $(\bar{x}-p)^2 < \frac{p(1-p)}{n}z^2$, es decir, $\bar{x}^2 - 2\bar{x}p + p^2 - \frac{pz^2}{n} + \frac{p^2z^2}{n} < 0$, y entonces $p^2(1+\frac{z^2}{n}) - p(2\bar{x}+\frac{z^2}{n}) + \bar{x}^2 < 0$.

Al resolver la ecuación de segundo grado en p, $p^2(1+\frac{z^2}{n}) - p(2\bar{x}+\frac{z^2}{n}) + \bar{x}^2 = 0$ se tiene que,

$$p = \frac{2\bar{x} + \frac{z^2}{n} \pm \sqrt{4\bar{x}\frac{z^2}{n} + \frac{z^4}{n^2} - \frac{4\bar{x}^2z^2}{n}}}{2(1+\frac{z^2}{n})}.$$

De esta expresión se llega a

$$p = \frac{\bar{x} + \frac{z^2}{2n} \pm z\sqrt{\frac{\bar{x}(1-\bar{x})}{n} + \frac{z^4}{n^2}}}{1+\frac{z^2}{n}}.$$

Cuando n es grande $\frac{z^2}{n}$ y $\frac{z^4}{n^2}$ tienden más rápidamente a 0 que cualquier otro valor dividido por n, así que

$$p \cong \bar{x} \pm z\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}.$$

Como $\hat{p} = \bar{x}$, un intervalo del $100\delta\,\%$ de confianza, aproximado para p es

$$\left(\hat{p} - z_{\frac{1+\delta}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + z_{\frac{1+\delta}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right). \tag{9.15}$$

Ejemplo 9.15. Se preguntó a 1000 estudiantes de una universidad sobre la calidad de sus profesores y 630 respondieron que estos eran buenos. Obtenga un intervalo de confianza aproximado para la proporción de bueno.

Solución. Sea p la proporción de *buenos*. Como $n = 1000$ entonces $\hat{p}$=0.63. Si δ=0.90, entonces $z_{\frac{1+\delta}{2}} = z_{0.95}$=1.645. De esta manera,
$\hat{p} - z_{\frac{1+\delta}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$=0.605 y $\hat{p} + z_{\frac{1+\delta}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$=0.655

Luego, un intervalo (aproximado) del 90 % de confianza para p es $(0.605,\ 0.655)$.

Supóngase ahora que, $X_1, X_2, ..., X_n$ es una muestra de $f(x; p_1) = p_1^x(1-p_1)^{1-x}I_{\{0,1\}}(x)$ y $Y_1, Y_2, ..., Y_m$ es una muestra de $f(y; p_2) = p_2^y(1-p_2)^{1-y}I_{\{0,1\}}(y)$. Si las muestras son grandes e independientes, se desea obtener un intervalo para la diferencia $p_1 - p_2$.

Se conoce que las estimaciones de máxima verosimilitud de p_1 y p_2 son respectivamente, $\hat{p}_1 = \bar{x}$ y $\hat{p}_2 = \bar{y}$, además de que

$$E(\bar{X}) = p_1,\ V(\bar{X}) = \frac{p_1(1-p_1)}{n},\ E(\bar{Y}) = p_2,\ \text{y}\ \ V(\bar{Y}) = \frac{p_2(1-p_2)}{m}.$$

De esta menera $E(\bar{X} - \bar{Y}) = p_1 - p_2$ y como las muestras son independientes $V(\bar{X} - \bar{Y}) = V(\bar{X}) + V(\bar{Y}) = \frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}$.

Para n y m grandes la variable

$$Z = \frac{\bar{X} - \bar{Y} - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}}}.$$

tiene distribución normal $(0, 1$ aproximada). Luego dado δ, $0 < \delta < 1$, existe $z = z_{\frac{1+\delta}{2}}$ tal que $P(-z < Z < z) = \delta$.

De manera análoga como se trató la cantidad pivotal Z para p, se tiene que

$$P\left(-z\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}+\frac{\hat{p}_2(1-\hat{p}_2)}{m}} < \bar{X}-\bar{Y}-(p_1-p_2) < z\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}+\frac{\hat{p}_2(1-\hat{p}_2)}{m}}\right)=\delta.$$

Luego un intervalo de confianza del $100\delta\,\%$, aproximado, para p_1-p_2 es

$$\left(\hat{p_1}-\hat{p_2}-z_{\frac{1+\delta}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}+\frac{\hat{p}_2(1-\hat{p}_2)}{m}},\hat{p_1}-\hat{p_2}+z_{\frac{1+\delta}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}+\frac{\hat{p}_2(1-\hat{p}_2)}{m}}\right). \tag{9.16}$$

9.6. Ejercicios

1. Se observó que $\bar{x}$ de una muestra aleatoria de tamaño 25 de una distribución normal $(\mu, 36)$ es 45.7. Encuentre un intervalo de confianza del 90 % para μ.

2. Una muestra de tamaño 28 de una distribución normal (μ, σ^2) produjo $\bar{x}$ =12.35 y s^2 =7.24. Determine un intervalo de confianza del 95 % para μ.

3. Sea $X_1, X_2, ..., X_n$ una muestra de $X \sim N(\mu, 16)$. Encuentre el valor de n tal que $P(\bar{X}-\frac{1}{2}<\mu<\bar{X}+\frac{1}{2})=0.95$.

4. Encuentre un intervalo de confianza del 90 % para θ con base en una muestra aleatoria de tamaño 10, de una población que tiene función de probabilidad $f(x;\theta)=\dfrac{2x}{\theta^2}$, para $0<x<\theta$ y $\theta>0$.

5. Demostrar que la longitud de un intervalo de confianza para σ^2 de una población normal tiende a cero cuando el tamaño de muestra crece.

6. Encuentre una cantidad pivotal con base en una muestra de tamaño n de una población normal (θ, θ), $\theta>0$. Use esta cantidad pivotal para obtener un intervalo de confianza del $100\delta\,\%$ para θ.

7. Sea $X \sim B(225, p)$. Si el valor observado de X es $x = 60$, encuentre un intervalo de confianza aproximado del 95 % para p.

8. Una muestra aleatoria de 500 ships de compuadores se probaron y se encontró que 45 estaban defectuosos. Hallar un intervalo de confianza del 95 % para estimar la fracción verdadera de ships defectuosos.

9. Si 3.2, 4.2, 5.9, 3.6, 3.9, 3.8, 4.3, 4.0, 4.2, 4.9 son los valores de una muestra aleatoria de tamaño 10 de una distribución normal $(4, \sigma^2)$, encuentre un intervalo de cofianza del 95 % para σ^2.

10. Una muestra de tamaño 20 de una distribución normal (μ, σ^2) produjo $\bar{x} = 4.5$ y $s^2 = 3.8$. Determine un intervalo de confianza del 90 % para σ^2.

11. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim W(\alpha, 2)$. Probar que $Q = \frac{2\sum X_i^2}{\alpha^2} \sim \chi^2_{2n}$ y luego use esto para encontrar un intervalo de confianza para α.

12. Si $L(x)$ y $U(x)$ son funciones tales que $L(x) \leq U(x)$ para todo x y satisfacen

$$P(L(X) \leq \theta) = 1 - \alpha_1 \ \text{ y } \ P(U(X) \geq \theta) = 1 - \alpha_2.$$

Probar que $\big(L(X),\ U(X)\big)$ es un intervalo del $100(1 - \alpha_1 - \alpha_2)$ % para θ.

13. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim W(\alpha, \beta)$, con β conocido. Use el método general para obtener un intrvalo de confianza del 100δ % para α.

    a) Basado en el estadístico $T_1 = X_1$.

    b) Basado en el estadístico $T_2 = \sum X_i^\beta$

14. Suponga que X tiene distribución de Laplace $f(x;\theta) = \frac{1}{2\theta} e^{-\frac{|x|}{\theta}}$, con $x \in R$ y $\theta > 0$). Dado δ, $0 < \delta < 1$, obtenga un intervalo de confianza del 100δ % para θ y basado en X.

15. Sea $X_1, X_2, ..., X_n$ una muestra de la distribución de Rayleigh con parámetro θ (ver 4.26). Dado δ, $0 < \delta < 1$ obtenga un intervalo de confianza del 100δ % para θ.

16. Para los dos conjuntos de datos del ejercicio 1.17 obtenga un intervalo de confianza del 95 % para la diferencia de promedios entre la rentabilidad mensual y la semestral.

17. Sean $X_1, X_2, ..., X_m$ y $Y_1, Y_2, ..., Y_n$ dos muestras independientes de $X \sim N(\mu_1, \sigma_1^2)$ y de $Y \sim N(\mu_2, \sigma_2^2)$, respectivamente. Si μ_1 y μ_2 son conocidos, obtenga un intervalo de confianza del 100δ % para $\frac{\sigma_2^2}{\sigma_1^2}$.

18. Sean $X_1, X_2, ..., X_m$ y $Y_1, Y_2, ..., Y_n$ dos muestras independientes de $X \sim E(\frac{1}{\lambda_1})$ y de $Y \sim E(\frac{1}{\lambda_2})$, respectivamente. Probar que $\dfrac{\lambda_2 \bar{X}}{\lambda_1 \bar{Y}} \sim F(2m, 2n)$ y luego obtenga un intervalo de confianza para $\frac{\lambda_2}{\lambda_1}$.

19. Sean $\bar{X}$ y $\bar{Y}$ las medias de dos muestras independientes, cada una de tamaño n, de $X \sim N(\mu_1, \sigma^2)$ y $Y \sim N(\mu_2, \sigma^2)$ respectivamente y donde la varianza común es conocida. Encuentre n de manera que $P(\bar{Y} - \bar{X} - \frac{\sigma}{5} < \mu_2 - \mu_1 < \bar{Y} - \bar{X} + \frac{\sigma}{5}) = 0.90$.

20. Sea $X_1, X_2, ..., X_n$ una muestra de $X \sim B(1, p)$, y suponga que P tiene una distribución a priori uniforme en $(0, 1)$. Obtenga un intervalo bayesiano del $100\delta\,\%$ para p.

Capítulo 10

Pruebas de hipótesis

En la investigación experimental, algunas veces el objetivo es estimar parámetros. Pero más a menudo el propósito final está en incluir algún uso de la estimación. Por ejemplo, se puede desear determinar si un nuevo método de fabricar bombillas eléctricas aumenta la vida de estas, de manera que si el promedio de vida del proceso de fabricación estándar es 1600 horas, es interesante ver si la media del nuevo proceso es mayor que este valor. Para averiguar esto el investigador se puede fundamentar en lo que se conoce como una *prueba de hipótesis*.

10.1. Conceptos básicos

Definición 10.1. Hipótesis estadística. Una hipótesis estadística es una afirmación o conjetura que se hace acerca de una o más variables aleatorias.

Una hipótesis estadística se denota por $\boldsymbol{\mathcal{H}}$ seguida por dos puntos y la afirmación. Si la hipótesis especifica completamente la distribución, se dice simple; de otra manera se dice compuesta.

Ejemplo 10.1. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu, 9)$. La hipótesis de que la media es 12, se escribe $\boldsymbol{\mathcal{H}} : \mu = 12$, y es simple. Mientras que la hipótesis de que la media es menor o igual a 12 se escribe $\boldsymbol{\mathcal{H}} : \mu \leq 12$, y es compuesta.

Definición 10.2. Prueba de una hipótesis estadística. Una prueba Υ, de una hipótesis $\boldsymbol{\mathcal{H}}$ es una regla o procedimiento para decidir si esta se rechaza.

Ejemplo 10.2. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu, 9)$. Considere la hipótesis $\boldsymbol{\mathcal{H}} : \mu \leq 12$. Una prueba posible Υ es: rechazar $\boldsymbol{\mathcal{H}}$ si $\bar{X} > 12 + \frac{3}{\sqrt{n}}$. Otra prueba posible Υ^* es rechazar $\boldsymbol{\mathcal{H}}$ si al lanzar una moneda aparece cara.

La primera prueba se llama no aleatorizada y la última prueba aleatorizada. El nombre de esta última se debe a que la decisión de rechazar la hipótesis depende de un experimento ajeno a la muestra.

Notación
$\Omega = \{(x_1, x_2, ..., x_n) : (x_1, x_2, ..., x_n) \in (X_1, X_2, ..., X_n)\}$ se llama el espacio muestral de observaciones.

Definición 10.3. Prueba no aleatorizada y región crítica. Una prueba Υ de la hipótesis $\boldsymbol{\mathcal{H}}$ definida por: Rechazar $\boldsymbol{\mathcal{H}}$ si y solo si $(x_1, x_2, ..., x_n) \in C_\Upsilon$, donde $C_\Upsilon \subset \Omega$, se llama prueba no aleatorizada y C_Υ se llama Región Crítica de la prueba Υ.

Ejemplo 10.3. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu, 9)$. Entonces Ω=R^n. Si se considera $\boldsymbol{\mathcal{H}} : \mu \leq 12$ y la prueba Υ : rechazar $\boldsymbol{\mathcal{H}}$ si y solo si $\bar{x} > 12 + \frac{3}{\sqrt{n}}$. Entonces Υ es una prueba no aleatorizada y la región crítica es C_Υ=$\{(x_1, x_2, ...x_n) : \bar{x} > 12 + \frac{3}{\sqrt{n}}\}$.

Nota
Una prueba Υ no aleatorizada de $\boldsymbol{\mathcal{H}}$ descompone a Ω en C_Υ y C'_Υ tal que si $(x_1, x_2, ...x_n) \in C_\Upsilon$, la hipótesis $\boldsymbol{\mathcal{H}}$ es rechazada. De esta manera, una prueba no aleatorizada está especificada por su correspondiente región crítica.

En la mayoría de los problemas de pruebas de hipótesis, dos hipótesis son discutidas: la hipótesis a ser probada que se llama hipótesis nula, denotada por $\boldsymbol{\mathcal{H}}_0$ y una segunda que se llama hipótesis alternativa la cual se denota por $\boldsymbol{\mathcal{H}}_1$. La idea es que si la hipótesis nula es falsa, entonces la hipótesis alternativa es verdadera, y viceversa. A menudo se dice que $\boldsymbol{\mathcal{H}}_0$ es probada *contra* o *versus* $\boldsymbol{\mathcal{H}}_1$. Si $\boldsymbol{\mathcal{H}}_0$ no es rechazada, se dice que es aceptada.

Puede ocurrir que $\boldsymbol{\mathcal{H}}_0$ sea verdadera y se rechace o sea falsa y se acepte, así que pueden cometerse dos tipos de error.

Definición 10.4. Tipos de error y tamaño de error Rechazar $\boldsymbol{\mathcal{H}}_0$ cuando es verdadera se llama un error tipo I. Aceptar $\boldsymbol{\mathcal{H}}_0$ cuando es falsa se llama un error tipo II.

La probabilidad de cometer un error tipo I se llama el tamaño del error tipo I y la probabilidad de cometer un error tipo II se llama el tamaño del error tipo II. Algunas veces estas probabilidades se denotan por α y β, respectivamente.

Para cualquier prueba Υ de $\boldsymbol{\mathcal{H}}_0$ hay asociada una función llamada función potencia. Esta da la evaluación del ajuste de una prueba y puede usarse en la comparación de dos pruebas. Una función potencia ideal, es una función que es 0 para aquellos θ correspondientes a $\boldsymbol{\mathcal{H}}_0$ y es 1 para los θ correspondientes a $\boldsymbol{\mathcal{H}}_1$. La idea es que no se quiere rechazar $\boldsymbol{\mathcal{H}}_0$ si es verdadera y se quiere rechazar esta cuando es falsa.

Definición 10.5. Función potencia. Sea Υ una prueba de $\boldsymbol{\mathcal{H}}_0$, la función potencia de la prueba Υ, denotada por $\Pi_\Upsilon(\theta)$ o $\beta(\theta)$, se define como la probabilidad de que $\boldsymbol{\mathcal{H}}_0$ sea rechazada cuando la muestra obtenida viene de una distribución que está parametrizada por θ.

Es decir, $\Pi_\Upsilon(\theta) = \beta(\theta) = P_\theta(Rechazar \boldsymbol{\mathcal{H}}_0)$, donde θ es el valor verdadero del parámetro.

Si Υ es una prueba no aleatorizada, entonces $\Pi_\Upsilon(\theta) = P_\theta[(X_1, X_2, \cdots, X_n) \in C_\Upsilon]$, donde C_Υ es la región crítica asociada con la prueba Υ.

Ejemplo 10.4. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu, 9)$. Considere $\boldsymbol{\mathcal{H}}_0 : \mu \leq 12$ y la prueba: Rechazar $\boldsymbol{\mathcal{H}}_0$ si $\bar{x} > 12 + \frac{3}{\sqrt{n}}$. La función potencia es

$$\begin{aligned}\Pi_\Upsilon(\mu) &= P_\mu\left(\bar{X} > 12 + \frac{3}{\sqrt{n}}\right)\\ &= P\left(\frac{\bar{X}-\mu}{\frac{3}{\sqrt{n}}} > \frac{12 + \frac{3}{\sqrt{n}} - \mu}{\frac{3}{\sqrt{n}}}\right)\\ &= 1 - \Phi\left(\frac{12 + \frac{3}{\sqrt{n}} - \mu}{\frac{3}{\sqrt{n}}}\right).\end{aligned}$$

Para $n = 9$ se tiene que $\Pi_\Upsilon(\mu) = 1 - \Phi(13 - \mu)$. En al figura 10.1 observe que si $\mu > 15$ la prueba Υ es casi segura que rechace $\boldsymbol{\mathcal{H}}_0$, mientras que si $\mu < 11$ la prueba Υ es casi segura que no la rechace, como debe ser. Si $12 < \mu < 13$ (así que $\boldsymbol{\mathcal{H}}_0$ es falsa), la prueba Υ tiene probablidad menor de 0.5 de rechazar $\boldsymbol{\mathcal{H}}_0$.

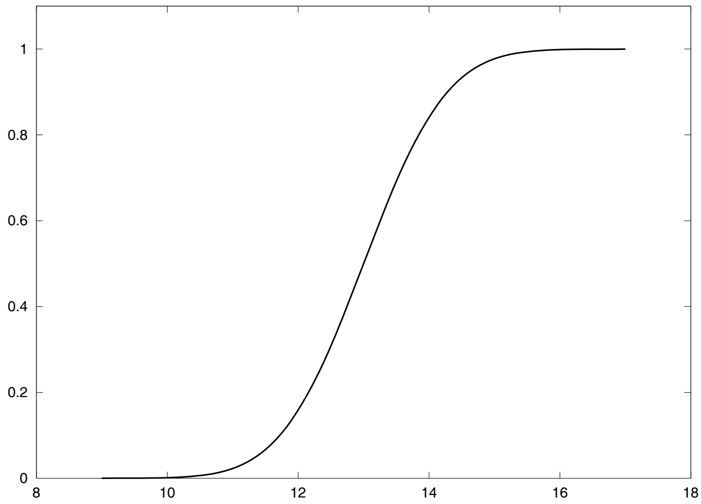


Figura 10.1: Función potencia ejemplo 9.4

Definición 10.6. Tamaño de prueba. Sea Υ una prueba de hipótesis de $\boldsymbol{\mathcal{H}}_0 : \theta \in \Theta_0$, donde $\Theta_0 \subset \Theta$ y Θ es el espacio de parámetros. El tamaño de la prueba Υ de $\boldsymbol{\mathcal{H}}_0$ se define como el

$$\sup_{\theta \in \Theta_0} \Pi_\Upsilon(\theta).$$

De manera similar, si $0 < \alpha < 1$, α es el tamaño de una prueba Υ que tiene función potencia $\beta(\theta)$ si

$$\alpha = \sup_{\theta \in \Theta_0} \beta(\theta). \tag{10.1}$$

El tamaño de la prueba para una prueba no aleatorizada $\boldsymbol{\mathcal{H}}_0$ también se conoce como el *tamaño de la región crítica* o el *nivel de significancia*.

Para algunos autores el nivel de significancia se define por $\alpha \leq \sup\limits_{\theta \in \Theta_0} \Pi_\Upsilon(\theta)$.

Ejemplo 10.5. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu, 9)$, considere $\boldsymbol{\mathcal{H}}_0 : \mu \leqslant 12$ y la prueba Υ : Rechazar $\boldsymbol{\mathcal{H}}_0$ si $\bar{x} > 12 + \dfrac{3}{\sqrt{n}}$. Entonces $\Theta_0 = \{\mu : \mu \leqslant 12\}$ y el tamaño de la prueba Υ para $n = 9$ es

$$\alpha = \sup_{\mu \in \Theta_0} \Pi_\Upsilon(\mu) = \sup_{\mu \leqslant\ 12} [1 - \Phi(13 - \mu)] = 1 - \Phi(1) \equiv 0.159.$$

En general, para probar hipótesis se emplean dos tipos de pruebas: hipótesis simple *vs* hipótesis simple, e hipótesis compuesta *vs* hipótesis compuesta. Esta última es la más común.

10.2. Prueba de hipótesis simple vs alternativa simple

Supóngase que $X_1, X_2, ..., X_n$ es una muestra aleatoria de $f(x;\theta)$ donde $\Theta = \{\theta_0, \theta_1\}$, se quiere probar $\boldsymbol{\mathcal{H}}_0 : \theta = \theta_0$ *vs* $\boldsymbol{\mathcal{H}}_1 : \theta = \theta_1$. Primero se considera una prueba llamada de la razón de *verosimilitud simple* y luego un criterio para obtener una buena prueba. Enseguida se presenta otra prueba en la que Θ es una variable aleatoria y que se conoce como prueba de Bayes.

10.2.1. Prueba de la razón de verosimilitud simple

Definición 10.7. Razón de verosimilitud simple. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$, donde $\theta = \theta_0$ o $\theta = \theta_1$, se define la *razón de verosimilitud simple* por,

$$\lambda = \lambda(x_1, ...x_n) = \frac{L(\theta_0; x_1, ...x_n)}{L(\theta_1; x_1, ...x_n)} = \frac{L_0}{L_1} = \frac{\Pi f(x_i;\theta_0)}{\Pi f(x_i;\theta_1)}. \tag{10.2}$$

Definición 10.8. Prueba de la razón de verosimilitud simple (p.r.v.s). Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$, donde $\theta = \theta_0$ o $\theta = \theta_1$, una prueba Υ de $\boldsymbol{\mathcal{H}}_0 : \theta = \theta_0$ *vs* $\boldsymbol{\mathcal{H}}_1 : \theta = \theta_1$ se dice una p.r.v.s si Υ se define por,

$$\textit{Rechazar } \boldsymbol{\mathcal{H}}_0 \text{ si } \lambda < k, \ \textit{aceptar } \boldsymbol{\mathcal{H}}_0 \text{ si } \lambda > k$$

y

$$\textit{aceptar } \boldsymbol{\mathcal{H}}_0, \textit{rechazar } \boldsymbol{\mathcal{H}}_0 \text{ o } \textit{aleatorizar} \text{ si } \lambda = k,$$

donde k es una constante no negativa. Para cada k diferente se tienen diferentes pruebas y para un k dado la prueba nos dice que se rechace $\boldsymbol{\mathcal{H}}_0$ si la razón de verosimilitud es pequeña; es decir, rechazar $\boldsymbol{\mathcal{H}}_0$ si es más posible que la muestra venga de $f(x;\theta_1)$ que de $f(x;\theta_0)$. Una buena prueba Υ de $\boldsymbol{\mathcal{H}}_0$ es una prueba para la cual,

$$\Pi_\Upsilon(\theta_0) = P(\textbf{Rechazar}\boldsymbol{\mathcal{H}}_0 | \boldsymbol{\mathcal{H}}_0 \text{ es verdadera}), \text{es pequeña, idealmente } 0$$

y

$$\Pi_{\Upsilon}(\theta_1) = P(\textbf{Rechazar}\boldsymbol{\mathcal{H}}_0|\boldsymbol{\mathcal{H}}_0 \text{ es falsa}), \text{ es grande, idealmente } 1.$$

Otra buena prueba de $\boldsymbol{\mathcal{H}}_0 : \theta = \theta_0$ vs $\boldsymbol{\mathcal{H}}_1 : \theta = \theta_1$ puede ser: hacer la suma de los dos errores lo más pequeña posible. Sin embargo, se usará el criterio de *prueba más potente*.

Definición 10.9. Prueba más potente Una prueba Υ^* de $\boldsymbol{\mathcal{H}}_0 : \theta = \theta_0$ *vs* $\boldsymbol{\mathcal{H}}_1 : \theta = \theta_1$ se dice que es una prueba más potente de tamaño α, con $0 < \alpha < 1$, si y solo si,

1. $\Pi_{\Upsilon^*}(\theta_0) = \alpha$
2. $\Pi_{\Upsilon^*}(\theta_1) \geq \Pi_{\Upsilon}(\theta_1)$ para cualquier otra prueba Υ para la cual $\Pi_{\Upsilon}(\theta_0) \leq \alpha$.

Luego, una prueba Υ^* es una prueba más potente de tamaño α si entre todas las pruebas de tamaño α o menor tiene la mayor potencia.

Teorema 10.1. Lema de Neyman-Pearson. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$, donde $\theta = \theta_0$ o $\theta = \theta_1$, y sea α dado, $0 < \alpha < 1$. Sea $k^* > 0$ y $C^* \subset \Omega$ la cual satisface:

1. $P_{\theta_0}[(X_1, X_2, ..., X_n) \in C^*] = \alpha$.
2. $\lambda \leqslant k^*$ si $(x_1, ..., x_n) \in C^*$ y
   $\lambda \geq k^*$ si $(x_1, ..., x_n) \in C^{*'}$,

donde λ es la razón de verosimilitud simple. Entonces, la prueba Υ^* correspondiente a la región crítica C^* es una prueba más potente de tamaño α de $\boldsymbol{\mathcal{H}}_o : \theta = \theta_0$ vs $\boldsymbol{\mathcal{H}}_1 : \theta = \theta_1$.

Demostración. Para realizar la prueba, sea $X = (X_1, ..., X_n)$, y considere $\int_D f(x_1, ..., x_n)dx_1...dx_n = \int_D L_j$ para $j = 0, 1$ y para cualquier $D \subset \Omega$.

Supóngase que Υ^* es una prueba con k^* y C^* que satisfacen 1) y 2). Si no existe otra prueba de tamaño α o menor, entonces Υ^* es una prueba más potente. Si existe otra prueba, sea esta Υ con región crítica C. Luego $P_{\theta_0}((X_1, X_2, ..., X_n) \in C) \leq \alpha$. De manera que se debe mostrar que $\Pi_{\Upsilon^*}(\theta_1) \geqslant \Pi_{\Upsilon}(\theta_1)$.

Como $C^* = (C^* \cap C) \cup (C^* \cap C')$ y $C = (C^* \cap C) \cup (C^{*'} \cap C)$, entonces

$$\Pi_{\Upsilon^*}(\theta) = P_\theta(X \in C^* \cap C) + P_\theta(X \in C^* \cap C')$$

y

$$\Pi_{\Upsilon}(\theta) = P_\theta(X \in C^* \cap C) + P_\theta(X \in C^{*'} \cap C).$$

Luego,

$$\Pi_{\Upsilon^*}(\theta) - \Pi_{\Upsilon}(\theta) = P_\theta(X \in C^* \cap C') - P_\theta(X \in C^{*'} \cap C) = \int_{C^* \cap C'} L(\theta) - \int_{C^{*'} \cap C} L(\theta).$$

Observe de la definición 10.7 y del teorema 10.1 que

$$L_1 \geqslant \frac{L_0}{k^*} \text{ si } X \in C^* \text{ y } L_1 \leqslant \frac{L_0}{k^*} \text{ si } X \in C^{*'},$$

por tanto

$$\begin{aligned}
\Pi_{\Upsilon^*}(\theta_1) - \Pi_{\Upsilon}(\theta_1) &\geqslant \frac{1}{k^*}\int_{C^* \cap C'} L_0 - \frac{1}{k^*}\int_{C^{*'} \cap C} L_0 \\
&= \frac{1}{k^*}\left[\int_{C^* \cap C'} L_0 + \int_{C^* \cap C} L_0 - \left(\int_{C^{*'} \cap C} L_0 + \int_{C^* \cap C} L_0\right)\right] \\
&= \frac{1}{k^*}\left[\int_{C^*} L_0 - \int_{C} L_0\right] = \frac{1}{k^*}\big(\alpha - \textbf{tamaño de}\Upsilon\big) \geqslant 0.
\end{aligned}$$

De esta manera, se concluye que $\Pi_{\Upsilon^*}(\theta_1) \geqslant \Pi_{\Upsilon}(\theta_1)$.

Ejemplo 10.6. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\lambda)=\lambda e^{-\lambda x}$, para $x > 0$, donde $\lambda = \lambda_0$ o $\lambda = \lambda_1$, y son conocidos. Supóngase que se quiere probar $\boldsymbol{\mathcal{H}}_0 : \lambda = \lambda_0$ vs $\boldsymbol{\mathcal{H}}_1 : \lambda = \lambda_1$ cuando $\lambda_0 < \lambda_1$. Entonces

$$\lambda = \frac{\lambda_0^n e^{-\lambda_0 \sum x_i}}{\lambda_1^n e^{-\lambda_1 \sum x_i}} = \left(\frac{\lambda_0}{\lambda_1}\right)^n e^{-(\lambda_0 - \lambda_1)\sum x_i}.$$

De acuerdo al lema, se tiene que la prueba más potente de tamaño α es rechazar $\boldsymbol{\mathcal{H}}_0$ si,

$$\left(\frac{\lambda_0}{\lambda_1}\right)^n e^{-(\lambda_0 - \lambda_1)\sum x_i} \leq k^*,$$

o equivalentemente, rechazar $\boldsymbol{\mathcal{H}}_0$ si

$$\Sigma x_i \leqslant \frac{1}{\lambda_1 - \lambda_0} \ln\left[\left(\frac{\lambda_1}{\lambda_0}\right)^n k^*\right] = k'.$$

La condición 1 del lema es $P_{\lambda_0}(\Sigma X_i \leqslant k') = \alpha$, y como ΣX_i tiene distribución gama con parámetros n y λ, entonces k' se obtiene de

$$\alpha = P_{\lambda_0}\left(\Sigma X_i \leqslant k'\right) = \int_0^{k'} \frac{1}{\Gamma(n)} \lambda_0^n x^{n-1} e^{-\lambda_0 x} dx = F_{\lambda_0}(k').$$

Pero como n es un entero positivo, por el teorema 4.21

$$F_{\lambda_0}(k') = 1 - \sum_{j=0}^{n-1} \frac{e^{-\lambda_0 k'}(\lambda_0 k')^j}{j!}.$$

Luego k' se obtiene de

$$\sum_{j=0}^{n-1} \frac{e^{-\lambda_0 k'}(\lambda_0 k')^j}{j!} - (1-\alpha) = 0.$$

Por ejemplo, si $\lambda_0 = 1$ y $\alpha = 0.05$, para $n = 2$, k' es aproximadamente 0.025, es decir que se rechaza $\boldsymbol{\mathcal{H}}_0$ si $\Sigma_1^2 x_i \leq 0.025$. Para $n = 10$, k' es aproximadamente 5.43, es decir, se rechaza $\boldsymbol{\mathcal{H}}_0$ si $\Sigma_1^{10} x_i \leq 5.43$.

Ejemplo 10.7. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(0, \sigma^2)$. Se quiere probar $\boldsymbol{\mathcal{H}}_0 : \sigma^2 = \sigma_0^2$ *vs* $\boldsymbol{\mathcal{H}}_1 : \sigma^2 = \sigma_1^2$ cuando $\sigma_1^2 \geqslant \sigma_0^2$. Entonces

$$\lambda = \frac{\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{x_i^2}{2\sigma_0^2}}\right)}{\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{x_i^2}{2\sigma_1^2}}\right)} = \left(\frac{\sigma_1}{\sigma_0}\right)^n e^{\frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\Sigma x_i^2}.$$

De acuerdo al lema la prueba más potente de tamaño α es rechazar $\boldsymbol{\mathcal{H}}_0$ si

$$\lambda = \left(\frac{\sigma_1}{\sigma_0}\right)^n \exp\left(\frac{1}{2}\left[\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right]\Sigma x_i^2\right) \leqslant k.$$

Es decir, se rechaza $\boldsymbol{\mathcal{H}}_0$ si

$$\exp\left(\frac{1}{2}\left[\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right]\Sigma x_i^2\right) \leqslant k'$$

o equivalentemente: rechazar $\boldsymbol{\mathcal{H}}_0$ si $\Sigma x_i^2 \geqslant k^*$.

De manera que la prueba más potente de tamaño α tiene región crítica
$C^* = \{(x_1, ..., x_n) : \Sigma x_i^2 \geqslant k^*\}$. Esto implica que $P_{\sigma_0^2}(\Sigma X_i^2 \geqq k^*) = \alpha$.

Ahora bien, bajo $\boldsymbol{\mathcal{H}}_0$ la variable $Q = \frac{\Sigma X_i^2}{\sigma_0^2} \sim \chi_n^2$, entonces $P(Q \geq \frac{k^*}{\sigma_0^2}) = \alpha$. Lo que significa que, $\frac{k^*}{\sigma_0^2} = \chi^2_{n(1-\alpha)}$. De esta manera se rechaza $\boldsymbol{\mathcal{H}}_0$ si $\Sigma x_i^2 \geqslant \sigma_0^2 \chi^2_{n(1-\alpha)}$.

¿Cuál será la prueba si $\sigma_1^2 \leq \sigma_0^2$?

Se observa en estos ejemplos que cada prueba se fundamenta en un estadístico. $T = \Sigma X_i$ en el ejemplo 10.6 y $T = \Sigma X_i^2$ en el ejemplo 10.7. Estos estadísticos se conocen como *estadísticos de prueba*.

10.2.2. Prueba de Bayes

Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$, donde $\theta \in \Theta$ y $\Theta = \{\theta_0, \theta_1\}$ es una variable aleatoria discreta. Suponga que se conoce la distribución de Θ; sea esta $g = P(\Theta = \theta_1) = 1 - P(\Theta = \theta_0)$. Al probar $\boldsymbol{\mathcal{H}}_0 : \theta = \theta_0$ *vs* $\boldsymbol{\mathcal{H}}_1 : \theta = \theta_1$ se puede tomar o no la decisión correcta, y es por eso que se desea una prueba con el menor riesgo esperado posible. Esta prueba es la prueba de Bayes.

Definición 10.10. Función de pérdida. En la prueba de $\boldsymbol{\mathcal{H}}_0 : \theta = \theta_0$ *vs* $\boldsymbol{\mathcal{H}}_1 : \theta = \theta_1$, la función de pérdida $l(d_i; \theta_j)$ es la pérdida incurrida cuando la decisión d_i se toma (decisión que $\boldsymbol{\mathcal{H}}_i$ es correcta) y θ_j es el valor verdadero del parámetro para $i = 0, 1$ y $j = 0, 1$. Por convención,

$$l(d_i, \theta_j) = 0, \quad \text{para } i = j \ \text{ y } l(d_i, \theta_j) > 0, \quad \text{para } i \neq j \tag{10.3}$$

Definición 10.11. Pérdida esperada. En la prueba de $\boldsymbol{\mathcal{H}}_0 : \theta = \theta_0$ *vs* $\boldsymbol{\mathcal{H}}_1 : \theta = \theta_1$ si Υ es la prueba con región crítica C_Υ la pérdida esperada o costo esperado o la función de riesgo de Υ se define por

$$R_\Upsilon(\theta) = E[l(d;\theta)] \tag{10.4}$$

Observe que,

$$\begin{aligned} R_\Upsilon(\theta) &= \int \cdots \int_{C_\Upsilon} l(d_1;\theta) \Pi f(x_i;\theta) dx_i + \int \cdots \int_{C'_\Upsilon} l(d_0;\theta) \Pi f(x_i;\theta) dx_i \\ &= l(d_1;\theta)\Pi_\Upsilon(\theta) + l(d_0;\theta)[1 - \Pi_\Upsilon(\theta)] \\ &= l(d_0;\theta) + [l(d_1;\theta) - l(d_0;\theta)]\Pi_\Upsilon(\theta). \end{aligned}$$

De esta forma

$$R_\Upsilon(\theta_0) = l(d_1;\theta_0)\Pi_\Upsilon(\theta_0) \ \text{ y } \ R_\Upsilon(\theta_1) = l(d_0;\theta_1)[1 - \Pi_\Upsilon(\theta_1)]. \tag{10.5}$$

Definición 10.12. Prueba de Bayes. Una prueba Υ' de $\boldsymbol{\mathcal{H}}_0 : \theta = \theta_0$ *vs* $\boldsymbol{\mathcal{H}}_1 : \theta = \theta_1$ se dice que es una prueba de Bayes respecto a la distribución a priori dada por $g = P(\Theta = \theta_1)$ si y solo si

$$(1-g)R_{\Upsilon'}(\theta_0) + gR_{\Upsilon'}(\theta_1) \leq (1-g)R_{\Upsilon}(\theta_0) + gR_{\Upsilon}(\theta_1) \tag{10.6}$$

para cualquier otra prueba Υ.

Teorema 10.2. La prueba de Bayes, Υ' de $\boldsymbol{\mathcal{H}}_0 : \theta = \theta_0$ *vs* $\boldsymbol{\mathcal{H}}_1 : \theta = \theta_1$ con respecto a la distribución a priori dada por $g = P(\Theta = \theta_1)$ tiene región crítica definida por

$$C' = \left\{ (x_1, x_2, \cdots, x_n) : \lambda = \frac{L_0}{L_1} < \frac{gl(d_0;\theta_1)}{(1-g)l(d_1;\theta_0)} \right\}. \tag{10.7}$$

Demostración. Sea Υ' la prueba de Bayes de $\boldsymbol{\mathcal{H}}_0 : \theta = \theta_0$ *vs* $\boldsymbol{\mathcal{H}}_1 : \theta = \theta_1$ respecto a la distribución a priori $g = P(\Theta = \theta_1)$, y sea C' su región crítica entonces,

$$\begin{aligned}
(1-g)R_{\Upsilon'}(\theta_0) + gR_{\Upsilon'}(\theta_1) &= (1-g)l(d_1;\theta_0)\Pi_{\Upsilon'}(\theta_0) + gl(d_0;\theta_1)[1 - \Pi_{\Upsilon'}(\theta_1)] \\
&= (1-g)l(d_1;\theta_0)\Pi_{\Upsilon'}(\theta_0) + gl(d_0;\theta_1) - gl(d_0;\theta_1)\Pi_{\Upsilon'}(\theta_1) \\
&= gl(d_0;\theta_1) + (1-g)l(d_1;\theta_0)\int_{C'} L_0 - gl(d_0;\theta_1)\int_{C'} L_1 \\
&= gl(d_0;\theta_1) + \int_{C'} \big[(1-g)l(d_1;\theta_0)L_0 - gl(d_0;\theta_1)L_1\big].
\end{aligned}$$

Esta expresión es mínima si C' está definida para todo $(x_1, x_2, \cdots, x_n)$ para lo cual $(1-g)l(d_1;\theta_0)L_0 - gl(d_0;\theta_1)L_1 < 0$, es decir, si $(1-g)l(d_1;\theta_0)L_0 < gl(d_0;\theta_1)L_1$.

Lo que significa que,

$$\frac{L_0}{L_1} < \frac{gl(d_0;\theta_1)}{(1-g)l(d_1;\theta_0)} \text{ y } C' = \left\{ (x_1, x_2, \cdots, x_n) : \lambda = \frac{L_0}{L_1} < \frac{gl(d_0;\theta_1)}{(1-g)l(d_1;\theta_0)} \right\}.$$

Ejemplo 10.8. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu, \sigma^2)$ donde σ^2 es conocido. Sea $\Theta = \{\mu_0, \mu_1\}$ una variable aleatoria. Se desea probar $\boldsymbol{\mathcal{H}}_0 : \mu = \mu_0$ vs $\boldsymbol{\mathcal{H}}_1 : \mu = \mu_1$, con $\mu_0 > \mu_1$.

Solución. Sea $g = P(\Theta = \mu_1)$ y $1 - g = P(\Theta = \mu_0)$. Se encuentra ahora $\lambda = \frac{L_0}{L_1}$.

De $L_j = \Pi f(x_i;\mu_j)$, para $j = 0, 1$, se tiene que

$$L_0 = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu_0)^2} \quad \text{y} \quad L_1 = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu_1)^2}.$$

Luego,

$$\lambda = e^{-\frac{1}{2\sigma^2}[\sum(x_i-\mu_0)^2-\sum(x_i-\mu_1)^2]} = e^{\frac{n}{\sigma^2}\bar{x}(\mu_0-\mu_1)-\frac{n}{2\sigma^2}(\mu_0^2-\mu_1^2)}.$$

De manera que la región crítica de la prueba está dada por,

$$\begin{aligned} C &= \left\{(x_1, x_2, \cdots, x_n) : e^{\frac{n}{\sigma^2}\bar{x}(\mu_0-\mu_1)-\frac{n}{2\sigma^2}(\mu_0^2-\mu_1^2)} < \frac{gl(d_0;\mu_1)}{(1-g)l(d_1;\mu_0)}\right\} \\ &= \left\{(x_1, x_2, \cdots, x_n) : \frac{n}{\sigma^2}\bar{x}(\mu_0-\mu_1) - \frac{n}{2\sigma^2}(\mu_0^2-\mu_1^2) < \ln \frac{gl(d_0;\mu_1)}{(1-g)l(d_1;\mu_0)}\right\} \\ &= \left\{(x_1, x_2, \cdots, x_n) : \bar{x} < \frac{\mu_0+\mu_1}{2} + \frac{\sigma^2}{n(\mu_0-\mu_1)} \ln \frac{gl(d_0;\mu_1)}{(1-g)l(d_1;\mu_0)}\right\} \end{aligned}$$

para $\mu_0 > \mu_1$.

10.3. Prueba de hipótesis compuesta

Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$, donde $\theta \in \Theta$. Se quiere probar $\boldsymbol{\mathcal{H}}_0 : \theta \in \Theta_0$ vs $\boldsymbol{\mathcal{H}}_1 : \theta \in \Theta_1$, donde $\Theta_0, \Theta_1 \subset \Theta$ y forman una partición de Θ, es decir, $\Theta_0 \cap \Theta_1 = \Phi$ y $\Theta_0 \cup \Theta_1 = \Theta$. Una prueba de este contraste de hipótesis la da la prueba de la razón de verosimilitud generalizada, y otra es la prueba de la Razón de verosimilitud monótona. Estas pruebas se consideran en esta sección.

10.3.1. Prueba de la razón de verosimilitud generalizada

Definición 10.13. Razón de verosimilitud generalizada. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$, donde $\theta \in \Theta$ y sea $L(\theta; x_1, ..., x_n)$ la función de verosimilitud. Se define la razón de verosimilitud generalizada, λ o λ_n por,

$$\lambda = \lambda_n = \lambda(x_1, ..., x_n) = \frac{\sup\limits_{\theta\in\Theta_0} L(\theta; x_1, ..., x_n)}{\sup\limits_{\theta\in\Theta} L(\theta; x_1, ..., x_n)}. \tag{10.8}$$

Se observa que, $0 \leqslant \lambda \leqslant 1$ y que el denominador de λ es la función de verosimilitud evaluada en el estimador de máxima verosimilitud.

Una prueba de $\boldsymbol{\mathcal{H}}_0 : \theta \in \Theta_0$ vs $\boldsymbol{\mathcal{H}}_1 : \theta \in \Theta_1 = \Theta - \Theta_0$ la da **el principio de la razón de verosimilitud generalizada** el cual establece: *rechace* $\boldsymbol{\mathcal{H}}_0$ *si y solo si* $\lambda \leqslant \lambda_0$, donde λ_0 es una costante fija y que satisface $0 \leqslant \lambda_0 \leqslant 1$.

Definición 10.14. Prueba uniformemente más potente. Una prueba Υ^* de $\boldsymbol{\mathcal{H}}_0 : \theta \in \Theta_0$ vs $\boldsymbol{\mathcal{H}}_1 : \theta \in \Theta - \Theta_0$ se dice que es una prueba uniformemente más potente (PUMP) de tamaño α si y solo si

1. $\sup\limits_{\theta \in \Theta_0} \Pi_{\Upsilon^*}(\theta) = \alpha$

2. $\Pi_{\Upsilon^*}(\theta) \geqslant \Pi_{\Upsilon}(\theta)$, para todo $\theta \in \Theta - \Theta_0$ y para cualquier prueba Υ con tamaño α o menor.

Los dos teoremas que se enuncian a continuación permiten obtener PUMP.

Teorema 10.3. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta)$, donde $\theta \in \Theta$ y Θ es algún intervalo. Supóngase que $f(x;\theta)$ pertenece a la familia exponencial y sea $t(x_1, ...x_n) = \sum\limits_{i=1}^{n} d(x_i)$:

1. Si $c(\theta)$ es una función monótona creciente en Θ y si existe un k^* tal que $P_{\theta_0}[t(X_1, ...X_n) \geq k^*] = \alpha$ entonces la prueba Υ^* con región crítica dada por $C^* = \{(x_1, ...x_n) : t(x_1, ...x_n) \geq k^*\}$ es una PUMP de tamaño α de $\boldsymbol{\mathcal{H}}_0 : \theta \leqslant \theta_0$ vs $\boldsymbol{\mathcal{H}}_1 : \theta > \theta_0$ o de $\boldsymbol{\mathcal{H}}_0 : \theta = \theta_0$ vs $\boldsymbol{\mathcal{H}}_1 : \theta > \theta_0$.

2. Si $c(\theta)$ es una función monótona decreciente en Θ y si existe un k^* tal que $P_{\theta_0}[t(X_1, ...X_n) \leq k^*] = \alpha$ entonces la prueba Υ^* con región crítica dada por $C^* = \{(x_1, ...x_n) : t(x_1, ...x_n) \leq k^*\}$ es una PUMP de tamaño α de $\boldsymbol{\mathcal{H}}_0 : \theta \leqslant \theta_0$ vs $\boldsymbol{\mathcal{H}}_1 : \theta > \theta_0$ o de $\boldsymbol{\mathcal{H}}_0 : \theta = \theta_0$ vs $\boldsymbol{\mathcal{H}}_1 : \theta > \theta_0$.

Ejemplo 10.9. Sea $X_1, ...X_n$ una muestra aleatoria de $X \sim B(1, p)$, $0 < p < 1$. Obtenga una PUMP de tamaño α de $\boldsymbol{\mathcal{H}}_0 : p \leqslant p_0$ vs $\boldsymbol{\mathcal{H}}_1 : p > p_0$.

Solución. Del ejemplo 7.13 $f(x,p) = p^x(1-p)^{1-x}I_{\{0,1\}}$ pertenece a la familia exponencial con $c(p) = \ln \frac{p}{1-p}$ y $d(x) = x$. Además, $c(p)$ es una función monotona creciente de p, luego una PUMP de tamaño α de $\boldsymbol{\mathcal{H}}_0 : p \leqslant p_0$ vs $\boldsymbol{\mathcal{H}}_1 : p > p_0$ o de $\boldsymbol{\mathcal{H}}_0 : p = p_0$ vs $\boldsymbol{\mathcal{H}}_1 : p > p_0$ es: rechazar $\boldsymbol{\mathcal{H}}_0$ si $\Sigma d(x_i) \geq k^*$. Es decir, rechazar $\boldsymbol{\mathcal{H}}_0$ si $\Sigma x_i \geq k^*$. Aquí k^* es un entero positivo y se obtiene de $\alpha = P_{p_0}(\Sigma X_i \geq k^*) = 1 - P_{p_0}(\Sigma X_i < k^*)$.

Bajo $\boldsymbol{\mathcal{H}}_0$, $\Sigma X_i \sim B(n, p_0)$, luego $\alpha = 1 - P_{p_0}(\Sigma X_i < k^*) = 1 - P(\Sigma X_i \leq k^* - 1)$. Esto significa que $k^* - 1 = \xi_{1-\alpha} b(n, p_0)$ o $k^* = 1 + \xi_{1-\alpha} b(n, p_0)$, donde $\xi_{1-\alpha} b(n, p_0)$ es el percentil del $100 p_0\,\%$ de la distribución binomial.

Si $\alpha = 0.05$, $n = 15$ y $p_0 = 0.3$, se rechaza $\mathcal{H}_0$ si $\Sigma x_i \geq 8$. Sin embargo, para una muestra de esta distribución no siempre es posible obtener k^*, de manera que una prueba no es posible. Observe que si $\alpha = 0.1$, $n = 15$ y $p_0 = 0.2$ no existe k^*, entero positivo, tal que $0.9 = P_{0.2}(\Sigma X_i \leq k^* - 1)$.

A pesar de esto, no se debe alarmar, pues el teorema 10.5 da un criterio para obtener pruebas de hipótesis en este caso.

Ejemplo 10.10. Sea $X_1, ...X_n$ una muestra aleatoria de $X \sim E(\theta)$, obtenga una PUMP de tamaño α de $\mathcal{H}_0 : \theta \leqslant \theta_0 \ vs \ \mathcal{H}_1 : \theta \ > \theta_0$.

Solución. Como $f(x;\theta) = \theta e^{-\theta x}$, $x > 0$ y $\theta > 0$, entonces

$$\begin{aligned} f(x;\theta) &= \theta e^{-\theta x} I_{(0,\infty)}(x) \\ &= \theta I_{(0,\infty)} e^{-\theta x} = a(\theta) b(x) e^{c(\theta)d(x)}, \end{aligned}$$

donde $c(\theta) = -\theta$ y $d(x) = x$.

Además, como $c(\theta)$ es monótona decreciente y $\Sigma d(x_i) = \sum x_i$, entonces por el teorema anterior una PUMP de tamaño α de $\mathcal{H}_0 : \theta \leqslant \theta_0 \ vs \ \mathcal{H}_1 : \theta \ > \theta_0$ es rechazar H_0 si $\Sigma x_i < k^*$ donde k^* se obtiene de,

$$\alpha = P_{\theta_0} \left[\Sigma X_i < k^* \right].$$

Pero bajo $\mathcal{H}_0$, $\Sigma X_i \sim G(n, \theta_0)$, luego k^* se obtendrá de

$$\alpha = \int_0^{k^*} \frac{1}{\Gamma(n)} \theta_0 u^{n-1} e^{-\theta_0 u} du.$$

Como en el ejemplo 10.6 el valor k^* se calcula al resolver $\alpha = 1 - \sum_{i=0}^{n-1} \frac{e^{-\theta_0 k^*} (\theta_0 k^*)^i}{i!}$.

Otra prueba de $\mathcal{H}_0 : \theta \in \Theta_0$ vs $\mathcal{H}_1 : \theta \in \Theta_1$ la da la *prueba de la razón de verosimilitud monótona.*

10.3.2. Prueba de la razón de verosimilitud monótona

Definición 10.15. Razón de verosimilitud monótona Una familia de densidades $\{f(x;\theta) : \theta \in \Theta$, y Θ es un intervalo$\}$ se dice que tiene una razón de verosimilitud monótona si existe un estadístico $T = t(X_1, ..., X_n)$, tal que la razón

$L(\theta_1; x_1, ..., x_n)/L(\theta_2; x_1, ..., x_n)$ es una función no creciente o una función no decreciente de $t(x_1, ..., x_n)$ para todo $\theta_1 < \theta_2$.

Ejemplo 10.11. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(0, \sigma^2)$. Entonces, $f(x, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$ tiene una razón de verosimilitud monótona.

En efecto,

$$\frac{L(\sigma_1^2; x_1, ..., x_n)}{L(\sigma_2^2; x_1, ..., x_n)} = \left(\frac{\sigma_2}{\sigma_1}\right)^n e^{-\frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right)\sum x_i^2}$$

es una función no creciente de Σx_i^2, para todo $\sigma_1^2 < \sigma_2^2$.

Teorema 10.4. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x; \theta)$, donde Θ es algún intervalo. Sea $\{f(x; \theta) : \theta \in \Theta\}$ una familia que tiene una razón de verosimilitud monótona en $t(x_1, ..., x_n)$:

1. Si la razón de verosimilitud monótona es no decreciente en $t(x_1, ..., x_n)$ y si k^* es tal que $P_{\theta_0}[t(X_1, ...X_n) < k^*] = \alpha$ entonces la prueba Υ^* con región crítica $C^* = \{(x_1, ...x_n) : t(x_1, ...x_n) < k^*\}$ es una PUMP de tamaño α de $\boldsymbol{\mathcal{H}}_0 : \theta \leqslant \theta_0 \ vs \ \boldsymbol{\mathcal{H}}_1 : \theta > \theta_0$.

2. Si la razón de verosimilitud monótona es no creciente en $t(x_1, ..., x_n)$ y si k^* es tal que $P_{\theta_0}[t(X_1, ...X_n) > k^*] = \alpha$ entonces la prueba Υ^* con región crítica $C^* = \{(x_1, ...x_n) : t(x_1, ...x_n) > k^*\}$ es una PUMP de tamaño α de $\boldsymbol{\mathcal{H}}_0 : \theta \leqslant \theta_0 \ vs \ \boldsymbol{\mathcal{H}}_1 : \theta > \theta_0$.

Ejemplo 10.12. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(0, \sigma^2)$. Se quiere probar $\boldsymbol{\mathcal{H}}_0 : \sigma^2 \leqslant \sigma_0^2 \ vs \ \boldsymbol{\mathcal{H}}_1 : \sigma^2 > \sigma_0^2$.

Solución. Del ejemplo anterior se conoce que $f(x, \sigma^2)$ tiene una razón de verosimilitud monótona no creciente de Σx_i^2. Así que una PUMP de tamaño α es rechazar $\boldsymbol{\mathcal{H}}_0$ si $\Sigma x_i^2 > k^*$, donde k^* se obtiene de $P\left(\Sigma X_i^2 > k^*\right) = \alpha$.

Como bajo $\boldsymbol{\mathcal{H}}_0$ el estadístico $Q = \frac{\sum X_i^2}{\sigma_0^2} \sim \chi_n^2$, entonces $\alpha = P_{\sigma_0^2}\left(Q > \dfrac{k^*}{\sigma_0^2}\right)$ implica que $\frac{k^*}{\sigma_0^2} = \chi^2_{n(1-\alpha)}$. Luego una PUMP de tamaño α es rechazar $\boldsymbol{\mathcal{H}}_0$ si $\Sigma x_i^2 > \sigma_0^2 \chi^2_{n(1-\alpha)}$.

Este resultado coincide con el del ejemplo 10.7, como debe ser.

Para finalizar esta sección se define lo que es una prueba insesgada y en la siguiente se muestra una de estas pruebas.

Definición 10.16. Prueba insesgada. Una prueba Υ de $\boldsymbol{\mathcal{H}}_0 : \theta \in \Theta_0$ vs $\boldsymbol{\mathcal{H}}_0 : \theta \in \Theta_1$ con función potencia $\beta(\theta)$ es una prueba insesgada si $\beta(\theta') \geq \beta(\theta'')$ para todo $\theta' \in \Theta_1$ y $\theta'' \in \Theta_0$.

10.4. Pruebas de la distribución normal

Por la importancia que juega la distribución normal en teoría estadística se consideran en esta sección pruebas acerca de los parámetros μ y σ^2 de esta distribución.

10.4.1. Pruebas de la media

Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu, \sigma^2)$, se obtendrán pruebas de la forma $\boldsymbol{\mathcal{H}}_0 : \mu \leqslant \mu_0$ vs $\boldsymbol{\mathcal{H}}_1 : \mu > \mu_0$ y de $\boldsymbol{\mathcal{H}}_0 : \mu = \mu_0$ vs $\boldsymbol{\mathcal{H}}_1 : \mu \neq \mu_0$, considerando en ambos casos si σ^2 es conocida o no.

1. Para probar $\boldsymbol{\mathcal{H}}_0 : \mu \leqslant \mu_0$ vs $\boldsymbol{\mathcal{H}}_1 : \mu > \mu_0$, considérese primero que σ^2 es conocido, sea esta σ_0^2. Entonces

$$L(\mu; x_1, ..., x_n) = \frac{1}{(\sqrt{2\pi}\sigma_0)^n} e^{-\frac{1}{2\sigma_0^2}\Sigma(x_i-\mu)^2}.$$

   Sea $\mu_1 < \mu_2$, entonces

$$\frac{L(\mu_1)}{L(\mu_2)} = e^{-\frac{1}{2\sigma_0^2}\sum\left[(x_i-\mu_1)^2-(x_i-\mu_2)^2\right]} = e^{\frac{\mu_1-\mu_2}{\sigma_0^2}\sum x_i} e^{-\frac{n}{2\sigma_0^2}(\mu_1^2-\mu_2^2)}$$

   Esta es una función decreciente de Σx_i así es que se rechaza $\boldsymbol{\mathcal{H}}_0$ si y solo si $\Sigma x_i > k^*$, donde k^* se obtiene de

$$\alpha = P_{\mu_0}\left[\Sigma X_i > k^*\right] = P_{\mu_0}\left[\frac{\Sigma X_i - n\mu_0}{\sqrt{n}\sigma_0} > \frac{k^* - n\mu_0}{\sqrt{n}\sigma_0}\right].$$

   Luego $1 - \alpha = \Phi\left(\frac{k^* - n\mu_0}{\sqrt{n}\sigma_0}\right)$; es decir que $z_{1-\alpha} = \frac{k^* - n\mu_0}{\sqrt{n}\sigma_0}$, o sea que $k^* = \sqrt{n}\sigma_0 z_{1-\alpha} + n\mu_0$.

Así que una PUMP de tamaño α es: rechazar $\mathcal{H}_0$ si $\Sigma x_i > n\mu_0 + \sqrt{n}\sigma_0 z_{1-\alpha}$, o equivalentemente, rechazar $\mathcal{H}_0$ si

$$\bar{x} > \mu_0 + \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha} \tag{10.9}$$

La función potencia de esta prueba es $\beta(\mu) = P(\bar{X} > \mu + \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha})$. Esta es una función creciente de μ y por tanto $\beta(\mu) > \beta(\mu_0) = \sup_{t \leq \mu_0} \beta(t)$ para todo $\mu > \mu_0$. Luego, la prueba es una prueba *insesgada* de tamaño α.

-Si σ^2 es desconocido, entonces la prueba es equivalente a probar $\mathcal{H}_0 : \theta \in \Theta_0$ vs $\mathcal{H}_1 : \theta \in \Theta - \Theta_0$, donde $\theta = (\mu, \sigma^2)$ y $\Theta = \{(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\}$.

Para obtener una prueba se busca algún estadístico que tenga comportamiento diferente bajo las dos hipótesis. Sea este $T = \dfrac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$, el cual tiene distribución t con $n-1$ grados de libertad.

Bajo $\mathcal{H}_0$ es $T < \dfrac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$ y bajo $\mathcal{H}_1$ es $T > \dfrac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$, es decir, T tiende a ser mayor para $\mu > \mu_0$ que para $\mu \leqslant \mu_0$. Luego, una prueba basada en T está dada por: rechazar $\mathcal{H}_0$ si $T > k$, donde k se obtiene de $\alpha = P_{\mu_0}(T > k)$. Como $T \sim t_{n-1}$, entonces $k = t_{(n-1,\ 1-\alpha)}$. Por tanto se rechaza $\mathcal{H}_0$ si $\dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} > t_{(n-1,\ 1-\alpha)}$, o equivalentemente rechazar $\mathcal{H}_0$ si,

$$\bar{x} > \mu_0 + \frac{s}{\sqrt{n}} t_{(n-1,\ 1-\alpha)}. \tag{10.10}$$

El estadístico de prueba en ambos casos es $T = \bar{X}$.

2. Para probar $\mathcal{H}_0 : \mu = \mu_0$ vs $\mathcal{H}_1 : \mu \neq \mu_0$, considérese primero que $\sigma^2 = \sigma_0^2$ y σ_0^2 es conocido. En este caso se sabe de 9.6 que un intervalo del $100\delta\,\%$ para μ es $(\bar{X} - \frac{\sigma_0}{\sqrt{n}} z_{\frac{1+\delta}{2}}, \bar{X} + \frac{\sigma_0}{\sqrt{n}} z_{\frac{1+\delta}{2}})$. Así que una prueba de tamaño $\alpha = 1 - \delta$ es rechazar $\mathcal{H}_0$ si μ_0 no pertenece a este intervalo. Una forma equivalente de la prueba es rechazar $\mathcal{H}_0$ si,

$$\bar{x} < \mu_0 - \frac{\sigma_0}{\sqrt{n}} z_{\frac{1+\delta}{2}} \ \text{ o } \ \bar{x} > \mu_0 + \frac{\sigma_0}{\sqrt{n}} z_{\frac{1+\delta}{2}}. \tag{10.11}$$

Puede observarse que $\frac{1+\delta}{2} = 1 - \frac{\alpha}{2}$, por tanto se rechaza $\boldsymbol{\mathcal{H}}_0$ si,

$$\bar{x} < \mu_0 - \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \text{ o } \bar{x} > \mu_0 + \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}}. \qquad (10.12)$$

-Si σ^2 es desconocido, se sabe también de 9.7 que un intervalo del $100\delta\,\%$ para μ es $(\bar{X} - \frac{S}{\sqrt{n}} t_{\frac{1+\delta}{2}}, \bar{X} + \frac{S}{\sqrt{n}} t_{\frac{1+\delta}{2}})$. Luego, una prueba de $\boldsymbol{\mathcal{H}}_0 : \mu = \mu_0$ vs $\boldsymbol{\mathcal{H}}_1 : \mu \neq \mu_0$ es rechazar $\boldsymbol{\mathcal{H}}_0$ si μ_0 no pertenece a este intervalo y el tamaño de la prueba es $\alpha = 1 - \delta$. El $t_{\frac{1+\delta}{2}}$ se obtiene de la distribución t con $n-1$ grados de libertad. De manera similar al caso anterior, una prueba de tamaño $\alpha = 1 - \delta$ es rechazar $\boldsymbol{\mathcal{H}}_0$ si,

$$\bar{x} < \mu_0 - \frac{s}{\sqrt{n}} t_{\frac{1+\delta}{2}} \text{ o } \bar{x} > \mu_0 + \frac{s}{\sqrt{n}} t_{\frac{1+\delta}{2}}. \qquad (10.13)$$

O de forma equivalente, se rechaza $\boldsymbol{\mathcal{H}}_0$ si,

$$\bar{x} < \mu_0 - \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2}} \text{ o } \bar{x} > \mu_0 + \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2}}. \qquad (10.14)$$

Se puede demostrar que esta prueba es también una prueba de la razón de verosimilitud generalizada.

Ejemplo 10.13. Para una muestra de tamaño 25 de una población normal con varianza $\sigma^2 = 36$ se encontró que la media $\bar{x} = 11.7$, se desea probar para $\alpha =0.05$, $\boldsymbol{\mathcal{H}}_0 : \mu \leq 10$ vs $\boldsymbol{\mathcal{H}}_1 : \mu > 10$.

Solución. De acuerdo a 10.9 se rechaza $\boldsymbol{\mathcal{H}}_0$ si, $\bar{x} > \mu_0 + \frac{\sigma_0}{\sqrt{n}} z_{0.95}$. Pero, $\mu_0 + \frac{\sigma_0}{\sqrt{n}} z_{0.95} = 10 + \frac{6}{5} * 1.645 =11.974$ y como $\bar{x}$ no es mayor que este número no se rechaza $\boldsymbol{\mathcal{H}}_0$.

10.4.2. Pruebas de la varianza

Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu, \sigma^2)$, se obtendrán pruebas de la forma $\boldsymbol{\mathcal{H}}_0 : \sigma^2 \leqslant \sigma_0^2$ vs $\boldsymbol{\mathcal{H}}_1 : \sigma^2 > \sigma_0^2$ y $\boldsymbol{\mathcal{H}}_0 : \sigma^2 = \sigma_0^2$ vs $\boldsymbol{\mathcal{H}}_1 : \sigma^2 \neq \sigma_0^2$, considerando en ambos casos si μ es conocida o no.

1. Para probar $\boldsymbol{\mathcal{H}}_0 : \sigma^2 \leqslant \sigma_0^2$ vs $\boldsymbol{\mathcal{H}}_1 : \sigma^2 > \sigma_0^2$, considere primero el caso en que $\mu = \mu_0$ y μ_0 es conocido. En el ejemplo 10.12 se encontró una prueba para σ^2 cuando $\mu = 0$. Sin embargo, $f(x,\theta)$=$f(x,\sigma^2)$ pertenece a la familia exponencial, pues se puede escribir en la forma $a(\sigma^2)b(x)e^{c(\sigma^2)d(x)}$, donde $a(\sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}}$, $b(x) = 1$, $c(\sigma^2) = -\frac{1}{2\sigma^2}$ y $d(x) = (x - \mu_0)^2$.

Como $c(\sigma^2)$ es una función creciente de σ^2, entonces una PUMP de tamaño α es rechazar $\boldsymbol{\mathcal{H}}_0$ si y solo si,

$$\Sigma(x_i - \mu_0)^2 \geq \sigma_0^2 \chi^2_{n(1-\alpha)}. \tag{10.15}$$

Observe que el estadístico de prueba es $T = \Sigma(X_i - \mu_0)^2$.

-Si μ_0 es desconocido, el estadístico $V = (n-1)\frac{S^2}{\sigma_0^2}$ que tiene distribución χ^2_{n-1} se comporta diferente bajo las dos hipótesis. Bajo $\boldsymbol{\mathcal{H}}_1$ tiende a ser mayor que bajo $\boldsymbol{\mathcal{H}}_0$. Así que una prueba de tamaño α es rechazar $\boldsymbol{\mathcal{H}}_0$ si $V > k$, donde k se obtiene de $\alpha = P_{\sigma_0^2}(V > k)$, es decir, que $k = \chi^2_{(n-1)(1-\alpha)}$. Por tanto se rechaza $\boldsymbol{\mathcal{H}}_0$ si $(n-1)\frac{s^2}{\sigma_0^2} > \chi^2_{(n-1)(1-\alpha)}$, o equivalentemente si,

$$s^2 > \left(\frac{\sigma_0^2}{n-1}\right)\chi^2_{(n-1)(1-\alpha)}. \tag{10.16}$$

o simplificando se rechaza $\boldsymbol{\mathcal{H}}_0$ si,

$$\Sigma(x_i - \bar{x})^2 > \sigma_0^2 \chi^2_{(n-1)(1-\alpha)}.$$

Es claro que $T = \Sigma(X_i - \bar{X})^2$ es el estadístico de prueba.

Esta es también una prueba de la razón de verosimilitud generalizada.

2. Para probar $\boldsymbol{\mathcal{H}}_0 : \sigma^2 = \sigma_0^2$ vs $\boldsymbol{\mathcal{H}}_1 : \sigma^2 \neq \sigma_0^2$, primero se considera que $\mu = \mu_0$, siendo μ_0 conocido. En este caso, de 9.9 se sabe que un intervalo del $100\delta\,\%$ para σ^2 es

$$\left(\frac{\Sigma(X_i - \mu_0)^2}{\chi^2_{n(\frac{1+\delta}{2})}}, \frac{\Sigma(X_i - \mu_0)^2}{\chi^2_{n(\frac{1-\delta}{2})}}\right).$$

De esta manera una prueba de tamaño $\alpha = 1 - \delta$ es rechazar $\boldsymbol{\mathcal{H}}_0$ si σ_0^2 no pertenece a este intervalo. Es decir, se rechaza $\boldsymbol{\mathcal{H}}_0$ si

$$\sigma_0^2 < \frac{\Sigma(x_i - \mu_0)^2}{\chi^2_{n(1-\alpha/2)}} \quad \text{o} \quad \sigma_0^2 > \frac{\Sigma(x_i - \mu_0)^2}{\chi^2_{n(\alpha/2)}}$$

-Si μ es desconocido, también de 9.11 se conoce que un intervalo de confianza del $100\delta\,\%$ para σ^2 es

$$\left(\frac{(n-1)S^2}{\chi^2_{(n-1)(\frac{1+\delta}{2})}}, \frac{(n-1)S^2}{\chi^2_{(n-1)(\frac{1-\delta}{2})}}\right).$$

Luego, una prueba de tamaño $\alpha = 1 - \delta$ de $\mathcal{H}_0 : \sigma^2 = \sigma_0^2$ vs $\mathcal{H}_1 : \sigma^2 \neq \sigma_0^2$ es rechazar $\mathcal{H}_0$ si σ_0^2 no pertenece a este intervalo. O equivalentemente rechazar $\mathcal{H}_0$ si

$$\sigma_0^2 < \frac{(n-1)s^2}{\chi^2_{(n-1)(1-\alpha/2)}} \quad \text{o} \quad \sigma_0^2 > \frac{(n-1)s^2}{\chi^2_{(n-1)\alpha/2}}.$$

Ejemplo 10.14. Considere una muestra de tamaño 16 de una población normal y suponga que $\bar{x} = 8$ y $s^2 = 16$. Probar $\mathcal{H}_0 : \sigma^2 \leqslant 12$ vs $\mathcal{H}_1 : \sigma^2 > 12$ con α =0.01.

Solución. Por lo que no se da información acerca de μ, para una prueba de $\mathcal{H}_0 : \sigma^2 \leqslant 12$ vs $\mathcal{H}_1 : \sigma^2 > 12$ con α =0.01 se debe usar 10.16. Es decir, se rechaza $\mathcal{H}_0$ si, $s^2 > \frac{12}{15}\chi^2_{15(0.99)}$. Pero $\chi^2_{15(0.99)}$ =30.6. Luego se rechaza $\mathcal{H}_0$ si s^2 >24.48. Puesto que $s^2 = 16$ no es mayor que 24.48, no se rechaza $\mathcal{H}_0$.

10.4.3. Pruebas de dos medias

En muchas situaciones es necesario comparar dos o más medias cuando no son conocidas. Aquí se considera el caso de dos medias de poblaciones normales.

Sean $X_{11}, X_{12}, ..., X_{1n}$ una muestra de $X_1 \sim N(\mu_1, \sigma_1^2)$ y $X_{21}, X_{22}, ..., X_{2m}$ una muestra de $X_2 \sim N(\mu_2, \sigma_2^2)$. Se desea probar hipótesis de la forma,

1. $\mathcal{H}_0 : \mu_1 \leq \mu_2$ vs $\mathcal{H}_1 : \mu_1 > \mu_2$,
2. $\mathcal{H}_0 : \mu_1 \geq \mu_2$ vs $\mathcal{H}_1 : \mu_1 < \mu_2$,
3. $\mathcal{H}_0 : \mu_1 = \mu_2$ vs $\mathcal{H}_1 : \mu_1 \neq \mu_2$.

Si las muestras son independientes y $\sigma_1^2 = \sigma_2^2$ se sabe del teorema 7.11 y su corolario que el estadístico,

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{1/n + 1/m}} \tag{10.17}$$

tiene distribución t con $m + n - 2$ grados de libertad ($T \sim t_{m+n-2}$). En particular, cuando $\mu_1 = \mu_2$

$$T^* = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{1/n + 1/m}} \sim t_{m+n-2} \tag{10.18}$$

donde

$$S_p^2 = \frac{\Sigma(X_{1i} - \bar{X}_1)^2 + \Sigma(X_{2i} - \bar{X}_2)^2}{m + n - 2}.$$

En la prueba de $\mathcal{H}_0 : \mu_1 \leq \mu_2$ vs $\mathcal{H}_1 : \mu_1 > \mu_2$, se observa que $T^* \leq T$ bajo $\mathcal{H}_0$, luego una prueba de tamaño α es: rechazar $\mathcal{H}_0$ si,

$$T^* > t_{(m+n-2)(1-\alpha)}. \tag{10.19}$$

En la prueba de $\mathcal{H}_0 : \mu_1 \geq \mu_2$ vs $\mathcal{H}_1 : \mu_1 < \mu_2$, se nota que $T^* > T$ bajo $\mathcal{H}_0$, luego una prueba de tamaño α es: rechazar $\mathcal{H}_0$ si,

$$T^* < t_{(m+n-2)(1-\alpha)}. \tag{10.20}$$

En la prueba de $\mathcal{H}_0 : \mu_1 = \mu_2$ vs $\mathcal{H}_1 : \mu_1 \neq \mu_2$ una prueba de tamaño α es: rechazar $\mathcal{H}_0$ si,

$$T^* \leq -t_{(m+n-2)(1-\alpha/2)} \text{ o } T^* \geq t_{(m+n-2)(1-\alpha/2)} \tag{10.21}$$

T^* dado por la expresión 10.18 es el estadístico de prueba.

Si las varianzas no son iguales, una prueba aproximada puede ser construida basada en la aproximación de Welch del estadístico T.

Ejemplo 10.15. Sean $X_1, X_2, ..., X_n$ y $Y_1, Y_2, ..., Y_m$ dos muestras aleatorias independientes de dos poblaciones normales $X \sim N(\mu_1, \sigma_1^2)$ y de $Y \sim N(\mu_2, \sigma_2^2)$, respectivamente. Si $n = 16, m = 12, \bar{x} = 15, \bar{y} = 9, s_1^2 = 25$ y $s_2^2 = 38$, probar bajo el supuesto de varianzas iguales $\mathcal{H}_0 : \mu_1 = \mu_2$ vs $\mathcal{H}_1 : \mu_1 \neq \mu_2$ con α =0.05.

Solución. Por 10.21 se rechaza $\mathcal{H}_0$ si $T^* \leq -t_{(m+n-2)(1-\alpha/2)}$ o $T^* \geq t_{(m+n-2)(1-\alpha/2)}$, donde T^* se obtiene 10.18. Con los datos del problema $T^* = \frac{6}{5.52\sqrt{\frac{1}{16}+\frac{1}{12}}}$ =2.85. Como $t_{(m+n-2)(1-\alpha/2)} = t_{26(0.975)} = 2.056$ y T^* es mayor que este valor se rechaza $\mathcal{H}_0$ con α =0.05.

10.4.4. Pruebas de dos varianzas

Sea $X_{11}, X_{12}, ..., X_{1n}$ una muestra de $X_1 \sim N(\mu_1, \sigma_1^2)$, y $X_{21}, X_{22}, ..., X_{2m}$ una muestra de $X_2 \sim N(\mu_2, \sigma_2^2)$. Se quiere probar hipótesis acerca de dos varianzas. En este texto se consideran pruebas de:

1. $\mathcal{H}_0 : \sigma_2^2 \leq \sigma_2^2$ vs $\mathcal{H}_1 : \sigma_1^2 > \sigma_2^2$,
2. $\mathcal{H}_0 : \sigma_2^2 \geq \sigma_2^2$ vs $\mathcal{H}_1 : \sigma_1^2 < \sigma_2^2$,
3. $\mathcal{H}_0 : \sigma_2^2 = \sigma_2^2$ vs $\mathcal{H}_1 : \sigma_1^2 \neq \sigma_2^2$.

Si las muestras son independientes, se sabe del corolario del teorema 7.12 que,

$$R = \frac{\Sigma(X_{1i} - \bar{X}_1)^2/(n-1)\sigma_1^2}{\Sigma(X_{2i} - \bar{X}_2)^2/(m-1)\sigma_2^2} \tag{10.22}$$

tiene una distribuion F con $n-1$ y $m-1$ grados de libertad, y en particular, el estadístico

$$R^* = \frac{(m-1)\Sigma(X_{1i} - \bar{X}_1)^2}{(n-1)\Sigma(X_{2i} - \bar{X}_2)^2} \tag{10.23}$$

tiene distribución F con $n-1$ y $m-1$ grados de libertad cuando $\sigma_1^2 = \sigma_2^2$.

Al probar $\boldsymbol{\mathcal{H}}_0 : \sigma_1^2 \leq \sigma_2^2$ vs $\boldsymbol{\mathcal{H}}_1 : \sigma_1^2 > \sigma_2^2$, se observa que $R = \frac{(m-1)\sum(X_{1i}-\bar{X}_1)^2\sigma_2^2}{(n-1)\sum(X_{2i}-\bar{X}_2)^2\sigma_1^2} \geq R^*$ si $\sigma_1^2 \leq \sigma_2^2$, una prueba de tamaño α es rechazar H_0 si,

$$R^* > F_{(n-1,m-1)(1-\alpha)}, \tag{10.24}$$

donde este F es el percentil de orden $1-\alpha$ de la distribución F con $n-1$ y $m-1$ grados de libertad.

En la prueba de $\boldsymbol{\mathcal{H}}_0 : \sigma_1^2 \geq \sigma_2^2$ vs $\boldsymbol{\mathcal{H}}_1 : \sigma_1^2 < \sigma_2^2$, se nota que $R \leq R^*$ si $\sigma_1^2 \geq \sigma_2^2$, así que una prueba de tamaño α es rechazar H_0 si,

$$R^* < F_{(n-1,m-1)(1-\alpha)}. \tag{10.25}$$

Una prueba de $\boldsymbol{\mathcal{H}}_0 : \sigma_1^2 = \sigma_2^2$ vs $\boldsymbol{\mathcal{H}}_1 : \sigma_1^2 \neq \sigma_2^2$, que es bilateral, es rechazar $\boldsymbol{\mathcal{H}}_0$ si R^* no pertenece al intervalo (k_1, k_2), donde $k_1 = F_{(n-1,m-1)(\frac{\alpha}{2})}$ y $k_2 = F_{(n-1,m-1)(1-\frac{\alpha}{2})}$.

Ejemplo 10.16. Con los datos del ejemplo anterior, probar $\boldsymbol{\mathcal{H}}_0 : \sigma_1^2 \geq \sigma_2^2$ vs $\boldsymbol{\mathcal{H}}_1 : \sigma_1^2 < \sigma_2^2$ con α =0.01.

Solución. Se calcula R^* que por 10.23 es $R^* = \frac{25}{38}$ =0.658. Se obtiene también del apéndice D que $F_{(m-1,n-1)(1-\alpha)} = F_{(15,11)(0.99)}$ ≈3.7. Luego por 10.25 se rechaza $\boldsymbol{\mathcal{H}}_0$ si $R^* < 3.7$. Como efectivamente R^* es menor que 3.7 se rechaza $\boldsymbol{\mathcal{H}}_0$ con α =0.01.

10.5. Pruebas de unión-intersección

El método de prueba de hipótesis de unión-intersección resulta útil cuando la hipótesis nula se puede expresar de manera conveniente como una intersección, tal como

$$\boldsymbol{\mathcal{H}}_0 : \theta \in \cap_{j \in I} \Theta_j, \tag{10.26}$$

donde I es un conjunto de índices que puede ser finito o infinito, dependiendo del problema.

Por ejemplo si $\boldsymbol{\mathcal{H}}_0 : \theta = \theta_0$, entonces $\boldsymbol{\mathcal{H}}_0 : \{\theta \leq \theta_0\} \cap \{\theta \geq \theta_0\}$ e $I = \{1, 2\}$.

Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x; \theta)$. Suponga que existen pruebas para cada uno de los problemas de pruebas $\boldsymbol{\mathcal{H}}_{0j} : \theta \in \Theta_j$ vs $\boldsymbol{\mathcal{H}}_{1j} : \theta \in \Theta'_j$. Si la región crítica para la prueba de $\boldsymbol{\mathcal{H}}_{0j}$ está determinada por el estadístico de prueba T_j, es decir, por el conjunto $\{(x_1, x_2, ..., x_n) : T_j(x_1, x_2, ..., x_n) \in \Omega_j\}$, entonces la región crítica para la prueba unión-intersección es

$$\cup_{j \in I} \{(x_1, x_2, ..., x_n) : T_j(x_1, x_2, ..., x_n) \in \Omega_j\}. \tag{10.27}$$

En efecto, si una de las hipótesis $\boldsymbol{\mathcal{H}}_{0j}$ es rechazada, entonces $\boldsymbol{\mathcal{H}}_0$ (la cual es verdadera si y solo si $\boldsymbol{\mathcal{H}}_{0j}$ es verdadera para toda j), debe ser rechazada. Solo si cada una de las hipótesis $\boldsymbol{\mathcal{H}}_{0j}$ es aceptada como verdadera, entonces la intersección $\boldsymbol{\mathcal{H}}_0$ será aceptada como verdadera.

Ejemplo 10.17. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $X \sim N(\mu, \sigma_0^2)$, donde σ_0^2 es conocido. Se quiere probar $\boldsymbol{\mathcal{H}}_0 : \mu = \mu_0$ vs $\boldsymbol{\mathcal{H}}_1 : \mu \neq \mu_0$ y μ_0 es un número específico.

Solución. Usando la prueba de la unión-intersección, la hipótesis nula se puede escribir como

$$\boldsymbol{\mathcal{H}}_0 : \{\mu : \mu \leq \mu_0\} \cap \{\mu : \mu \geq \mu_0\}.$$

Ahora bien, en la prueba de $\boldsymbol{\mathcal{H}}_{01} : \mu \leq \mu_0$ vs $\boldsymbol{\mathcal{H}}_{11} : \mu > \mu_0$ una prueba de tamaño α es rechazar $\boldsymbol{\mathcal{H}}_{01}$ si $\bar{x} > \mu_0 + \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha}$ (ver sección 10.4.1).

Similarmente, en la prueba $\boldsymbol{\mathcal{H}}_{02} : \mu \geq \mu_0$ vs $\boldsymbol{\mathcal{H}}_{12} : \mu < \mu_0$, una prueba de tamaño α es rechazar $\boldsymbol{\mathcal{H}}_{02}$ si $\bar{x} < \mu_0 - \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha}$.

Luego, la prueba de la unión-interseccón de $\boldsymbol{\mathcal{H}}_0 : \mu = \mu_0$ vs $\boldsymbol{\mathcal{H}}_1 : \mu \neq \mu_0$ es rechazar $\boldsymbol{\mathcal{H}}_0$ si

$$\bar{x} < \mu_0 - \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha/2} \quad \text{o} \quad \bar{x} > \mu_0 + \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha/2}. \tag{10.28}$$

En situaciones especiales se puede encontrar una expresión simple para la región crítica de la prueba. En particular, si cada prueba de $\boldsymbol{\mathcal{H}}_{0j}$ tiene una región crítica de la forma $\{(x_1, x_2, ..., x_n) : T_j(x_1, x_2, ..., x_n) > c\}$, donde c es una constante y no depende de j, entonces la región crítica de la prueba unión-intersección, dada 10.27 se puede expresar por

$$\cup_{j\in I}\{x : T_j(x) > c\} = \{x : \sup_{j\in I} T_j(x) > c\}, \quad \text{donde } x = (x_1, x_2, ..., x_n).$$

De manera que el estadístico de prueba para probar $\boldsymbol{\mathcal{H}}_0$ es $T(x) = \sup_{j\in I} T_j(x)$.

10.6. Pruebas binomiales

Sea $X_1, ..., X_n$ una muestra aleatoria de $X \sim B(1, p)$. Se quiere obtener pruebas para p. Estas se basan en el estadístico suficiente $S = \Sigma X_i \sim B(n, p)$.

Teorema 10.5. Sea $X_1, ..., X_n$ una muestra aleatoria de $X \sim B(1, p)$ y sea $S = \Sigma X_i$. Sea $Z_0 = \dfrac{S - np_0}{\sqrt{np_0(1-p_0)}}$, para n grande.

1. Una prueba aproximada de tamaño α de $\boldsymbol{\mathcal{H}}_0 : p \leqslant p_0$ *vs* $\boldsymbol{\mathcal{H}}_1 : p > p_0$ es rechazar $\boldsymbol{\mathcal{H}}_0$ si
$$z_0 > z_{1-\alpha}$$
o equivalentemente rechazar $\boldsymbol{\mathcal{H}}_0$ si,
$$\Sigma x_i > np_0 + \sqrt{np_0(1-p_0)}z_{1-\alpha} \quad \text{o si} \quad \bar{x} > p_0 + \sqrt{\frac{p_0(1-p_0)}{n}}z_{1-\alpha}.$$

2. Una prueba aproximada de tamaño α de $\boldsymbol{\mathcal{H}}_0 : p \geqslant p_0$ *vs* $\boldsymbol{\mathcal{H}}_1 : p < p_0$ es rechazar $\boldsymbol{\mathcal{H}}_0$ si
$$z_0 < -z_{1-\alpha}$$
o equivalentemente rechazar $\boldsymbol{\mathcal{H}}_0$ si,
$$\Sigma x_i < np_0 - \sqrt{np_0(1-p_0)}z_{1-\alpha} \quad \text{o si} \quad \bar{x} < p_0 - \sqrt{\frac{p_0(1-p_0)}{n}}z_{1-\alpha}.$$

3. Una prueba aproximada de tamaño α de $\boldsymbol{\mathcal{H}}_0 : p = p_0$ *vs* $\boldsymbol{\mathcal{H}}_0 : p \neq p_0$ es rechazar $\boldsymbol{\mathcal{H}}_0$ si
$$z_0 < -z_{1-\frac{\alpha}{2}} \quad \text{o} \quad z_0 > z_{1-\frac{\alpha}{2}}.$$

Este teorema es fácil de demostrar, pues cuando $n \to \infty$, $Z_0 \sim N(0, 1)$.

Teorema 10.6. Sea $X_1, ..., X_n$ es una muestra aleatoria de $X \sim B(1, p)$ y sea $S = \Sigma X_i$. Si $F_S(s) = P(S \leq s)$ entonces,

1. Una prueba conservativa[1] de tamaño α de $\mathcal{H}_0 : p \leqslant p_0$ *vs* $\mathcal{H}_1 : p > p_0$ es rechazar $\mathcal{H}_0$ si $1 - F_S(s-1) \leqslant \alpha$.

2. Una prueba conservativa de tamaño α de $\mathcal{H}_0 : p \geqslant p_0$ *vs* $\mathcal{H}_1 : p < p_0$ es rechazar $\mathcal{H}_0$ si $F_S(s) \leqslant \alpha$.

3. Una prueba conservativa de tamaño α de $\mathcal{H}_0 : p = p_0$ *vs* $\mathcal{H}_1 : p \neq p_0$ es rechazar $\mathcal{H}_0$ si $F_S(s) \leqslant \frac{\alpha}{2}$ o $1 - F_S(s-1) \leqslant \frac{\alpha}{2}$.

Ejemplo 10.18. Sea $X_1, ..., X_n$ es una muestra aleatoria de $X \sim B(1,p)$. Obtener una prueba conservativa de tamaño α=0.1 para probar $\mathcal{H}_0 : p \leq 0.2$ *vs* $\mathcal{H}_1 : p > 0.2$.

Solución. En el ejemplo 10.9 no se pudo obtener una PUMP de tamaño α=0.1 cuando se quiso probar $\mathcal{H}_0 : p \leqslant 0.2$ *vs* $\mathcal{H}_1 : p > 0.2$ y para una muestra de tamaño 15. De acuerdo a este teorema una prueba conservativa es rechazar $\mathcal{H}_0$ si $1 - P_{p=0.2}(\Sigma X_i \leq k^* - 1) \leq 0.1$ o equivalentemente rechazar $\mathcal{H}_0$ si $P_{p=0.2}(\Sigma X_i \leq k^* - 1) \geq 0.9$. Comparando con la tabla binomial para $n = 15$, se tiene que $k^* - 1 = 5$. De manera que se rechaza $\mathcal{H}_0$ si $\Sigma x_i \geq 6$.

10.7. Una prueba chi-cuadrado

Se presenta en esta sección una prueba de $\mathcal{H}_0 : p_i = p_{i0}$ para $i = 1, 2, \cdots, k$, donde $X_1, ..., X_k$ son variables aleatorias independientes tales que $X_i \sim B(n, p_i)$; $\Sigma_{i=1}^{k} p_i = 1$ y $\Sigma_{i=1}^{k} X_i = n$. Fue propuesta en 1910 por Karl Pearson [19] , antes que la teoría general de las pruebas de hipótesis fuera desarrollada. Esta prueba usa el estadístico

$$Q_{k-1,0} = \sum_{i=1}^{k} \frac{(X_i - np_{i0})^2}{np_{i0}}$$

el cual tiende a ser pequeño cuando $\mathcal{H}_0$ es verdadero y grande cuando $\mathcal{H}_0$ es falsa.

Teorema 10.7. Suponga que los resultados posibles de un experimento se incluyen en k conjuntos $A_1, A_2, \cdots, A_k$ tales que $A_i \cap A_j = \Phi, i \neq j$, y sea $P(A_i) = p_i$, para $i = 1, 2, \cdots, k$. Si el experimento se repite n veces sea X_i el número de resultados

[1] Prueba conservativa en el sentido según el cual el tamaño de la prueba puede ser menor que α para todos los valores del parámetro bajo $\mathcal{H}_0$

que caen en el conjunto A_i, de manera que $\Sigma_{i=1}^{k} X_i = n$, entonces

$$Q_{k-1} = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i} \tag{10.29}$$

tiene una distribución límite chi-cuadrado con $k-1$ grados de libertad cuando n tiende a infinito.

Demostración. Se hace la prueba para $k = 1$, la cual en si es elaborada como se observará.

Sea $Y = \frac{X_1 - np_1}{\sqrt{np_1(1-p_1)}}$. Por el teorema central del límite la distribución de Y tiende a la normal $(0,1)$, cuando n tiende a infinito.

Sea $Q_1 = Y^2$, entonces la distribución de Q_1 es aproximadamente chi-cuadrado con un grado de libertad.

En efecto, sea F_n la función de distribución de Y para todo entero positivo n, entonces $\lim\limits_{n\to\infty} F_n(y) = \Phi(y)$, $-\infty < y < \infty$.

Sea G_n la función de distribución de Q_1 para todo entero positivo n. Si $q_1 > 0$ entonces,

$$\begin{aligned} G_n(q_1) &= P(Q_1 \leqslant q_1) = P(Y^2 \leqslant q_1) = P(-\sqrt{q_1} \leqslant Y \leqslant \sqrt{q_1}) \\ &= F_n(\sqrt{q_1}) - F_n(-\sqrt{q_1}-) \ \text{ donde } F_n(w-) = P(Y < w). \end{aligned}$$

Como Φ es absolutamente continua entonces,

$$\begin{aligned} \lim_{n\longrightarrow\infty} G_n(q_1) &= \Phi(\sqrt{q_1}) - \Phi(-\sqrt{q_1}) \\ &= \int_{-\infty}^{\sqrt{q_1}} \phi(z)dz - \int_{-\infty}^{-\sqrt{q_1}} \phi(z)dz \\ &= \int_{-\sqrt{q_1}}^{\sqrt{q_1}} \phi(z)dz = 2\int_0^{\sqrt{q_1}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz. \end{aligned}$$

Si $v = z^2$, entonces $dv = 2zdz$; para $z = 0$ es $v = 0$ y para $z = \sqrt{q_1}$ es $v = q_1$, luego

$$\lim_{n\longrightarrow\infty} G_n(q_1) = \int_0^{q_1} \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{\pi}} \cdot \frac{1}{v^{\frac{1}{2}}} e^{-\frac{1}{2}v} dv = \int_0^{q_1} \frac{1}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} v^{-\frac{1}{2}} e^{\frac{1}{2}v} dv.$$

Si $q_1 < 0$, $G_n(q_1) \to 0$ cuando $n \to \infty$. De manera que el limite anterior es la función de distribución de una variable aleatoria que es χ^2_1.

Sean $X_2 = n - X_1$ y $p_2 = 1 - p_1$, luego $(X_1 - np_1)^2 = (n - X_2 - n + np_2)^2 = (X_2 - np_2)^2$. Así que

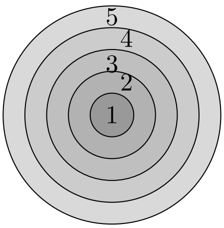
$$\begin{aligned} Q_1 = \frac{(X_1 - np_1)^2}{np_1(1-p_1)} &= \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_1 - np_1)^2}{n(1-p_1)} \\ &= \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - np_2)^2}{np_2} = \sum_{i=1}^{2} \frac{(X_i - np_i)^2}{np_i}. \end{aligned}$$

Es decir, Q_1 tiene una distribución aproximada chi-cuadrado con un grado de libertad.

Es de anotar que la aproximación de Q_{k-1} dada por 10.29 es buena cuando n es suficientemente grande de manera que $np_i \geqslant 5, i = 1, 2, \cdots, k$.

De esta manera, al probar $\boldsymbol{\mathcal{H}}_0 : p_1 = p_{10}$, $p_2 = p_{20}$, $\cdots$, $p_{k-1} = p_{k-1,0}$, $p_k = p_{k0} = 1 - p_{10} - p_{20} - \cdots - p_{k-1,0}$ contra cualquier otra alternativa, (donde $p_{10}, p_{20}, \cdots, p_{k-1,0}$ son conocidos), una prueba aproximada de tamaño α es: rechazar $\boldsymbol{\mathcal{H}}_0$ si $Q_{k-1} > c$, donde c se obtiene de $P(Q_{k-1} > c) = \alpha$. Es decir, se rechaza $\boldsymbol{\mathcal{H}}_0$ si el valor observado $q_{k-1,0} > \chi^2_{(k-1,\ 1-\alpha)}$.

Ejemplo 10.19. Una diana consta de 5 círculos concéntricos de radios 5, 10, 15, 20 y 25 centímetros, como se muestra en la figura. Un tirador experto dispara al blanco con probabilidades $0.15, 0.15, 0.25, 0.25$ y 0.20 de dar en $1, 2, 3, 4$ y 5, respectivamente. Si el tirador dispara 100 veces al blanco en forma independiente y acierta 10 veces en 1, 18 en 2, 26 en 3, 30 en 4 y 16 en 5, ¿hay alguna evidencia de rechazar la hipótesis $\boldsymbol{\mathcal{H}}_0 : p_1 = 0.15$, $p_2 = 0.15$, $p_3 = 0.25$, $p_4 = 0.25$,$p_5 = 0.20$? Use $\alpha = 0.05$.



Solución. Sea X_i la frecuencia con que el tirador da en el blanco $i, i = 1, 2, 3, 4, 5$. Entonces,

$$Q_{4,0} = \sum_{i=1}^{5} \frac{(X_i - np_{i0})^2}{np_{i0}}$$

tiene una distribución χ^2_4 aproximada, y

$$q_{4,0} = \frac{(10-15)^2}{15} + \frac{(18-15)^2}{15} + \frac{(26-25)^2}{25} + \frac{(30-25)^2}{25} + \frac{(16-20)^2}{20} = 4.107.$$

Se rechaza $\boldsymbol{\mathcal{H}}_0$ si $q_{4,0} > c$ donde c se obtiene de $P(Q_{4,0} > c) = 0.05$. De la tabla de la distribución χ^2 se observa que $c = 9.49$. Como $q_{4,0} < 9.49$ no se rechaza $\boldsymbol{\mathcal{H}}_0$.

10.8. Ejercicios

1. Sea $X_1, X_2, ..., X_{10}$ una muestra aleatoria de una distribución de Bernoulli con parámetro p.

   a) Encuentre la prueba mas potente de tamaño $\alpha = 0.0547$ de la hipótesis $\boldsymbol{\mathcal{H}}_0 : p = 1/2$ vs $\boldsymbol{\mathcal{H}}_1 : p = 1/4$.

   b) Para probar $\boldsymbol{\mathcal{H}}_0 : p \leq 1/2$ vs $\boldsymbol{\mathcal{H}}_1 : p > 1/2$, encontrar el tamaño y la función potencia de la prueba que rechaza H_0 si $\Sigma_{i=1}^{10} X_i \geq 6$.

2. Sean $X_1, ..., X_{10}$ variables aleatorias independientes tales que $X_i \sim B(1, p)$.

   a) Encontrar la prueba más potente de tamaño $\alpha = 0.0547$ de $\boldsymbol{\mathcal{H}}_0 : p = 1/2$ vs $\boldsymbol{\mathcal{H}}_1 : p = 1/4$. Encuentre la potencia de esta prueba.

   b) Para probar $\boldsymbol{\mathcal{H}}_0 : p \leq 1/2$ vs $\boldsymbol{\mathcal{H}}_1 : p > 1/2$, halle el tamaño de la prueba y dibuje la función potencia de la prueba que rechaza $\boldsymbol{\mathcal{H}}_0$ si $\sum_{i=1}^{10} x_i \geq 6$.

3. Sea $X_1, ..., X_n$ una muestra aleatoria de $f(x : \lambda) = \frac{\lambda^x e^\lambda}{x!}$ para $x = 0, 1, ...$, se quiere probar $\boldsymbol{\mathcal{H}}_0 : \lambda = \lambda_0$ vs $\boldsymbol{\mathcal{H}}_1 : \lambda = \lambda_1$ con $\lambda_1 < \lambda_2$. Obtenga la prueba más potente.

4. Una caja contiene 6 fusibles de autos, θ de 15 y $6 - \theta$ de 20, probar $\boldsymbol{\mathcal{H}}_0 : \theta = 3$ vs $\boldsymbol{\mathcal{H}}_1 : \theta \neq 3$ como sigue: tomar 3 fusibles con reemplazamiento y rechazar $\boldsymbol{\mathcal{H}}_0$ si los tres son del mismo color; de otra manera no se rechaza.

   a) Calcular la probabilidad del error tipo I.

   b) Hacer lo mismo si los fusibles son tomados sin reemplazamiento.

5. Sea $X_1, ..., X_n$ una muestra aleatoria de $f(x; \theta) = \theta x^{\theta - 1}$ si $0 \leq x \leq 1$ y $\theta > 0$.

   a) Si $n = 1$ obtenga la prueba más potente de $\boldsymbol{\mathcal{H}}_0 : \theta = 1$ vs $\boldsymbol{\mathcal{H}}_1 : \theta = 2$ con $\alpha = 0.05$. Calcule la potencia de la prueba para $\boldsymbol{\mathcal{H}}_0$.

   b) Obtenga la prueba más potente para la muestra.

6. Sea X una variable aleatoria discreta. Obtenga la prueba más potente de tamaño $\alpha = 0.0975$ de $\boldsymbol{\mathcal{H}}_0 : X \sim G(0.05)$ vs $\boldsymbol{\mathcal{H}}_1 : X \sim P(0.95)$ con base en un valor observado de x.

7. Sea $X_1, ..., X_n$ una muestra aleatoria de $X \sim U(0, \theta)$. Encuentre una PUMP de tamaño α de $\boldsymbol{\mathcal{H}}_0 : \theta \geq \theta_0$ vs $\boldsymbol{\mathcal{H}}_1 : \theta < \theta_0$.

8. Para el ejemplo 10.7 hacer la prueba considerando ahora que $\sigma_1^2 \leq \sigma_2^2$.

9. Sea $X_1, ..., X_n$ una muestra aleatoria de $f(x; \theta)$ y sea S un estadístico suficiente para θ. Demuestre que una prueba más potente de $\boldsymbol{\mathcal{H}}_0 : \theta = \theta_0$ vs $\boldsymbol{\mathcal{H}}_1 : \theta = \theta_1$ se puede expresar en términos de S.

10. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una distribución normal con media μ y varianza $\sigma^2 = 25$. Se desea probar $\boldsymbol{\mathcal{H}}_0 : \mu = 10$ vs $\boldsymbol{\mathcal{H}}_1 : \mu = 5$. Encuentre el tamaño muestral n de manera que la prueba más potente tenga $\alpha = \beta = 0.025$, donde α y β son los tamaños de error tipo I y II respectivamente.

11. Sea $X_1, ..., X_n$ una muestra aleatoria de $f(x; \theta) = \theta x^{\theta - 1}$ si $0 \leq x \leq 1$ y $\theta > 0$. Encuentre la prueba de Bayes correspondiente a la distribución a priori dada por $g = \frac{2}{3}$ de $\boldsymbol{\mathcal{H}}_0 : \theta = 1$ vs $\boldsymbol{\mathcal{H}}_1 : \theta = 2$, usando la función de pérdida $l(d_0; \theta_1) = \frac{1}{2}$ y $l(d_1; \theta_0) = 1$.

12. Sea $X_1, ..., X_{36}$ una muestra aleatoria de $X \sim N(\mu, 4)$. Probar $\boldsymbol{\mathcal{H}}_0 : \mu = 25$ vs $\boldsymbol{\mathcal{H}}_1 : \mu = 26$.

13. Para una muestra de tamaño 40 de una población normal con media μ y varianza σ^2 conocida, graficar la función potencia de la prueba de la razón de verosimilitud de $\boldsymbol{\mathcal{H}}_0 : \mu \leq 0$ vs $\boldsymbol{\mathcal{H}}_1 : \mu > 0$.

14. Sea $X_1, ..., X_{36}$ una muestra aleatoria de $X \sim N(\mu, 4)$. Encontrar una PUMP de

    a) $\boldsymbol{\mathcal{H}}_0 : \mu = \mu_0$ vs $\boldsymbol{\mathcal{H}}_1 : \mu < \mu_0$.
    b) $\boldsymbol{\mathcal{H}}_0 : \mu = \mu_0$ vs $\boldsymbol{\mathcal{H}}_1 : \mu > \mu_0$.
    c) Demostrar que no existe una PUMP de $\boldsymbol{\mathcal{H}}_0 : \mu = \mu_0$ vs $\boldsymbol{\mathcal{H}}_1 : \mu \neq \mu_0$.

15. Sea X una variable aleatoria continua con función de densidad $f(x; \theta) = \left[1 - (x - \frac{1}{2})\theta^2\right] I_{(0,1)}(x)$, con $-1 < \theta < 1$. Probar que una PUMP de tamaño α de $\boldsymbol{\mathcal{H}}_0 : \theta = 0$ vs $\boldsymbol{\mathcal{H}}_1 : \theta \neq 0$, basado en un valor observado x de X es rechazar $\boldsymbol{\mathcal{H}}_0$ si $x \leq \alpha$.

16. Dada una muestra de tamaño 100 con $\bar{x} = 2.7$ y $\sum_{i=1}^{100}(x_i - \bar{x})^2 = 225$, probar la hipótesis nula $\boldsymbol{\mathcal{H}}_0 : \mu = 3$ y $\sigma^2 = 2.5$ en el nivel $\alpha = 0.01$. Suponga que la población es normal.

17. Dada las muestras 2.3, 1.6, 2.8, 1.2, 1.6 y 3.4, 2.9, 5.0, 0.8, 1.7, 4.7 de poblaciones normales, probar la hipótesis de que las varianzas son iguales.Tome $\alpha = 0.05$.

18. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta) = \frac{1}{\theta}x^{\frac{1-\theta}{\theta}}I_{(0,1)}(x)$. Obtenga una prueba uniformemente más potente de tamaño α de $\mathcal{H}_0 : \theta \leq \theta_0$ vs $\mathcal{H}_1 : \theta > \theta_0$.

19. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de $f(x;\theta) = \frac{3x^2}{\theta}e^{-\frac{x^3}{\theta}}I_{(0,\infty)}(x)$. Obtenga la forma de la región crítica para una prueba uniformemente mas potente de tamaño α de $\mathcal{H}_0 : \theta = \theta_0$ vs $\mathcal{H}_1 : \theta > \theta_0$.

20. Sea $X_1, ..., X_9$ una muestra aleatoria de $X \sim N(\mu_1, \sigma_1^2)$ y sea $Y_1, ..., Y_{16}$ una muestra aleatoria de $Y \sim N(\mu_2, \sigma_2^2)$ y suponga que las muestras son independientes. Sea $\bar{x} = 16$, $\bar{y} = 20$, $S_1^2 = 36$ y $S_2^2 = 64$.

    a) Asumiendo varianzas iguales, probar $\mathcal{H}_0 : \mu_1 = \mu_2$ vs $\mathcal{H}_1 : \mu_1 \neq_2$ con $\alpha = 0.05$.
    b) Probar $\mathcal{H}_0 : \sigma_1^2 \geq \sigma_2^2$ vs $\mathcal{H}_1 : \sigma_1^2 < \sigma_2^2$ con $\alpha = 0.1$.

21. Sea $X_1, ..., X_n$ una muestra aleatoria de $X \sim E(\theta)$ y sea $Y_1, ..., Y_m$ una muestra aleatoria de $Y \sim E(\lambda)$. Si las muestras son independientes,

    a) Encuentre la prueba de la razón verosimilitud de $\mathcal{H}_0 : \theta = \lambda$ vs $\mathcal{H}_1 : \theta \neq \lambda$.
    b) Demuestre que la prueba anterior está basada en el estadístico $T = \frac{\sum X_i}{\sum X_i + \sum Y_j}$.

22. Para los dos conjuntos de datos del ejercicio 1.17 probar: la media de la rentabilidad mensual es menor o igual a la media de la rentabilidad semestral. Use $\alpha = 0.05$.

23. Una moneda se lanza 20 veces y se observan 6 caras. Sea $p = P(cara)$. Obtener una prueba de tamaño a lo sumo de 0.1 de $\mathcal{H}_0 : p \geq 0.5$ vs $\mathcal{H}_1 : p < 0.5$. ¿Cual es la potencia de una prueba de tamaño $\alpha = 0.0577$ de $\mathcal{H}_0 : p \geq 0.5$.

Apéndice A

Análisis combinatorio

Muchos problemas en teoría de probabilidad pueden ser resueltos por conteo simple del número de caminos que en un cierto evento pueden ocurrir. La teoría de conteo es conocida formalmente como *Análisis combinatorio*.

Si $a_1, a_2, \cdots, a_n$ son números, entonces

1. La suma $a_1 + a_2 + \cdots + a_n$ se representa por $\sum\limits_{i=1}^{n} a_i$

2. El producto $a_1 a_2 \cdots a_n$ se simboliza por $\prod\limits_{i=1}^{n} a_i$

Algunas sumas útiles son

$$\sum_{i=1}^{n} i = \tfrac{n(n+1)}{2}, \quad \sum_{i=1}^{n} i^2 = \tfrac{n(n+1)(2n+1)}{6},$$
$$\sum_{i=1}^{n} i^3 = \left(\tfrac{n[n+1]}{2}\right)^2 \text{ y } \sum_{k=0}^{n-1} a^k = \tfrac{1-a^n}{1-a}.$$

A.1. Principio básico de conteo

Este principio de conteo es básico para toda esta sección y establece que:

"Si un experimento consta de k etapas tal que la i-ésima $(i = 1, 2, ..., k-1)$ tiene n_i resultados posibles, y si con cada resultado de esta etapa, la etapa $i+1$ tiene

n_{i+1} *resultados posibles, entonces el experimento consta de* $n_1\ n_2 \cdots n_k$ *resultados posibles".*

Ejemplo. Un comité de personas consiste de un hombre y una mujer, estos van a ser escogidos de un grupo de 7 hombres y 6 mujeres. Cuántos comités posibles pueden ser escogidos?

Solución. Mirando la escogencia del hombre como el resultado del experimento 1 y la escogencia de la mujer como el resultado del experimento 2, se ve del principio básico de conteo que hay $7 \times 6 = 42$ comités posibles.

Ejemplo. Un colegio plantea comités consistentes de un estudiante de cada clase: 3 de primero, 4 de segundo, 5 de tercero, 2 de cuarto y 3 de quinto. Cuántos subcomités pueden construirse?

Solución. Del principio básico de conteo generalizado, se sigue que hay $3 \times 4 \times 5 \times 2 \times 3 = 360$ comités posibles.

Ejemplo. ¿Cuántas placas de automóviles se pueden hacer si las tres primeras posiciones son formadas por letras y las tres restantes por números?

Solución. Las tres primeras posiciones pueden ser ocupadas por cada una de las 26 letras del alfabeto y las tres siguientes pueden ser ocupadas cada una por 10 dígitos. Aplicando el principio de conteo generalizado, hay $26 \times 26 \times 26 \times 10 \times 10 \times 10 = 17'576.000$ placas.

¿Cuántas placas son posibles si la repetición entre letras y números fueran prohibidas?

A.2. Permutaciones

Cada arreglo diferente que se puede hacer de un conjunto de n objetos diferentes, se llama *permutación*. Por ejemplo, de las letras a, b, c se pueden hacer los siguientes arreglos: $abc, acb, bca, bac, cab, cba$. Este número de arreglos diferentes se puede obtener por el principio básico de conteo, ya que la primera posición puede ser ocupada por las tres letras, y una vez ocupada la primera, la segunda puede ser ocupada por dos letras y finalmente, la tercera por 1, es decir, hay $3 \times 2 \times 1 = 6$ permutaciones posibles.

Si tenemos n objetos diferentes, el razonamiento previo es aplicable y por tanto el número de permutaciones que hay es $n(n-1)\cdot ... \cdot 2 \cdot 1 = n!$.

El número de permutaciones de n objetos tomadolos todos, se nota por $P_{n,n}$ y es igual a $P_{n,n} = n!$.

Ejemplo. La clase de estadística consiste de 8 hombres y 5 mujeres. Un exámen es realizado y los estudiantes son ordenados de acuerdo a su nota. Si ningún estudiante tiene su calificación repetida con otros:

a) Cuántos arreglos diferentes son posibles?

b) Si los hombres son arreglados entre sí y las mujeres arregldas entre sí, ¿cuántos ordenes son posibles?

Solución.

a) En este caso hay 13 notas diferentes y por tanto el número de órdenes posibles es $P_{13,13} = 13! = 6.227'020.800$.

b) En el segundo caso, el número de órdenes entre los hombres es $P_{8,8} = 8! = 40.320$; y el número de órdenes entre las mujeres es $P_{5,5} = 5! = 120$. Ahora, por el pricipio básico de conteo, hay $P_{8,8} \times P_{5,5} = 4'838.400$.

Ahora, si del número n de objetos diferentes son tomados r objetos a un tiempo para cada permutación, entonces el número de permutaciones de r objetos tomados del conjunto de n objetos, se un fijo nota $P_{n,r}$ y es:

$$P_{n,r} = n \cdot (n-1) \cdot (n-2) \cdot ... \cdot (n-(r-1)) \tag{A.1}$$

Puede mostrarse que

$$P_{n,r} = \frac{n!}{(n-r)!} \tag{A.2}$$

Ejemplo. Cuántos órdenes son posibles de las letras a, b, c, d tomadas de dos en dos?

Solución. Hay 4 letras en el conjunto y se van a ordenar de dos en dos. Luego, hay:

$$P_{4,2} = \frac{4!}{2!} = \frac{24}{2} = 12$$

Por lo tanto hay 12 órdenes diferentes.

Ejemplo. Tres profesores van a ser elegidos para el fondo de empleados de la Universidad para los cargos de presidente, vicepresidente y secretario, de 25 miembros. En cuantos caminos pueden ser elegidos los profesores?

Solución. Este problema se puede resolver por el principio básico de conteo, pues para presidente hay 25 posibilidades, una vez elegido hay 24 posibilidades de elegir vicepresidente y finalmente, quedan 23 opciones de elegir secretario. En total hay $25 \times 24 \times 23 = 13.800$ formas para elegir los tres profesores.

Por otro lado este es un problema de permutación, puesto que el orden es tenido en cuenta, ya que si el presidente es A, el vicepresidente es B y es secretario es C, es diferente al conjunto si el presidente es B, el vicepresidente C y el secretario A.

Luego si $n = 25$ y $r = 3$, se tiene que:

$$P_{25,3} = \frac{25!}{22!} = 25 \times 24 \times 23 = 13.800$$

Entonces hay 13.800 formas para elegir los tres profesores.

Además, se puede establecer que si n objetos existen, de los cuales n_1 son iguales entre sí, n_2 son iguales entre sí y, ..., n_r son iguales entre sí, entonces el número de permutaciones diferentes es:

$$P_{n;n_1,n_2,\ldots,n_r} = \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_r!} \tag{A.3}$$

Ejemplo. ¿De cuántas maneras pueden ser arregladas 4 monedas de 100 pesos, 3 de 200 pesos y 5 de 500 pesos, en una fila, si las monedas similares son indistinguibles unas de otras?

Solución. Se observa que hay 12 monedas de las cuales 4 son de 100 pesos, 3 de 200 pesos y 5 de 500 pesos, por lo tanto:

$$P_{12;4,3,5} = \frac{12!}{4! \cdot 3! \cdot 5!} = 27.720$$

De ahí que existen 27.720 arreglos posibles.

A.3. Combinaciones

Una combinación es un subconjunto o un arreglo de todos o parte de los objetos de un conjunto en particular, sin importar el orden. El número total de combinaciones posibles de un conjunto de objetos tomados todos a la vez, es 1.

El número de combinaciones posibles de un conjunto de objetos diferentes tomados parte a la vez, puede ser obtenido encontrando el número de permutaciones totales y luego encontrar las permutaciones con los minsmos objetos como una combinación. El siguiente ejemplo lo ilustra.

Ejemplo. Encontrar el número de combinaciones del conjunto de 3 letras tomadas de a dos.

Solución. El número total de permutaciones de las tres letras tomadas a la vez es:

$$P_{3,2} = \frac{3!}{1!} = 6$$

Cada combinación consiste de dos letras. El número total de permutaciones de las mismas dos letras tomadas a la vez es $P_{2,2} = 2! = 2$.

Las dos permutaciones consistentes de las mismas letras son consideradas como una sola combinación. Así que el número de combinaciones es:

$$\frac{P_{3,2}}{P_{2,2}} = \frac{6}{2} = 3$$

El número de combinaciones de n objetos de un conjunto, tomando r elementos a la vez, se denota por $C_{n,r}$ y razonando como anteriormente:

$$C_{n,r} = \frac{P_{n,r}}{P_{r,r}} = \frac{n!}{r! \cdot (n-r)!} = \binom{n}{r} \tag{A.4}$$

Ejemplo. Un comité de cuatro miembros es formado de un grupo de 15 personas. Cuántos comités diferentes son posibles?

Solución. No importa el orden, ya que el comité formado por las personas A, B, C, D es el mismo que el formado por D, C, A, B, o por cualquier otro arreglo de estas. Por lo tanto:

$$C_{15,4} = \frac{15!}{4! \cdot 11!} = 1365$$

Hay entonces, 1365 comités diferentes.

A.4. Teorema del binomio

Si a y b son reales y n es un entero positivo, entonces

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Consecuencias de este teorema son

1. $2^n = \sum_{k=0}^{n} \binom{n}{k}$

2. $\sum_{k=0}^{n} \binom{a}{k} \binom{b}{n-k} = \binom{a+b}{n}$

3. Una generalización del teorema del binomio es

$$\left(\sum_{j=1}^{n} a_j\right)^2 = \left(\sum_{i=1}^{n} a_i\right)\left(\sum_{j=1}^{n} a_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j$$

4. También

$$\left(\sum_{i=1}^{m} a_i\right)\left(\sum_{j=1}^{n} a_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i a_j$$

Apéndice B

Conceptos del cálculo

Con el objeto de facilitar el estudio de este texto se dan algunos conceptos fundamentales del cálculo necesarios.

Definición de función. Una función f de A en B es una regla (ley, fórmula, receta) que asocia a cada punto $a \in A$ con uno y sólo un punto $b \in B$. Si f es una función de A en B, y $y \in B$ tal que corresponde a algún $x \in A$, entonces $y = f(x)$.

Algunas funciones especiales que se cosideran en este texto se dan a cotinuación.

Función indicadora. Sea S cualquier espacio (conjunto) con puntos s y A cualquier subconjunto de S. La función indicadora de A, denotada por $I_A(.)$ es la función que tiene dominio en S y codominio en el conjunto $\{0, 1\}$ y está definida por:

$$I_A(s) = \begin{cases} 1, & \text{si } s \in A \\ 0, & \text{si } s \notin A \end{cases} \tag{B.1}$$

Propiedades
Sea S cualquier espacio y sea $\mathcal{F}$ una colección de subconjuntos de S, entonces

a) $I_A(s) = 1 - I_{A'}(s)$, para todo $A \in \mathcal{F}$.

b) $I_{A_1 \cap A_2 \cap A_3 \ldots \cap A_n}(s) = I_{A_1}(s) \cdot I_{A_2}(s) \ldots I_{A_n}(s)$, para cada $A_1, A_2, ..., A_n \in \mathcal{F}$.

c) $I_{A_1 \cup A_2 \cup A_3 \ldots \cup A_n}(s) = \text{máx}\, \{I_{A_1}(s), I_{A_2}(s), ..., I_{A_n}(s)\}$, para cada $A_1, A_2, ...,$ $A_n \in \mathcal{F}$.

d) $I_A^2(s) = I_A(s)$, para todo $A \in \mathcal{F}$.

Ejemplo. La función $f(.)$ definida por, indicadora.en términos de la función

$$f(x) = \begin{cases} 0, & \text{para } x \leq 0 \\ x, & \text{para } 0 < x \leq 1 \\ 2-x, & \text{para } 1 < x \leq 2 \\ 0, & \text{para } x > 2 \end{cases}$$

en término de la función indicadora es . Luego,

$$f(x) = xI_A(x) + (2-x)I_B(x)$$

donde, $A = \{x|0 < x \leq 1\},\ B = \{x|1 < x \leq 2\}$
Es decir, que:

$$f(x) = xI_{(0,1)}(x) + (2-x)I_{(1,2)}(x)$$

.

Función convexa. Una función $g: R \to R$ continua se dice convexa si para todo $x_0, x_1 \in R$,

$$g(\lambda x_0 + (1-\lambda)x_1) \leq \lambda g(x_0) + (1-\lambda)g(x_1) \quad \text{para todo} \quad \lambda \in [0,1]$$

Son funciones convexas x^2, e^{-x} y $\sqrt{x}$ para $x > 0$.

Puede probarse que una función g dos veces diferenciable es convexa si $g''(x) \geq 0$ para todo x.

Función parte entera. La función $f: R \longrightarrow N$ definida por

$$f(x) = [x]$$

es el mayor entero que es menor o igual a x, es decir

$$[x] \leq x < [x] + 1$$

Función beta. La función $f: R^+ x R^+ \longrightarrow R$ definida por

$$\beta(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$$

se llama la función beta(de Euler).

Función gama. La función $f : R^+ \longrightarrow R$ definida por

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$$

se llama la función gama(de Euler).

La función gama satisface,

a) $\Gamma(x+1) = t\Gamma(x)$, para todo $x > 0$.
Esta propiedad permite extender la definición para los $x < 0$ con $x \neq -1, -2, -3, \cdots$.
por ejemplo para $x \in (-1, 0)$

$$\Gamma(x) = \frac{\Gamma(x+1)}{x}$$

b) $\Gamma(n+1) = n!$, si $n \in \mathbb{N}$

c) $2^{2x-1}\Gamma(x)\Gamma(x+\frac{1}{2}) = \sqrt{\pi}\Gamma(2x)$ (Fórmula de Duplicación).
De este resultado se tiene que $\Gamma(1/2) = \sqrt{\pi}$.

d) $\beta(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$.

Número e de Euler. El número trascendente e es el resultado del siguiente limite:

$$\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x, \text{ o, } \lim_{x\to 0}\left(1+x\right)^{\frac{1}{x}}$$

Serie de Taylor. Sea f una función con $n+1$ derivadas continuas alrededor de $x = a$, entonces

$$f(x) = \sum_{k=0}^{n} \frac{f'(a)(x-a)^k}{k!} + R_n$$

donde $f^0(a) = f(a)$, $f^i(a) = \frac{d^i f(x)}{dx^i}|_{x=a}$ y R_n es el residuo.

Si $a = 0$ la serie se convierte en la serie de Mac-Laurin

$$f(x) = \sum_{k=0}^{n} \frac{f'(0)x^k}{k!} + R_n$$

Ejemplo. La función e^x se representa por la serie

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

Cuando $x = 1$ se tiene que una aproximación de e con 7 cifras decimales es 2.7182818

Constante de Euler. La constante γ de Euler es el resultado de,

$$\gamma = \lim_{n\to\infty} \left[\sum_{k=1}^{n} \frac{1}{k} - ln(n) \right] = \int_1^{\infty} \left(\frac{1}{[x]} - \frac{1}{x} \right)$$

Un valor aproximado de γ con 7 cifras decimales es 0.5772157.

Apéndice C

Resumen de distribuciones especiales

En esta tabla se hace un resumen de todas las distribuciones contenidas en el capítulo IV. Se encuentran la función de masa o densidad (según el caso), la de distribución, la esperanza, la varianza, la función generadora de momentos y en algunos casos el momento de orden k.

Resumen de Distribuciones Especiales	
Nombre	**Características**
Uniforme discreta	$p_X(x) = \frac{1}{N} I_{\{1,2,\cdots,N\}}(x)$ $E(X) = \frac{N+1}{2},\ V(X) = \frac{N^2-1}{12}$
Bernoulli(p)	$p_X(x) = p^x(1-p)^{1-x} I_{\{0,1\}}(x)$ $E(X) = p$, $V(X) = p(1-p)$ $M(t) = pe^t + 1 - p$
Binomial(n, p)	$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} I_{\{0,1,\cdots,n\}}(x)$ $E(X) = np$, $V(X) = np(1-p)$ $M(t) = [pe^t + 1 - p]^n$
Poisson(λ)	$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!} I_{\{0,1,\cdots,\}}(x)$ $E(X) = \lambda$, $V(X) = \lambda$ $M(t) = e^{\lambda e^{(t-1)}}$
Geométrica(p)	$p_X(x) = p(1-p)^{x-1} I_{\{1,2,\cdots,\}}(x)$ $E(X) = \frac{1}{p}\ ,\ V(X) = \dfrac{1-p}{p^2}$ $M(x) = \frac{pe^t}{1-(1-p)e^t}$
Binomial negativa(r, p)	$p_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} I_{\{r,r+1,\cdots,\}}(x)$ $E(X) = \frac{r}{p},\ V(X) = \frac{r(1-p)}{p^2}$ $M(t) = \left[\frac{pe^t}{1-(1-p)e^t}\right]^r$
Hipergeométrica(M, k, n)	$p_X(x) = \dfrac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} I_{\{0,1,\cdots,min\{K,n\}\}}(x)$ $E(X) = \frac{nK}{M}$, $V(X) = \dfrac{nK(M-K)(M-n)}{M^2(M-1)}$
Uniforme(a, b)	$f_X(x) = \frac{1}{b-a} I_{(a,b)(x)}$ $E(X) = \frac{a+b}{2}$, $V(X) = \frac{(b-a)^2}{12},\ M(t) = \frac{e^{bt}-e^{at}}{t(b-a)}$

Resumen de Distribuciones Especiales	
Nombre	**Características**
Normal(μ,σ^2)	$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(x-\mu)^2},\ -\infty<x<\infty$ $E(X)=\mu,\ V(X)=\sigma^2$ $M(t)=e^{t\mu+\frac{1}{2}t^2\sigma^2},\ \mu_4=3\sigma^4$
Exponencial(λ)	$f_X(x)=\lambda e^{-\lambda x}I_{(0,\infty)}(x),\ F(x)=1-e^{-\lambda x}I_{[0,\infty)}(x)$ $E(X)=\frac{1}{\lambda},\ V(X)=\frac{1}{\lambda^2},\ M(t)=\frac{\lambda}{\lambda-t},\ \lambda>t$
Gama(r,λ)	$f_X(x)=\dfrac{\lambda e^{-\lambda x}(\lambda x)^{r-1}}{\Gamma(r)}I_{(0,\infty)}(x),$ $F(x)=\sum_{j=0}^{n-1}\frac{e^{-\lambda x}(\lambda x)^j}{j!},$ si $r=n,$ y , $n\in Z^+$ $E(X)=\frac{r}{\lambda},\ V(X)=\frac{r}{\lambda^2},\ M(t)=\left(\frac{\lambda}{\lambda-t}\right)^r$
Chi-cuadrado	$f_X(x)=\dfrac{e^{-\frac{x}{2}}x^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}I_{(0,\infty)}(x),\ E(X)=n,\ V(X)=2n$ $M(t)=\left(\frac{1}{1-2t}\right)^{\frac{n}{2}}$
Weibull(a,b)	$f_X(x)=abx^{b-1}e^{-ax^b},\ x>0,\ a>0,\ b>0\ ,$ $F(x)=1-e^{-ax^b},\ x>0,\ \mu'_k=\dfrac{\Gamma(1+\frac{k}{b})}{a^{k/b}},$ $E(X)=\frac{\Gamma(1+\frac{1}{b})}{a^{1/b}},\ V(X)=\frac{\Gamma(1+\frac{2}{b})-\Gamma(1+\frac{1}{b})}{a^{2/b}}$
Beta(a,b)	$f_X(x)=\frac{1}{B(a,b)}x^{n-1}(1-x)^{b-1},\ 0<x<1,a>0,\ b>0$ $E(X)=\frac{a}{a+b},\ V(X)=\frac{ab}{a+b+1}$
Lognormal(μ,σ^2)	$f_X(x)=\frac{1}{x\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left[\frac{lnx-\mu}{\sigma}\right]^2},x>0$ $\mu'_k=e^{k\mu+\frac{1}{2}k^2\sigma^2};E(X)=e^{\mu+\frac{\sigma^2}{2}},\ V(X)=e^{2\mu+\sigma^2}[e^{\sigma^2}-1]$
t(de Student)	$f_X(x)=\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}}\left(\frac{1+x^2}{n}\right)^{-(n+1)/2},\ -\infty<x<\infty$ $E(X)=0,\ n>1$ y $V(X)=\frac{n}{n-2},\ n>2$
F(de Snedecor)	$f_X(x)=\frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}\left(\frac{m}{n}\right)^{\frac{m}{2}}\left[\frac{x^{\frac{m}{2}-1}}{[1+\frac{m}{n}x]^{(m+n)/2}}\right],\ x\geq 0$ $E(X)=\frac{n}{n-2},\ n>2$ y $V(X)=\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)},\ n>4$

Apéndice D

Tablas

El objetivo esta en presentar las tablas de las distribuciones normal, chi-cuadrado, t de Student y F de Snedecor.

La tabla D.1 da los valores de $\Phi = P(Z \leq z)$,

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

donde $z = 0$ a $z = 3.49$ con paso de 0.1. Para obtener los valores de Φ para z negativo, se usa la relación $\Phi(-z) = 1 - \Phi(z)$. $\star$

La tabla D.2 da los valores t_n de la función de distribución de

$$F(t) = \int_{-\infty}^{t} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}\left(1 + \frac{x^2}{n}\right)^{(n+1)/2}} dx = \alpha$$

Con $n = 1, 2, \cdots, 30, 40, 60, 120, \infty$ y para $\alpha = 0.750$ 0.900 0.950 0.975 0.990 0.995 y 0.995. $\star$

La tabla D.3 da los valores χ^2 de la función de distribución de la chi-cuadrado,

$$F(x) = \int_{0}^{x} \frac{t^{(n-2)/2} e^{-t/2}}{2^{n/2}\Gamma(n/2)} dt = \alpha$$

con $n = 1, 2, \cdots, 30$ y para $\alpha = 0.005$ 0.010 0.025 0.050 0.100 0.0250 0.500 0.750 0.900 0.950 0.975 0.990 y 0.995. $\star$

La tabla D.4 da los valores f de la función de distribución de la F con m y n grados de libertad

$$F(y) = \int_0^y \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} \left[\frac{x^{\frac{m}{2}-1}}{[1+\frac{m}{n}x]^{(m+n)/2}}\right] dx$$

Para $m = n = 1, 2, \cdots, 10, 12, 15, 20, 30, 60, 120, \infty$ y para $\alpha = 0.900\ 0.950\ 0.975\ 0.990$ y 0.0995.

Para valores de $\alpha = 0.100\ 0.050\ 0.025\ 0.010$ y 0.005, se emplea la relación $F_{1-\alpha}(m, n) = \frac{1}{F_\alpha}(n, m)$. $\star$

Tabla D.1: VALORES DE $\Phi(x)$ y x DE LA DISTRIBUCIÓN NORMAL

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Tabla D.2: VALORES $t_n(\alpha)$ DE LA DISTRIBUCIÓN t

$$F(t) = \int_{-\infty}^{t} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}\left(1+\frac{x^2}{n}\right)^{(n+1)/2}} dx$$

F / n	.75	.90	.95	.975	.99	.995	.9995
1	1.000	3.078	6.314	12.706	31.821	63.657	636.619
2	.816	1.886	2.920	4.303	6.965	9.925	31.599
3	.765	1.638	2.353	3.182	4.541	5.841	12.924
4	.741	1.533	2.132	2.776	3.747	4.604	8.610
5	.727	1.476	2.015	2.571	3.365	4.032	6.869
6	.718	1.440	1.943	2.447	3.143	3.707	5.959
7	.711	1.415	1.895	2.365	2.998	3.499	5.408
8	.706	1.397	1.860	2.306	2.896	3.355	5.041
9	.703	1.383	1.833	2.262	2.821	3.250	4.781
10	.700	1.372	1.812	2.228	2.764	3.169	4.587
11	.697	1.363	1.796	2.201	2.718	3.106	4.437
12	.695	1.356	1.782	2.179	2.681	3.055	4.318
13	.694	1.350	1.771	2.160	2.650	3.012	4.221
14	.692	1.345	1.761	2.145	2.624	2.977	4.140
15	.691	1.341	1.753	2.131	2.602	2.947	4.073
16	.690	1.337	1.746	2.120	2.583	2.921	4.015
17	.689	1.333	1.740	2.110	2.567	2.898	3.965
18	.688	1.330	1.734	2.101	2.552	2.878	3.922
19	.688	1.328	1.729	2.093	2.539	2.861	3.883
20	.687	1.325	1.725	2.086	2.528	2.845	3.850
21	.686	1.323	1.721	2.080	2.518	2.831	3.819
22	.686	1.321	1.717	2.074	2.508	2.819	3.792
23	.685	1.319	1.714	2.069	2.500	2.807	3.768
24	.685	1.318	1.711	2.064	2.492	2.797	3.745
25	.684	1.316	1.708	2.060	2.485	2.787	3.725
26	.684	1.315	1.706	2.056	2.479	2.779	3.707
27	.684	1.314	1.703	2.052	2.473	2.771	3.690
28	.683	1.313	1.701	2.048	2.467	2.763	3.674
29	.683	1.311	1.699	2.045	2.462	2.756	3.659
30	.683	1.310	1.697	2.042	2.457	2.750	3.646
40	.681	1.303	1.684	2.021	2.423	2.704	3.551
60	.679	1.296	1.671	2.000	2.390	2.660	3.460
120	.677	1.289	1.658	1.980	2.358	2.617	3.373
∞	.674	1.282	1.645	1.960	2.326	2.576	3.291

Tabla D.3: VALORES $\chi^2_n(\alpha)$ DE LA DISTRIBUCIÓN CHI-CUADRADO

$$F(x) = \int_0^x \frac{t^{(n-2)/2} e^{-t/2}}{2^{n/2}\Gamma(n/2)} dt$$

n \ F	.005	.010	.025	.050	.100	.250	.500	.750	.900	.950	.975	.990	.995
1	$.0^4393$	$.0^3157$	$.0^3982$	$.0^2393$	.0158	.102	.455	1.32	2.71	3.84	5.02	6.63	7.88
2	.0100	.0201	.0506	.103	.211	.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	.0717	.115	.216	.352	.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	.207	.297	.484	.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	.412	.554	.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	.676	.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7

Tabla D.4: VALORES $F_{n,m}(\alpha)$ DE LA DISTRIBUCIÓN F

$$F(y) = \int_0^y \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}\left(\frac{m}{n}\right)^{\frac{m}{2}}\left[\frac{x^{\frac{m}{2}-1}}{[1+\frac{m}{n}x]^{(m+n)/2}}\right]dx$$

F	n \ m	1	2	3	4	5	6	7	8	9	10	12	15	20	30	60	120	∞
.9		39.9	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9	60.2	60.7	61.2	61.7	62.3	62.8	63.1	63.3
.95		161	200	216	225	230	234	237	239	241	242	244	246	248	250	252	253	254
.975	1	648	800	864	900	922	937	948	957	963	969	977	985	993	1,000	1,010	1,010	1,020
.99		4,050	5,000	5,400	5,620	5,760	5,860	5,930	5,980	6,020	6,060	6,110	6,160	6,210	6,260	6,310	6,340	6,370
.995		16,200	20,000	21,600	22,500	23,100	23,400	23,700	23,900	24,100	24,200	24,400	24,600	24,800	25,000	25,300	25,400	25,500
.9		8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.46	9.47	9.48	9.49
.95		18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5
.975	2	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.4	39.4	39.5	39.5	39.5	39.5
.99		98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5
.995		199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199
.9		5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.17	5.15	5.14	5.13
.95		10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.62	8.57	8.55	8.53
.975	3	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.3	14.3	14.2	14.1	14.0	13.9	13.9
.99		34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.5	26.3	26.2	26.1
.995		55.6	49.8	47.5	46.2	45.4	44.8	44.4	44.1	43.9	43.7	43.4	43.1	42.8	42.5	42.1	42.0	41.8
.9		4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.82	3.79	3.78	3.76
.95		7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.75	5.69	5.66	5.63
.975	4	12.2	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.46	8.36	8.31	8.26
.99		21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.8	13.7	13.6	13.5
.995		31.3	26.3	24.3	23.2	22.5	22.0	21.6	21.4	21.1	21.0	20.7	20.4	20.2	19.9	19.6	19.5	19.3
.9		4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.17	3.14	3.12	3.10
.95		6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.50	4.43	4.40	4.37
.975	5	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.23	6.12	6.07	6.02
.99		16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.38	9.20	9.11	9.02
.995		22.8	18.3	16.5	15.6	14.9	14.5	14.2	14.0	13.8	13.6	13.4	13.1	12.9	12.7	12.4	12.3	12.1
.9		3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.80	2.76	2.74	2.72
.95		5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.81	3.74	3.70	3.67
.975	6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.07	4.96	4.90	4.85
.99		13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.23	7.06	6.97	6.88
.995		18.6	14.5	12.9	12.0	11.5	11.1	10.8	10.6	10.4	10.3	10.0	9.81	9.59	9.36	9.12	9.00	8.88
.9		3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.56	2.51	2.49	2.47
.95		5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.38	3.30	3.27	3.23
.975	7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.36	4.25	4.20	4.14
.99		12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	5.99	5.82	5.74	5.65
.995		16.2	12.4	10.9	10.1	9.52	9.16	8.89	8.68	8.51	8.38	8.18	7.97	7.75	7.53	7.31	7.19	7.08
.9		3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.38	2.34	2.32	2.29
.95		5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.08	3.01	2.97	2.93
.975	8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.89	3.78	3.73	3.67
.99		11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.20	5.03	4.95	4.86
.995		14.7	11.0	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.01	6.81	6.61	6.40	6.18	6.06	5.95

F	n \ m	1	2	3	4	5	6	7	8	9	10	12	15	20	30	60	120	∞
.9		3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.25	2.21	2.18	2.16
.95		5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.86	2.79	2.75	2.71
.975	9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.56	3.45	3.39	3.33
.99		10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.65	4.48	4.40	4.31
.995		13.6	10.1	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.23	6.03	5.83	5.62	5.41	5.30	5.19
.9		3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.16	2.11	2.08	2.06
.95		4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.70	2.62	2.58	2.54
.975	10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.31	3.20	3.14	3.08
.99		10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.25	4.08	4.00	3.91
.995		12.8	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.66	5.47	5.27	5.07	4.86	4.75	4.64
.9		3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.01	1.96	1.93	1.90
.95		4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.47	2.38	2.34	2.30
.975	12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	2.96	2.85	2.79	2.72
.99		9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.70	3.54	3.45	3.36
.995		11.8	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.91	4.72	4.53	4.33	4.12	4.01	3.90
.9		3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.87	1.82	1.79	1.76
.95		4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.25	2.16	2.11	2.07
.975	15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.64	2.52	2.46	2.40
.99		8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.21	3.05	2.96	2.87
.995		1.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.25	4.07	3.88	3.69	3.48	3.37	3.26
.9		2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.74	1.68	1.64	1.61
.95		4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.04	1.95	1.90	1.84
.975	20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.35	2.22	2.16	2.09
.99		8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.78	2.61	2.52	2.42
.995		9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.68	3.50	3.32	3.12	2.92	2.81	2.69
.9		2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.61	1.54	1.50	1.46
.95		4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.84	1.74	1.68	1.62
.975	30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.07	1.94	1.87	1.79
.99		7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.39	2.21	2.11	2.01
.995		9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.18	3.01	2.82	2.63	2.42	2.30	2.18
.9		2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.48	1.40	1.35	1.29
.95		4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.65	1.53	1.47	1.39
.975	60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.82	1.67	1.58	1.48
.99		7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.03	1.84	1.73	1.60
.995		8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.74	2.57	2.39	2.19	1.96	1.83	1.69
.9		2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.41	1.32	1.26	1.19
.95		3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.55	1.43	1.35	1.25
.975	120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.69	1.53	1.43	1.31
.99		6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.86	1.66	1.53	1.38
.995		8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.54	2.37	2.19	1.98	1.75	1.61	1.43
.9		2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.34	1.24	1.17	1.00
.95		3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.46	1.32	1.22	1.00
.975	∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.57	1.39	1.27	1.00
.99		6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.70	1.47	1.32	1.00
.995		7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.36	2.19	2.00	1.79	1.53	1.36	1.00

Bibliografía

[1] BAIN, L. y ENGELHARDT, M. *Introduction to probability and mathematical statistical.* Duxbury press. Belmont, CA. 1992.

[2] BHATTACHARYYA, G. y JOHNSON, T. *Statistical concepts and methods.* John Wiley. New York. 1977.

[3] BLANCO, L.*Probabilidad.*Universidad Nacional de Colombia. Bogotá 2004

[4] BURINTON, G y MAY, D. *Handbook of probability and statistics with tables.* McGraw-Hill. New York. 1970.

[5] BURRIL, C. *Measure, integration and probability.* McGraw-Hill. New York. 1972.

[6] CASELLA, G. y BERGER, R. *Statistical inference.* Duxbury, Thompson Learning. Pacif Grove, CA. 2002

[7] DEKKING, F.M; KRAAIKAMP, C; LOPUAÄ, H.P y MEESTER, L.E. *A modern introduction to probability and statistics.* Springer. London. 2005

[8] GHAHRAMANI, S. *Fundamentals of probability.* Prentice Hall. New Jersey. 2001

[9] GREENWOOD, J. A. y DURAN, D. *Aids for fitting the gamma distribution by maximum likelihood.* Technometrics Vol. 2, No. 1, Pag 55-64. 1960

[10] HARNETT, D y MURPHY, J. *Introductory statistical analysis.* Addison-Weasley. Massachusetts. 1975.

[11] HOGG, R y CRAIG, A. *Indtroduction to mathematical statistics.* Collier McMillan. New Jersey. 1995.

[12] JACOB, J Y PROTTER, P. *Probability essentials.* Springer. Berlín. 2004.

[13] JOHNSON, N; KOTZ, S y BALAKRISHNAN, N. *Continous univariate distributions. Vol.1.* John Wiley & Sons.inc. New York. 1995

[14] JOHNSON, N; KOTZ, S y KEMP, A. *Univariate discrete distributions.* John Wiley & Sons.inc. New York. 1992.

[15] KOOP, G; POIRIER, D y TOBIAS, J. *Bayesian econometric methods.* Cambridge University press. New York. 2007.

[16] KREYSZIG, E. *Introducción a la estadística matemática: Principios y métodos.* Limusa. México. 1979.

[17] LOVE, M. *Probability theory I.* Springer-Verlag. New York. 1977

[18] MAYORGA, J. *Inferencia estadística.* Universidad Nacional. Bogotá. 2004

[19] MENDENHALL, W; SCHEAFFER, R y WACKERLY, D. *Mathematical statistics with applications.* Duxbury Press. Belmont, CA. 1981.

[20] MOOD, A; GRAYBILL, F y BOES, D. *Introduction to the theory of statistics.* International Student. McGraw Hill. Singapore. 1974

[21] PETROV, V y MORDECKI, E. *Teoría de probabilidades.* Matematnka. Moscú. 2003

[22] ROSS, S.*A First Course in Probability.* Prentice Hall. New Jersey. 1994

[23] ROSS, S. *Introduction to probability models.* Academic Press. New York. 2007

[24] ROSSENTHAL, J. *A first look at rigorous probabiliity theory.* World Scienttific. Singapore. 2008

[25] SHAO, J. *Mathematical statistics.* Springer. 1999

[26] WASSERMAN, L. *All of statistics.* Springer. Pittsburgh. 2004

Índice alfabético

La impresión de este libro se realizó en papel bond blanco 90 grs. para páginas interiores y propalcote de 280 grs. para la portada con plastificado mate. Con un tiraje de 200 ejemplares. *El libro PROBABILIDAD Y ESTADÍSTICA una introducción* del Autor: César A. Herazo Henriquez hace parte de la primera convocatoria para la publicación de libros de la colección de investigación Francisco José de Caldas de la Universidad de Cartagena. El diseño y diagramación se realizó en la Editorial Universitaria - Sección de Publicaciones de la Universidad de Cartagena y se terminó de imprimir en el año 2014 en la empresa Alpha Impresores, en la ciudad de Cartagena de Indias, Colombia.