Universidad Nacional Autónoma de México

Facultad de Ciencias

Notas de estadística descriptiva e inferencial para la carrera de Actuaría en línea

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Actuaría

PRESENTA:

Castellanos Rodríguez Sara

TUTOR

M. en I. María Isabel Escalante Membrillo

Hoja de Datos del Jurado

Datos del alumno
Castellanos
Rodríguez
Sara
Tel.: 55-3276-1256
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
305280466

Datos del tutor
M. en I.
María Isabel
Escalante
Membrillo

Datos del sinodal 1
Dra.
María del Pilar
Alonso
Reyes

Datos del sinodal 2
M. en C.
José Antonio
Flores
Díaz

Datos del sinodal 3
Act.
Francisco
Sánchez
Villarreal

Datos del sinodal 4
Act.
Alma Nayeli
Santos
Coria

Datos del trabajo escrito.
Notas de estadística descriptiva e inferencial para la carrera de actuaría en línea
147
2018

Índice general

Prólogo

La obtención y el manejo de datos no es algo nuevo, se sabe que desde los griegos ya se realizaban censos con fines tributarios, sociales y militares. Al principio sólo se recolectaban datos para tener un registro y descripción de éstos. No fue sino hasta mediados del siglo XVII con los trabajos de C. Huygens, B. Pascal, E. Halley, G. Neumann entre otros que esta información se empieza a utilizar con fines deductivos, por ejemplo Halley presenta la primera tabla de mortalidad para tratar de establecer el precio de las anualidades de las compañías de seguros. Y fue hasta el siglo XIX que la estadística entra en una nueva fase de desarrollo con los trabajos de Galton y Pearson pues a ellos se debe el paso de la estadística deductiva a la inductiva con el estudio de la relación entre variables, el primero desarrolla el método de correlación y el segundo el coeficiente de correlación.

El desarrollo de la teoría de probabilidades se da a principios del siglo XVIII, sin embargo es hasta comienzos del siguiente siglo cuando comienzan a asentar las bases de dicho teoría y es ahí donde se puede situar el inicio de la estadística moderna con R. Arnold Fisher su máximo representante ya que logro situar a la estadística como una herramienta para el análisis de experimentos además fue pionero en la introducción del método de máxima verosimilitud.

En la actualidad la estadística no sólo se trata de recolectar datos y generar tablas si no también de relacionar los datos, interpretar resultados y generar conclusiones con fundamentos matemáticos. Este conjunto de herramientas es usados en muchas áreas para el análisis de fenómenos sociales, económicos, biológicos y físicos. Con esto la importancia de esta rama y su estudio salen a flote.

El presente texto va dirigido principalmente a estudiantes de la carrera de actuaría, tanto de sistema abierto como escolarizado y está concebido para que el alumno desarrolle los temas de la primera mitad de un curso básico de estadística para licenciatura: Estadística descriptiva e Inferencial.

Debido al auge tecnológico en los últimos años la facilidad de acceso a la información en línea, que,entre otras cosas, ha permitido la generación de material didáctico electrónico útil al estudiante. Es por lo que se pensó en la creación de este texto, digital y abierto, como complemento a un curso escolarizado o bien para quien desee instruirse por sus propios medios.

Un prerrequisito a estás notas es un curso de cálculo diferencial e integral que incluya derivadas parciales e integral multiple así como conocimiento previo de probabilidad: conceptos como variable aleatoria, función de densidad y distribución,esperanza, varianza, etc, serán de gran ayuda para la comprensión de los temas.

Este trabajo se divide en dos capitulos: En el capítulo 1 se desarrolla la parte de estadística descriptiva, empezando por tipos de datos, seguido de su representación gráfica, para finalizar con la presentación de medidas de tendencia central, posición, dispersión y forma. El capítulo 2 corresponde a una parte de la inferencia estadística: comienza con estimación puntual, en particular por el método de momentos y de máxima verosimilitud. Después se estudian algunas propiedades importantes deseables para estos estimadores. Al final de ambos capítulos se encuentra una sección de ejercicios propuestos para complementar el desarrollo de ideas del alumno.

Estadística descriptiva.

En estadística descriptiva se muestran técnicas para el análisis datos él cual consiste en clasificarlos, ordenarlos y representarlos a través de gráficas y/o medidas, una de ellas ya bastante conocida: el promedio.

1.1 Datos, variables y escalas de medición

Al realizar cualquier tipo de experimento, medición u observación se obtienen datos los cuales se pueden ver como el rango o imagen de una función que en este contexto la llamaremos *variable*. Y al conjunto de valores que se obtuvieron del experimento u observación se le llamara *muestra*.

Una primera clasificación para el tipo de datos es la siguiente:

Se habla de una variable cuantitativa cuando se realiza una medición y el resultado es un número con el que se pueden realizar operaciones aritméticas y se entiende por variable cualitativa (o atributo) cuando solamente registran una cualidad del objeto o persona en estudio. Dado un colectivo de personas su edad y estatura son ejemplo de variables cuantitativas mientras que la afiliación política y el estado civil son atributos.

Dentro del conjunto de variables cuantitativas se distingue entre *discretas* y *continuas*. Se llama ***variable discreta*** cuando sólo puede tomar un número finito o numerable de valores y ***variable continua*** cuando la cantidad de valores que puede tomar es no numerable, es decir, si toma cualquier valor dentro de un intervalo (a,b) de la recta real. La estatura de un grupo de personas sería ejemplo de una variable continua mientras que el número de hijos por mujer sería de una variable discreta.

Aunque en la práctica todas las variables cuantitativas son discretas, debido a la limitación de los aparatos de medición, en general se trata a las variables que en teoría son continuas como tales, por razones que sólo se alcanzan a ver en el segundo capítulo.

Una tercera clasificación está en función de la relación que guardan entre si los datos.

Las variables cualitativas pueden ser clasificadas de acuerdo a dos escalas: escala nominal y ordinal. La primera unicamente pone nombre a una característica y no es posible crear alguna relación de orden o magnitud entre los posibles valores de la variable, el lugar de nacimiento de las personas o el grupo sanguíneo son un ejemplo de éstas. Las variables ordinales llevan asociado un orden entre sus respuestas y aunque se pueden emplear números para su clasificación no es posible hacer operaciones aritméticas entre estos valores, por ejemplo: para calificar el grado de satisfacción de un servicio se pueden crear la siguiente clasificación: 0 = Muy malo, 1 = malo, 2 = Regular, 3 = Bueno, 4 = Excelente. En este caso, es claro que existe un orden entre sus valores, pero al realizar operaciones aritméticas entre los valores no se genera un atributo de la variable. Dado lo anterior, no se puede decir que un valor malo y uno bueno hacen un valor excelente.

Las variables cuantitativas pueden clasificarse en escala intervalar y de razón. En la primera hay un orden entre los valores de la variable y se pueden realizar con ellos las operaciones de suma y resta lo que genera una noción de distancia. En esta escala el valor cero no es absoluto pues no refleja ausencia de la magnitud medida. La temperatura es una variable de esta clase aquí el posible valor cero depende de la escala que se use para medir la temperatura (Celsius, Kelvin, Fahrenheit). La segunda tiene las mismas propiedades que la escala de intervalo y añade a estas características la de incorporar un origen no arbitrario, de esta forma el cero se entiende como inexistencia. También es posible realizar cualquier tipo de operación lógica y aritmética y expresiones del tipo ‶una observación es el doble de la otra″tienen sentido. Por ejemplo, la variable edad estudiada en una población humana.

Es importante tener en cuenta el tipo de variable con la cual se trabaja así como el tipo de escala ya que de este dependerá el tipo de procedimiento estadístico que se llevará acabo.

1.2 Tablas de frecuencias

La organización de los datos constituye parte importante de su análisis ya que facilita cálculos posteriores y aunque actualmente, con el desarrollo de programas estadísticos, en la práctica éstas herramientas han perdido relevancia, siguen siendo importantes en la enseñanza por su gran valor conceptual.

La organización va a depender de la cantidad de datos que se tenga.

Caso 1) Cuando se tiene un número pequeño de observaciones, en su mayoría distintas es conveniente darlos por extensión.

Ejemplo **1.2.1.** Al lanzar un dado 7 veces se obtienes y observar el número obtenido, se obtienen las siguientes observaciones:

$$1, 1, 5, 6, 3, 4, 5$$

Caso 2) Cuando se tiene un gran número de observaciones pero muy pocas distintas se organiza en una tabla de frecuencias,la cual estará formada por cada uno de los valores distintos acompañados por el número de veces que se repite.

Ejemplo **1.2.2.** En una votación para elegir a un presidente de entre cuatro candidatos se han obtenido los siguientes resultados

Candidato	Número de votos
A	287
B	450
C	230
D	315

Caso 3) Si se tienen muchas observaciones y en su mayoría distintas, pueden disponerse agrupándolas en intervalos e indicando el número de datos que caen dentro de cada intervalo.

Ejemplo **1.2.3.** La siguiente tabla muestra los salarios diarios de 65 empleados de la mueblería Otlamex.

Salarios	Número de empleados
[250-260)	8
[260-270)	10
[270-280)	16
[280-290)	14
[290-300)	10
[300-310)	5
[310-320)	2

Al trabajar con variables cuantitativas o cualitativas de tipo ordinal y si la cantidad de datos es muy grande (caso 2 y 3), se suele trabajar con un arreglo de datos basado en intervalos conocido como tabla de frecuencias. No sólo ayuda a organizar los valores si no que es de gran ayuda para construir algunas de las representaciones gráficas que se verán más adelante.

Para la elaboración de tablas de frecuencia hay que seguir los siguientes pasos:

Paso 1) Los datos de la muestra deben de ordenarse y aunque es usual ordenar de menor a mayor no representa ningún problema efectuarlo al revés.

Paso 2) Se debe definir el número de clases o intervalos $\boldsymbol{K}$ en las que se agrupan los datos de la muestra. La raíz cuadrada del total de datos, $K = \sqrt{n}$, ó la regla de Sturges, $K = 1 + log_2(n)$, son los más usados para este fin.

Paso 3) Determinar la longitud del intervalo de clase (c), el cual se aproxima mediante la fórmula:

$$c = \frac{\text{valor máximo-valor mínimo}}{\text{número de clases}}.$$

A partir de este paso se generan los **límites de clase** que son los valores mayor (L_{i+1}) y menor (L_i) de cada intervalo. Éstos tienen la misma aproximación que los datos en la muestra o de la población y se denotan por L_i también las **marcas de clase** que son los puntos medios de los intervalos y son los valores que se usarán para representar a cada clase y se denotar por m_i.

Paso 4) Calcular los distintos tipos de frecuencia:

a) **Frecuencia absoluta observada**. Es el número de elementos en la muestra o en la población que pertenece a la clase en cuestión. Se denota por f_i.

b) **Frecuencia absoluta acumulada**. Es el número de datos en la muestra o población, que son menores o iguales que el límite superior del intervalo en cuestión. Se denota por F_i, y se obtiene sumando la frecuencia del intervalo actual con las frecuencias de los intervalos anteriores.

c) **Frecuencia relativa observada**. Es la proporción de datos que pertenecen a la clase en cuestión. Generalmente se denota por f_i' o f_i^*. Se obtiene realizando el cociente entre la frecuencia y el total de datos, es decir, $f_i^* = \dfrac{f_i}{n}$.

d) **Frecuencia relativa acumulada**. Es la proporción de los datos que son menores o iguales que el límite superior de la clase con la cual se está trabajando. Se denota por F_i' o F_i^* y se calcula como el cociente de la frecuencia acumulada y el número de datos: $F_i^* = \dfrac{F_i}{n}$.

Ejemplo **1.2.4.** Las calificaciones finales en Biologia de 80 estudiantes se observan en la siguiente tabla

68	84	75	82	68	90	62	88	76	93
73	79	88	73	60	93	71	59	85	75
61	65	75	87	74	62	95	78	63	72
66	78	82	75	94	77	69	74	68	60
96	78	89	61	75	95	60	79	83	71
79	62	67	97	78	85	76	65	71	75
65	80	73	57	88	78	62	76	53	74
86	67	73	81	72	63	76	75	85	77

Hallar la tabla de frecuencias asociada al conjunto de datos.

Paso 1) Los datos ordenados de mayor a menor quedan:

53	57	59	60	60	60	61	61	62	62
62	62	63	63	65	65	65	66	67	67
68	68	68	69	71	71	71	72	72	73
73	73	73	74	74	74	75	75	75	75
75	75	75	76	76	76	76	77	77	78
78	78	78	78	79	79	79	80	81	82
82	83	84	85	85	85	86	87	88	88
88	89	90	93	93	94	95	95	96	97

Paso 2) El número de clases que considerado para este ejemplo se obtiene con la raíz cuadrada del número de datos: $K = \sqrt{80} \approx 9$.

Paso 3) La longitud de las clases es $c = \dfrac{97-53}{9} \approx 4.88$, por lo que éstas quedan:

Clase 1	[53 - 57.88)
Clase 2	[57.88 -62.76)
Clase 3	[62.76-67.64)
Clase 4	[67.64 - 72.52)
Clase 5	[72.52-77.4)

Clase 6	[77.4-82.28)
Clase 7	[82.28-87.16)
Clase 8	[87.16-92.04)
Clase 9	[92.04-97)

Paso 4) Calcular los distintos tipos de frecuencia:

	Clase	Marca de clase	f_i	F_i	f_i^*	F_i^*
Clase 1	[53 - 57.88)	55.44	2	2	.025	.025
Clase 2	[57.88 -62.76)	60.32	10	12	.125	.015
Clase 3	[62.76-67.64)	65.2	8	20	.1	.25
Clase 4	[67.64 - 72.52)	70.08	9	29	.1125	.3625
Clase 5	[72.52-77.4)	74.96	20	49	.25	.6125
Clase 6	[77.4-82.28)	79.84	12	61	.15	.7625
Clase 7	[82.28-87.16)	84.72	7	68	.0875	.085
Clase 8	[87.16-92.04)	89.6	5	73	.0625	.9125
Clase 9	[92.04-97)	94.48	7	80	.0875	1

Ejemplo **1.2.5.** Los siguientes datos representan los pesos de 40 estudiantes hombres de una universidad. Construir su tabla de frecuencias con 7 y 10 intervalos.

138	164	150	132	144	125	149	157
146	158	140	147	136	148	152	144
168	126	138	176	163	199	154	165
146	173	142	147	135	153	140	135
161	145	135	142	150	156	145	128

Paso 1) Los datos ordenados de mayor a menor quedan:

125	126	128	132	135	135	135	136
138	138	140	140	142	142	144	144
145	145	146	146	147	147	148	149
150	150	152	153	154	156	157	158
161	163	164	165	168	173	176	199

Paso 2) El número de clases que considerado para este ejemplo será 7 y 10.

Con 7 clases:

Paso 3) La longitud de las clases es $c = \dfrac{199-125}{7} \approx 10.571$, por lo que éstas quedan:

Clase 1	[125 -135.5)
Clase 2	[135.5 -146.1)
Clase 3	[146.1 -156.7)
Clase 4	[156.7-167.2)

Clase 5	[167.2-177.8)
Clase 6	[177.8-188.4)
Clase7	[188.4-199)

Paso 4) Calcular los distintos tipos de frecuencia:

	Clase	Marca de clase	f_i	F_i	f_i^*	F_i^*
Clase 1	[125 -135.5)	130.2	7	7	.175	.025
Clase 2	[135.5 -146.1)	140.8	13	20	.325	.5
Clase 3	[146.1 -156.7)	151.4	10	30	.25	.75
Clase 4	[156.7-167.2)	161.9	6	36	.15	.9
Clase 5	[167.2-177.8)	172.5	3	39	.075	.975
Clase 6	[177.8-188.4)	183.1	0	39	.0	.975
Clase 7	[188.4-199)	193.7	1	40	.025	.1

Con 10 clases:

Paso 5) La longitud de las clases es $c = \dfrac{199-125}{10} \approx 7.4$, por lo que éstas quedan:

Clase 1	[125 -132.4)
Clase 2	[132.4-139.8)
Clase 3	[139.8-147.2)
Clase 4	[147.2-154.6)
Clase 5	[154.6-162)

Clase 6	[162-169.4)
Clase 7	[169.4-176.8)
Clase 8	[176.8-184.2)
Clase 9	[184.2-191.6)
Clase 10	[191.6-199)

Paso 6) Calcular los distintos tipos de frecuencia:

	Clase	Marca de clase	f_i	F_i	f_i^*	F_i^*
Clase 1	[125 -132.4)	128.7	4	4	.1	.1
Clase 2	[132.4-139.8)	136.1	6	10	.15	.25
Clase 3	[139.8-147.2)	143.5	12	22	.3	.55
Clase 4	[147.2-154.6)	150.9	7	29	.175	.725
Clase 5	[154.6-162)	158.3	4	33	.1	.825
Clase 6	[162-169.4)	165.7	4	37	.1	.925
Clase 7	[169.4-176.8)	173.1	2	39	.05	.975
Clase 8	[176.8-184.2)	180.5	0	39	0	.975
Clase 9	[184.2-191.6)	187.9	0	39	0	.975
Clase 10	[191.6-199)	195.3	1	40	.025	1

1.3 Descripción gráfica de los datos

El objetivo de los gráficos es revelar, visualmente, patrones de comportamiento de la variable de estudio. El tipo de representación utilizado no sólo estará en función del tipo de datos, si no también del concepto que se desee representar y/o conocer.

1. **Diagrama de pastel**: Ésta gráfica es común si lo que se busca es resaltar la distribución de frecuencias relativas. Éste muestra la cantidad de datos que pertenecen a cada categoría como una parte proporcional de un círculo y para obtener cada sector circular bastará multiplicar la frecuencia relativa de la categoría por 360 con lo cual se obtendrá el ángulo interior del sector.

Ejemplo **1.3.1.** Los datos del ejemplo 1.2.2 se pueden representar por medio de un diagrama de pastel.

Candidato	f_i	f_i^*	ángulo del sector circular
A	287	.2238	80.568
B	450	.3510	126.36
C	230	.1794	64.584
D	315	.2457	88.452

La Figura 1.1 muestra el gráfico solicitado, en el cual cada sector representa el porcentaje de población que voto por determinado candidato.

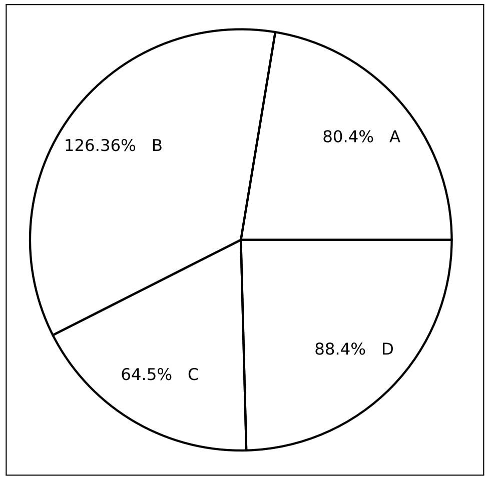


Figura 1.1: Gráfica de pastel

2. **Gráfica de barras**: Se utilizan para representar la frecuencia de las categorías de una variable cualitativa a través de rectángulos de áreas proporcionales a la frecuencia de cada atributo o categoría.

Ejemplo **1.3.2.** La siguiente figura muestra el diagrama de barras asociada al Ejemplo 1.2.2. Para construirlo se debe tener en cuenta que la base de los rectángulos son iguales y las alturas representan el número de personas que elegirá a determinado candidato.

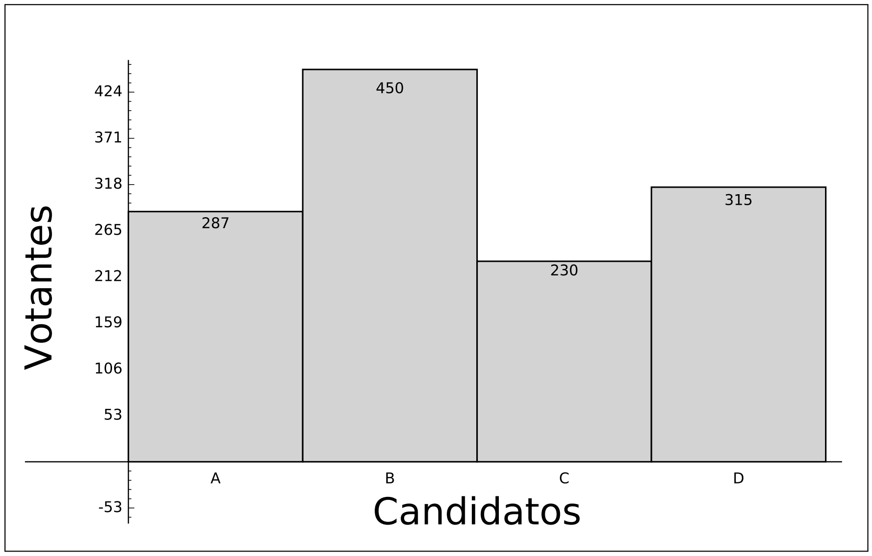


Figura 1.2: Gráfica de barras

Cuando el tamaño de la muestra es pequeña y la variable es cuantitativa u ordinal, el gráfico recomendado es:

3. **Diagrama de tallo y hojas**, el cual presenta las observaciones mediante el empleo de los dígitos que constituyen los valores de los datos. Cada uno se divide en dos partes: el(los) dígito(s) principal(es) se convierte(n) en el tallo y el(los) dígito(s) posterior(es) se convierte(n) en la hoja. Los tallos se escriben a lo largo del eje principal, y por cada dato se escribe una hoja para mostrar la distribución de éstos.

Ejemplo **1.3.3.** Construir un diagráma de tallo y hoja para el siguiente conjunto de calificaciones de 20 estudiantes.

82	74	88	66	58	74	78	84	96	76
62	68	72	92	86	76	52	76	82	78

Como todos los datos constan de 2 dígitos es conveniente utilizar el primer dígito como tallo y el segundo como hoja. Usualmente la representación se hace de manera vertical. Para ello se trazará una recta vertical y se colocaran todos los tallos distintos a la izquierda de ésta como se muestra a continuación.

$$\begin{array}{r|l} 5 & \\ 6 & \\ 7 & \\ 8 & \\ 9 & \end{array}$$

En seguida se coloca cada hoja junto a su tallo, esto se hace escribiendo el segundo dígito a lado derecho de la linea vertical. Las hojas no necesitan acomodarse con respecto a algún orden y pueden repetirse como se puede apreciar en la siguiente figura.

20 calificaciones de examen

5	8	2						
6	6	2	8					
7	4	4	8	6	2	6	6	8
8	2	8	4	6	2			
9	6	2						

Las principales gráficas que se apoyan en la tabla de frecuencias para su construcción son:

4. **Histograma** Permite la representación de la frecuencia de una variable a través de barras rectangulares cuyas bases están centradas en el punto medio de cada intervalo y las áreas de dichos rectángulos son proporcionales a la frecuencia (absoluta o relativa) de cada clase. El eje de las abscisas se refiere a los valores de la característica en estudio y el eje de las ordenas a la frecuencia observada para cada una de ellas.

Ejemplo **1.3.4.** El histograma para el Ejemplo1.2.4 está dado en las Figuras 1.3. El eje vertical representa la frecuencia absoluta de cada intervalo y el eje horizontal marca el limite superior e inferior de las clases. Al ser las bases de los rectángulos iguales las áreas son proporcionales a la frecuencia.

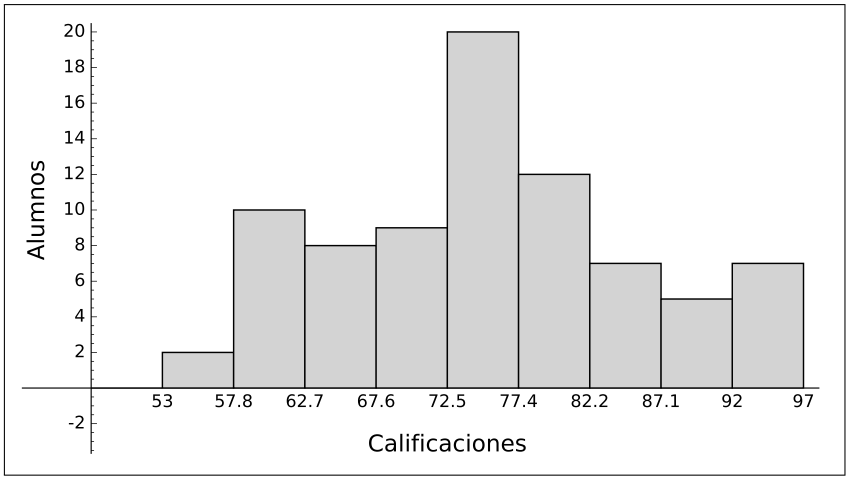
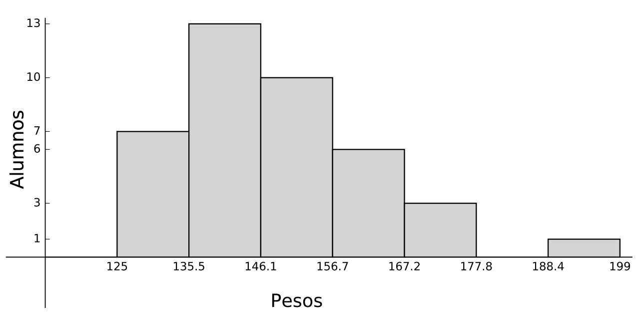
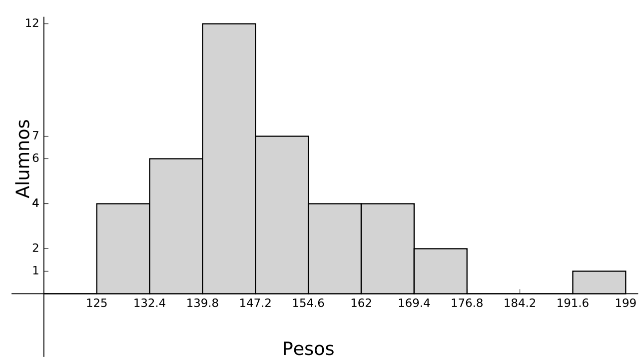


Figura 1.3: Calificaciones

Ejemplo **1.3.5.** En el Ejemplo1.2.5 se construyeron dos distribuciones de frecuencias, con 7 y 10 intervalos de clase, los histogramas asociados se pueden observar en las Figuras respectivamente. En esta ocasión se opto por representa a cada intervalo por sus límites de clase(eje horizontal).



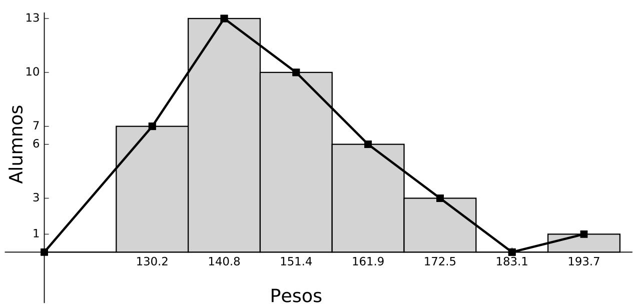
(a)



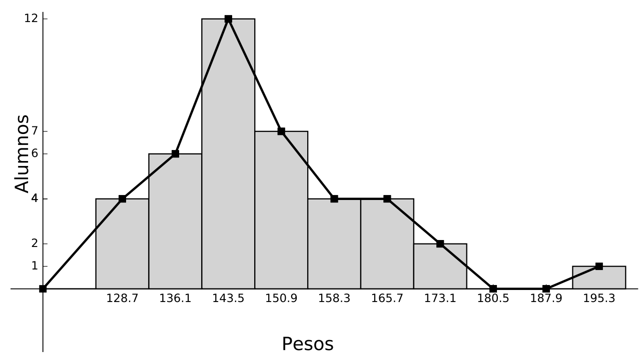
(b)

5. **Polígono de frecuencias**: Es una gráfica de líneas rectas que se realiza a partiendo del histograma, se forma a partir por la unión de los puntos cuya ordenada es la frecuencia respectiva y la abscisa es la marca de clase, es opcional dibujar una marca de clase antes de la primera y una posterior a la última , cada una de frecuencia cero.

Ejemplo **1.3.6.** Obtener el polígono de frecuencias asociado a los dos histogramas del ejemplo 1.3.5



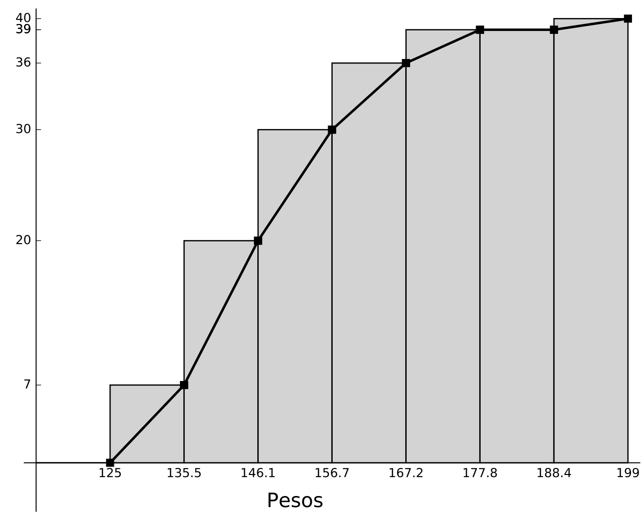
(a)



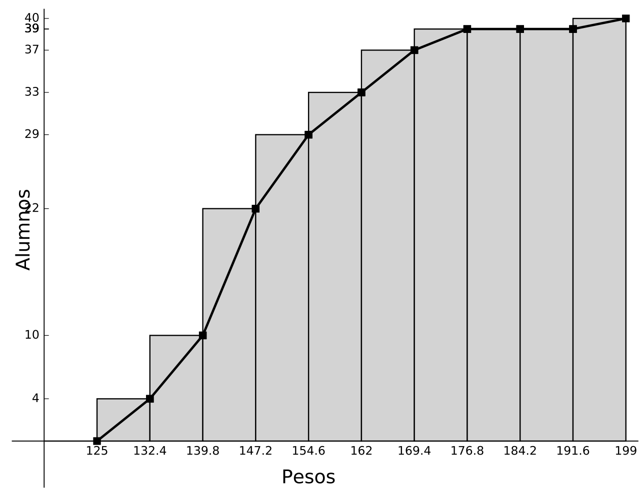
(b)

6. La **ojiva** o polígono de frecuencias acumuladas se dibuja utilizando las fronteras de los intervalos de clase contra las frecuencias acumuladas absolutas o relativas. La ojiva indica, para cada frontera, los elementos (o proporción de elementos) que son menores o iguales que esa frontera. Si se utiliza la frecuencia acumulada relativa se llama ojiva porcentual.

Ejemplo **1.3.7.** Obtener las gráficas de ojiva relacionadas a los histogramas del ejemplo 1.3.5



(a)



(b)

1.4 Descripción numérica de los datos.

Para las variables cuantitativas una vez organizados los datos se procede a obtener una serie de cálculos que resuman y representen a la muestra. Las más importantes se clasifican en: medidas de tendencia central, de dispersión y de forma.

En el apartado anterior ya se menciono que si se tiene una cantidad muy grande de observaciones se suele trabajar con intervalos de clase, si los valores se encuentran de esta manera se dice que son *datos agrupados*. En esta parte del texto se trabajará con los datos de manera extensiva y agrupada para cada una de las medidas.

A partir de ahora se denotara por $x_1, x_2, \ldots, x_n$ las n observaciones que se obtuvieron del experimento en cuestión.

Medidas de tendencia central.

Sea $x_1, x_2, ..., x_n$ una muestra. De entre todas las características numéricas que tiene este conjunto de datos algunas de las más importantes son las de tendencia central, las cuales se resumen por medio de un sólo número, el cual está determinado por el cálculo de las medidas que se definen en esta sección.

Definición **1.4.1.** La media muestral o promedio, $\bar{x}$, de la muestra $x_1, x_2, ..., x_n$ se define como:

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}. \tag{1.1}$$

Si los datos están ordenados en intervalos de clase (datos agrupados) la *media* es

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i \cdot m_i}{n} \tag{1.2}$$

donde f_i es la frecuencia absoluta; m_i es la marca de clase y n es el número de datos.

Ejemplo **1.4.1.** Los salarios anuales de 4 individuos son $\$15,000$, $\$16,000$, $\$16,500$ y $\$40,000$.

a) Hallar su media aritmética.

Usando la Fórmula 1.1, la media de los datos es

$$\bar{x} = \frac{\$15,000 + \$16,000 + \$16,500 + \$40,000}{4} = \frac{\$87,500}{4} = \$21,875$$

b) ¿Podría decirse que la media ‶representa″a la muestra?.

No, pues 4 de 5 trabajadores están muy debajo de ese valor. Este valor en el promedio se debe a que tenemos un valor atípico muy grande comparado con el resto de las observaciones.

Ejemplo **1.4.2.** En la siguiente tabla se muestran la distribución absoluta de las alturas de 100 estudiantes varones de la Facultad de ciencias.

Altura	Marca de clase (m_i)	Frecuencia absoluta (f_i)
$60-62$	61	5
$63-65$	64	18
$66-68$	67	42
$69-71$	70	27
$72-74$	73	8
Total		100

Halar la altura promedio para esta muestra.

Solución:

Usando la Fórmula 1.2 se obtiene

$$\bar{x} = \frac{61 \cdot 5 + 64 \cdot 18 + 67 \cdot 42 + 70 \cdot 27 + 73 \cdot 8}{100} = \frac{6745}{100} = 67.45$$

En cierto sentido la media representa el centro de gravedad de los datos observados o centro geométrico del conjunto de valores ya que equilibra las desviaciones positivas y negativas respecto a su valor

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

La media es la mejor de las medidas de representación pues toma en cuenta todos los elemento de la muestra aunque esto también la vuelve sensible a la presencia de valores atípicos, lo cual provoca que en existencia de datos extremos no es la mejor opción.

La siguiente cantidad tiene la ventaja sobre la media que, en muestras con datos extremos, representa mejor a la muestra , pues sólo depende de la posición de los datos y es poco sensible a estos valores; además es la medida de tendencia central más representativa en el caso de variables que sólo admiten la escala ordinal.

Definición **1.4.2.** La mediana x̃ de la muestra ordenada, la cual se denotará como $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, es aquel número real tal que el 50 % de los valores de la muestra son menores o iguales a él y el otro 50 % es mayor o igual.

El valor de la mediana está dado por:

$$\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{si } n \text{ es impar,} \\ \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} & \text{si } n \text{ es par.} \end{cases} \tag{1.3}$$

Ejemplo **1.4.3.** Las calificaciones de un estudiante en seis exámenes han sido 8.4, 9.1, 7.2, 6.8, 8.7 y 7.8. Hallar la mediana de esta muestra.

Solución:

Primero ordenemos los datos de menor a mayor magnitud:

$$6.8, 7.2, 7.8, 8.4, 8.7, 9.1$$

Como el número de datos es par, la Fórmula 1.3 indica que primero deben identificarse x_3 y x_4, es decir, los valores que se encuentran en la tercera y cuarta posición con respecto a la muestra ordenada, los cuales son 7.8 y 8.4 respectivamente. Una vez obtenidos estos valores se procede a promediarlos y así la mediana es:

$$x = \frac{7.8 + 8.4}{2}$$

Si se está trabajando con datos agrupados se hablará del intervalo mediano que es aquel cuya frecuencia absoluta acumulada es mayor o igual a la total. En tal caso, la mediana se puede obtener por interpolación como se ve en el ejemplo o bien si $M = [L_{inf}, L_{sup}]$ es el intervalo mediano, entonces L_i es su límite inferior; A y B sus frecuencias, absoluta acumulada y absoluta, respectivamente y C la amplitud de M, la mediana está dada por la siguiente fórmula:

$$\tilde{x} = L_{inf} + \frac{n/2 - A}{B} \cdot C. \tag{1.4}$$

Ejemplo **1.4.4.** La siguiente tabla muestra una distribución de frecuencias para las puntuaciones de 120 estudiantes en un examen final de álgebra. Hallar la mediana de las calificaciones por medio de interpolación y mediante la fórmula 1.4.

	Marca de clase	f_i	F_i
[30,40)	35	1	1
[40,50)	45	3	4
[50,60)	55	11	15
[60,70)	65	21	36
[70,80)	75	43	79
[80-90)	85	32	111
[90-100)	95	9	120

1. Por interpolación.

   Si se considera que el peso de cada estudiante puede tomar cualquier valor dentro de cada intervalo de la tabla y ya que la mediana es aquel peso para el cual la mitad de la frecuencia total ($120/2 = 60$) queda por encima y la mitad por abajo se puede proceder de la siguiente forma: Se buscará el intervalo para el cual la frecuencia acumulada (F_i) sea inmediatamente menor que la mitad de los datos, en este caso es el intervalo $[60,70)$ con $F_i = 36$. Luego para llegar al 60 deseado tomamos 24 calificaciones más de los 43 elementos de intervalo $[70,80)$, es decir, la mediana debe estar a 24/43 de camino entre 70 y 80; por lo tanto la mediana es

   $$70 + \frac{24}{43}(80-70) = 70 + \frac{240}{43} = 75.58$$

   .

2. Por fórmula:

   El intervalo mediano es $[70,80)$ ya que su frecuencia absoluta acumulada es inmediatamente mayor a la mitad de los datos. Entonces

$$
\begin{aligned}
L_{inf} &= \text{Limite inferior intervalo mediano} = 70 \\
n &= \text{número de datos} = 120 \\
A &= \text{Frecuencia acumulada del intervalo anterior} \; = 36 \\
B &= \text{Frecuencia del intervalo mediano} = 43 \\
C &= \text{Amplitud del intervalo mediano} = 10
\end{aligned}
$$

luego,

$$\tilde{x} = L_{inf} + \frac{n/2 - A}{B} \cdot C = 70 + \frac{60 - 36}{43} \cdot 10 = 75.58$$

La siguiente medida no se ve afectada por ninguna característica de la muestra como el tamaño de los datos o la existencia de valores extremos y aunque en dados casos esto se ve como una ventaja, en general es una medida bastante inestable ya que un pequeño cambio en las observaciones puede afectarle mucho.

Definición **1.4.3.** La moda, que se denota por m_0 es el dato que más se repite en la muestra y, a diferencia de las medidas anteriores, se puede usar en datos categóricos.

Para datos agrupados, el intervalo modal,M, será aquel con mayor frecuencia. Se denotará por L_{inf} al límite inferior de M, a D como la frecuencia de la clase posterior al intervalo modal, a E como la frecuencia del intervalo anterior a M y C la amplitud de la clase modal, entonces la moda para datos agrupados se puede obtener a través de la fórmula

$$m_o = L_{inf} + \frac{D}{D+E} \cdot C.$$

Ejemplo **1.4.5.** Para los datos del ejercicio 1.4.4 encontrar la calificación modal El intervalo con mayor frecuencia (clase modal) es $M = [70, 80)$ de esto se obtiene los datos:

$$\begin{aligned} L_{inf} &= \text{Limite inferior de clase modal} = 70 \\ D &= \text{Frecuencia de la clase anterior al intervalo modal} = 21 \\ E &= \text{Frecuencia de la clase posterior al intervalo modal} = 32 \\ C &= \text{Amplitud de clase modal} = 10 \end{aligned}$$

Entonces, la moda de la muestra es

$$m_0 = L_{inf} + \frac{D}{D+} \cdot C = 70 + \frac{21}{21+32} \cdot 10 = 73.96$$

Medidas de posición

Las medidas anteriores son una comparación de los datos con un sólo valor. Como complemento se tienen las medidas de posición, las cuales permiten comparar una porción de datos con el resto de la muestra. Se llaman medidas de posición o *cuantiles* a aquellos que dividen a la población en k partes iguales. La mediana es una caso particular de un cuantil ya que divide a la población en dos partes iguales. Los más usados son los *cuartiles* y *percentiles*.

Los *cuartiles* dividen a la muestra en 4 partes iguales y pueden o no ser parte de la misma. El primer cuartil (Q_1) es el número real tal que, el 25% de los datos son menores o iguales a él. El segundo cuartil o mediana (Q_2) es aquel número tal que el 50% de los elementos de la muestra es menor o igual. Por último, el tercer cuartil (Q_3) tiene la propiedad que a lo más el 75% de los valores son menores o iguales a Q_3. Para datos no agrupados el cuartil k, $k = 1, 2, 3$, es igual a

$$Q_k = x_{\left(\frac{k(n+1)}{4}\right)}$$

Es decir, Q_k es el valor ordenado que está en la posición $\frac{k(n+1)}{4}$ en caso de que este sea entero, de no ser así se promedian los valores que estén en las posiciones $\lfloor \frac{k(n+1)}{4} \rfloor$ y $\lfloor \frac{k(n+1)}{4} \rfloor + 1$.

Ejemplo **1.4.6.** Hallar los tres cuartiles para los siguientes datos

8	11	4	3	2	5	10	6	4	1
10	8	12	6	5	7				

Al ordenar la muestra se obtiene

1	2	3	4	4	5	5	6	6	7
8	8	10	10	11	12				

Como $\frac{k(n+1)}{4} = \frac{1(16+1)}{4} = 4.25$ no es un número entero, entonces Q_1 será el promedio de lo valores $x_{(4)} = 4$ y $x_{(5)} = 4$, es decir, $Q_1 = \frac{4+4}{2} = 4$. De manera análoga para el segundo cuartil, $\frac{2(n+1)}{4} = \frac{2(16+1)}{4} = 8.5$ que por no ser entero $Q_2 = \frac{x_{(8)}+x_{(9)}}{2} = \frac{6+6}{2} = 6$. Para Q_3, $\frac{k(n+1)}{4} = \frac{3(16+1)}{4} = 12.75$ por lo tanto, $Q_3 = \frac{x_{(12)}+x_{(13)}}{2} = \frac{8+10}{2} = 9$.

$$1\ 2\ 3\ 4\ \boxed{Q_1 = 4}\ 4\ 5\ 5\ 6\ \boxed{Q_2 = 6}\ 6\ 7\ 8\ 8\ \boxed{Q_3 = 9}\ 10\ 10\ 11\ 12$$

Por otro lado , para datos agrupados hay que considerar al intervalo que contiene a Q_k,el cual se denotará por I_{Q_k}, esté es el primer intervalo para el cual $F_i \geq \frac{n \cdot k}{4}$. Entonces el i-ésimo cuartil está dado por la fórmula

$$Q_k = L_i + \left(\frac{\frac{n \cdot i}{4} - A}{B} \right) \cdot C. \tag{1.5}$$

Donde,

L_i = Límite inferior del intervalo que contiene a Q_k.

A = Frecuencia acumulada del intervalo anterior a la clase I_{Q_k}.

B = Frecuencia del intervalo I_{Q_k}.

C = Amplitud del intervalo I_{Q_k}.

Ejemplo **1.4.7.** Haciendo uso de la tabla de frecuencias del Ejercicio1.4.4 encontrar los cuartiles Q_1, Q_2 y Q_3 e interpretar su significado.

- Para Q_1. En la tabla se puede apreciar que la primera clase cuya frecuencia acumulada, F_i, es mayor o igual a $\frac{k \cdot n}{4} = \frac{1 \cdot 120}{4} = 30$ es el intervalo $[60, 70)$. Una vez determinado este intervalo es posible encontrar los datos necesarios para hacer uso de la Fórmula 1.5.

Como,

$L_4 =$ Límite inferior del intervalo que contiene a $Q_1 = 60$

$A =$ Frecuencia acumulada del intervalo anterior a la clase $I_{Q_1} = 15$

$B =$ Frecuencia del intervalo $I_{Q_1} = 21$

$C =$ Amplitud del intervalo $I_{Q_1} = 10$

Entonces, el primer cuartil es

$$L_i + \left(\frac{\frac{120\cdot 1}{4} - A}{B}\right)\cdot C = 60 + \left(\frac{30-15}{21}\right)\cdot 10 = 67.1$$

- Para Q_2. El primer intervalo de clase para el cual la frecuencia acumulada es mayor o igual que $\frac{k\cdot n}{4} = \frac{2\cdot 120}{4} = 60$ es $I_{Q_2} = [70, 80)$ por lo que,

$L_5 =$ Límite inferior del intervalo que contiene a $Q_2 = 70$

$A =$ Frecuencia acumulada del intervalo anterior a la clase $I_{Q_2} = 36$

$B =$ Frecuencia del intervalo $I_{Q_2} = 43$

$C =$ Amplitud del intervalo $I_{Q_2} = 10$

Luego, el cuartil 2 es

$$L_i + \left(\frac{\frac{120\cdot 2}{4} - A}{B}\right)\cdot C = 70 + \left(\frac{60-36}{43}\right)\cdot 10 = 75.5$$

- Para Q_3. La sexta clase, $[80, 90)$, es la que contiene a Q_3 pues es el primer intervalo para el cual se cumple $F_6 = 111 \geq \frac{n\cdot k}{4} = \frac{120\cdot 3}{4} = 90$, así que,

$$
\begin{aligned}
L_6 &= \text{Límite inferior del intervalo que contiene a } Q_3 = 80\\
A &= \text{Frecuencia acumulada del intervalo anterior a la clase } I_{Q_3} = 79\\
B &= \text{Frecuencia del intervalo } I_{Q_3} = 32\\
C &= \text{Amplitud del intervalo } I_{Q_3} = 10
\end{aligned}
$$

Por lo tanto,

$$
Q_3 = L_i + \left(\frac{\frac{120\cdot 3}{4} - A}{B}\right)\cdot C = 80 + \left(\frac{90-79}{32}\right)\cdot 10 = 83.4
$$

Como $Q_1 = 67.1$, $Q_2 = 75.5$ y $Q_3 = 83.4$ entonces de los estudiantes el 25% obtuvo una calificación menor o igual a 67, el 50% tuvo 75 o más y sólo el 25% tuvo 83 o más.

Los ***percentiles*** son los valores que, al igual que los cuartiles, pueden o no pertenecer a la muestra y dividen a la muestra ordenada, en 100 partes iguales. Se llama P_k al k-ésimo percentil para el cual k% de los datos son menores o iguales a él, mientras que el $(100-k)$% son mayores o iguales. El k-ésimo percentil, con $k = 1, \ldots, 100$, para datos no agrupados está dado por

$$
P_k = x_{\left(\frac{k(n+1)}{100}\right)}
$$

Análogo al calculo de cuartiles, si $\frac{k(n+1)}{100}$ es entero, entonces P_k es el valor ordenado que está en la posición $\frac{k(n+1)}{100}$. En caso contrario, P_k es el promedio de los valores que estén en las posiciones $\lfloor\frac{k(n+1)}{100}\rfloor$ y $\lfloor\frac{k(n+1)}{100}\rfloor + 1$.

Para datos agrupados, si I_{P_k} es el intervalo que contiene al k-ésimo percentil y L_i el límite inferior de I_{P_k}, A la frecuencia acumulada del intervalo anterior a I_{P_k}, B la frecuencia del intervalo I_{P_k}, C la amplitud del intervalo que contiene a P_k. El percentil k se calcula de la siguiente forma:

$$P_k = L_i + \left(\frac{\frac{n \cdot k}{100} - A}{B} \right) \cdot C. \tag{1.6}$$

Equivalente a lo anterior, I_{P_k} es el primer intervalo para el cual su frecuencia acumulada F_i es mayor o igual que $\frac{n \cdot k}{100}$

Ejemplo **1.4.8.** De la tabla de frecuencias del Ejemplo 1.2.3 obtener las frecuencias acumuladas absolutas así como 30^{vo} percentil.

Salarios	f_i	F_i
[250-260)	8	8
[260-270)	10	18
[270-280)	16	34
[280-290)	14	48
[290-300)	10	58
[300-310)	5	63
[310-320)	2	65

- El intervalo donde se encuentra $P_3 0$ es $I_3 0 = [270, 280)$ pues es el primero tal que la frecuencia absoluta acumulada es mayor a $\frac{65 \cdot 30}{100} = 19.5$. Entonces,

$$L_i = 270$$

$$A = \text{frecuencia acumulada del intervalo anterior} = 18$$

$$B = \text{frecuencia del intervalo} I_3 0 = 16$$

$$c = \text{amplitud del intervalo} = 10$$

Por la Fórmula 1.6

$$P_3 0 = 270 + \left(\frac{19.5 - 18}{16}\right) \cdot 10 = 270.93$$

Esto significa que el 30 % de los trabajadores ganan 270.93 pesos o menos al día.

Medidas de dispersión

Las medidas de dispersión indican el nivel de concentración de los datos y como consecuencia determinan que tanto una medida central puede ser usada para representar a la muestra.

La medida de dispersión más simple es el rango:

$$\text{Rango} = \text{Valor máximo - Valor mínimo},$$

ya que para su cálculo sólo se necesitan el valor máximo y mínimo de la muestra. Ésta, generalmente, se usa como medida preliminar de dispersión, ya que tiene dos grandes desventajas: no da información acerca de la distribución que tienen los datos entre sí; y es extremadamente sensible a la presencia de valores extremos en la muestra.

Para estudiar la variabilidad de los datos se suelen utilizar medidas que involucran la diferencia de los datos con respecto a la media. De manera natural se podría pensar en $\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n}$, sin embargo $\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n} = \frac{\sum_{i=1}^{n} x_i}{n} - \frac{\sum_{i=1}^{n} \bar{x}}{n} = \bar{x} - \frac{n \cdot \bar{x}}{n} = 0$, por lo que no es útil como medida de dispersión. Al hablar de distancias se busca que las diferencias sean positivas por lo que dos maneras de solucionar este inconveniente son:

- Utilizar el valor absoluto, lo cual da lugar a la llamada *desviación media* (d.m.)

$$\text{d.m.} = \frac{\displaystyle\sum_{i=1}^{n} |x_i - \bar{x}|}{n}. \tag{1.7}$$

  Si se tienen datos agrupados, cada x_i se sustituirá por la marca de clase de cada intervalo y el valor absoluto de las diferencias se multiplicará por la frecuencia absoluta de la clase, lo cual da lugar a la fórmula

$$\text{d.m.} = \frac{\displaystyle\sum_{i=1}^{n} |m_i - \bar{x}| f_i}{n}. \tag{1.8}$$

- Si se sustituye el valor absoluto por el cuadrado de las diferencias se obtiene la *varianza muestral* (S^2) dada por

$$S^2 = \frac{\displaystyle\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}. \tag{1.9}$$

Para datos agrupados, la varianza se calcula como

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

A la raíz cudrada de esta suma se le conoce como desviación estándar (S).

Cabe mencionar que en algunos textos se sustituye el denominador $n-1$ por n. Véase (.....)

La varianza por ser una medida que depende de la media no será útil en la existencia de valores extremos.

El llamado ***Rango intercuartil***, se calcula como la diferencia entre los cuartiles 3 y 1; es decir, $R_Q = Q_3 - Q1$, también permite conocer la variabilidad de los datos entre sí pero, a diferencia de las anteriores, no es sensible a valores extremos pues sólo mide la variabilidad de la mitad central de los datos, si la amplitud intercuartil es muy pequeña, significa que los valores están muy próximos entre sí, es decir, hay poca dispersión o variabilidad y, por lo tanto, el valor central (que es justo la mediana) resulta muy representativo. Si la amplitud intercuartil es muy grande ocurre justo lo contrario.

Un elemento que permite visualizar gráficamente algunas medidas es el *diagrama de caja y brazos* con el cual se identifica de manera simple la mediana, los tres cuartiles, el valor máximo y mínimo de la muestra y la dispersión por cuartiles y la presencia de valores extremos. También es usado para comparar dos muestras entre si.

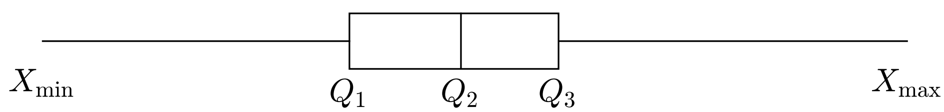


Figura 1.4: Caja y brazos

Como se muestra en la Figura b) los extremos de la caja representan el primer y último cuartil. La linea vertical corresponde a la mediana y el brazo izquierdo representa la distancia del valor mínimo a (Q_1) mientras que el derecho la distancia de Q_3 al valor máximo.

Ejemplo **1.4.9.** En un estudio sobre el tiempo diario (en minutos) que dedican a entrenar los deportistas de una academia se obtuvieron los siguientes datos:

60	66	77	70	66	68	57	70	66	52
75	65	69	71	58	66	67	74	61	63
69	80	59	66	70	67	78	75	64	71
81	62	64	69	68	72	83	56	65	74
67	54	65	65	69	61	67	73	57	62
67	68	63	67	71	68	76	61	62	63
76	61	67	67	64	72	64	73	79	58
67	71	68	59	69	70	66	62	63	66

a) calcule su media, moda y cuartiles.

b) Construya su gráfica de cajas.

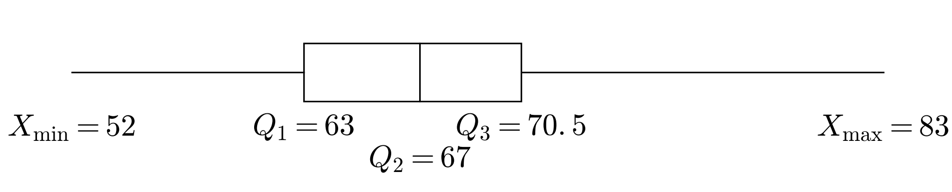
a) Al ordenar los datos se obtiene

52	54	56	57	57	58	58	59	59	60
61	61	61	61	62	62	62	62	63	63
63	63	64	64	64	64	65	65	65	65
66	66	66	66	66	66	66	67	67	67
67	67	67	67	67	67	68	68	68	68
68	69	69	69	69	69	70	70	70	70
71	71	71	71	72	72	73	73	74	74
75	75	76	76	77	78	79	80	81	83

La media es $\frac{\sum_{i=1}^{80} X_i}{80} = 67.025$, la moda es 67, el menor valor entre los datos es 52, el mayor 83, los tres cuartiles son

$$Q_1 = \frac{X_{20} + X_{21}}{2} = \frac{63 + 63}{2} = 63, \qquad Q_2 = \frac{X_{40} + X_{41}}{2} = \frac{67 + 67}{2} = 67$$
$$Q_3 = \frac{X_{60} + X_{61}}{2} = \frac{70 + 71}{2} = 70.5,$$

b) En la gráfica de cajas y brazos se puede observar el rango (31 unidades) y rango intercuartil (7.5 unidades). La mitad de los deportistas se dedican a entrenar entre 63 y 70.5 minutos al día mientras que el 25 % entrenan entre 52 y 63 minutos y el resto entre 70.5 y 83 minutos diarios.



Medidas de forma

Mientras mayor sea el tamaño de la muestra se espera poder escoger intervalos de clase muy pequeños y tener un número significante de observaciones en cada intervalo. Como consecuencia el polígono de frecuencias (absolutas o relativas) se verá casi como una curva continua, a la que se referirá como curva de distribución o frecuencia. En lo siguiente se analizará la "forma" de éstas a través de las siguientes medidas **asimetría** y **curtosis o apuntamiento**.

Para poder definirlas es necesario usar los momentos muestrales de un conjunto de datos observados.

El r-ésimo momento muestral con respecto al origen se define mediante

$$m_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r \tag{1.11}$$

mientras que para datos agrupados, distribuidos en N intervalos de clase,

$$m_r = \frac{1}{n}\sum_{i=1}^{N} x_i^r f_i \tag{1.12}$$

Por otro lado, el r-ésimo momento muestral con respecto a la media, para datos no agrupados y agrupados respectivamente, es

$$M_r = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^r \tag{1.13}$$

$$M_r = \frac{1}{n}\sum_{i=1}^{N} (x_i - \bar{x})^r f_i. \tag{1.14}$$

La forma de la distribución se puede determinar mediante el sesgo, el cual es una medida del grado de simetría, con respecto a la media, de los datos y tiene como finalidad mostrar hacia que lado de la media se ubican más valores. Para determinar el sesgo de una muestra se usa el *coeficiente de Fisher* $g_1 = \dfrac{M_3}{S^3}$, el cual es el cociente entre el tercer momento muestral con respecto a la media y la desviación estándar o el *coeficiente de Pearson* $g_2 = \dfrac{\bar{x} - m_0}{S}$ donde el numerador es la diferencia de la media con la moda y el denominador es la desviación estándar.

Ambos coeficientes se comparan con el valor cero bajo los siguientes criterios:

a) Si $g_i < 0$ con $i = 1, 2$ se dice que la distribución es asimétrica negativa o bien, es sesgada a la izquierda, es decir, hay mayor número de datos a la derecha de la media. Entonces se cumple la relación $m_o < med < \bar{x}$.

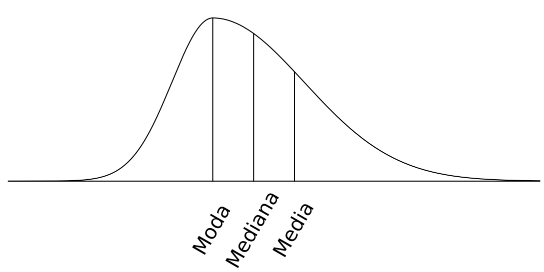


Figura 1.5: Asimetría a la izquierda

b) Si $g_i > 0$ con $i = 1, 2$ la distribución es asimétrica positiva o, dicho de otra forma, es sesgada a la derecha, es decir, hay más datos a la izquierda de la media y en este caso se cumple la relación $m_o > med > \bar{x}$.

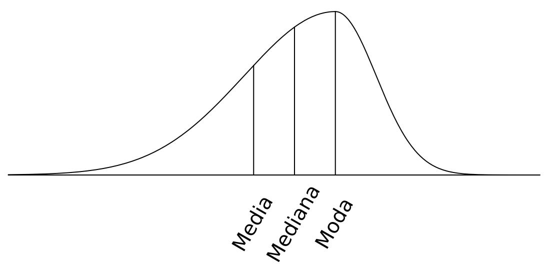


Figura 1.6: Asimetría a la derecha

c) Si $g_i = 0$ con $i = 1, 2$ entonces la distribución es simétrica, es decir, existe la misma cantidad de datos a ambos lados de la media y se cumple $m_o = med = \bar{x}$.

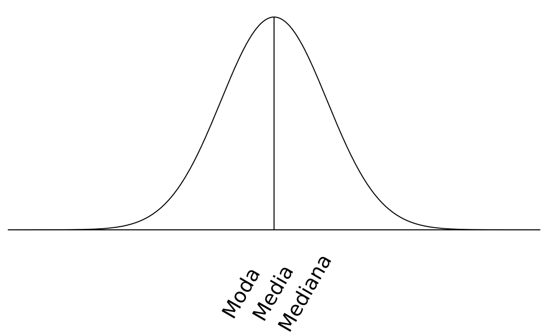


Figura 1.7: Asimetría central

Para distribuciones unimodales, moderadamente asimétricas y campaniformes, Fisher dio una medida de la curtosis o picudez de la distribución que indica mayor o menor concentración de la información. Una muestra con una importante curtosis supone que los datos están más cercanos a la media lo que a su vez implica una varianza pequeña. Por el contrario, mientras más aplastada sea la distribución o la curtosis sea pequeña significa que los datos se alejan más de la media. Esta medida se compara con la distribución normal (campana de Gauss), mediante el siguiente coeficiente $K = \frac{M_4}{S^4}$. El valor de comparación para este coeficiente es 3, por ser la curtosis de la distribución normal, bajo el siguiente criterio;

1. Si $K = 3$ se trata de una distribución mesocúrtica y entonces la distribución de los datos se verá como una campana de Gauss.

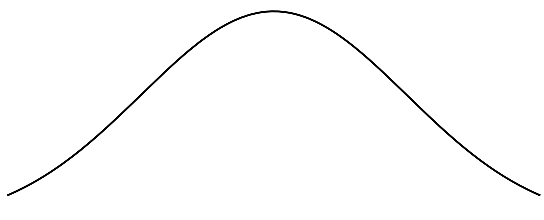


Figura 1.8: Mesocurtica

2. Si $K < 3$ los datos tienen una distribución aplanada y en este caso se dice que tiene forma Platicúrtica.

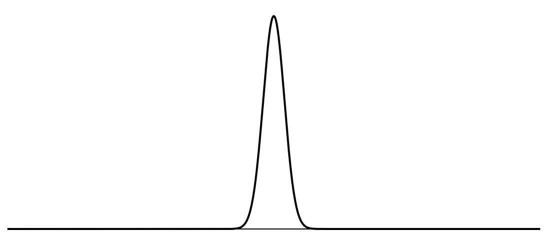


Figura 1.9: platicurtica

3. Si $K > 3$ la distribución es más puntiaguda que una normal. Se llama distribución Leptocúrtica.

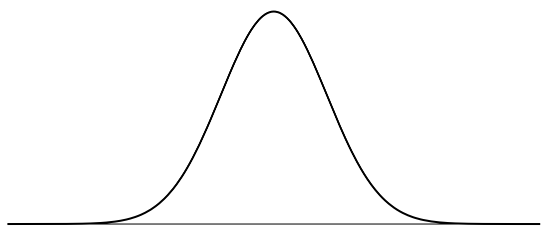


Figura 1.10: leptocurtica

Ejercicios

1. Clasifique las siguientes variables en cualitativas o cuantitativas. Después clasifique las variables cualitativas y las variables cuantitativas en discretas o continuas.

   a) El número de llamadas telefónicas recibidas en una oficina en un día.

   b) La religión de una persona.

   c) La velocidad de un auto en una carrera.

   d) Si una persona fuma o no fuma.

   e) El número de águilas que caen al lanzar 10 veces una moneda.

2. Considere los siguientes datos sobre el tipo de problemas de salud (J = hinchazón de las articulaciones, F = fatiga, B = dolor de espalda, M = debilidad muscular, C = tos, N = escurrimiento nasal/irritación, O = otro) que aquejan a los plantadores de árboles. Obtenga las frecuencias absolutas y relativas de las diversas categorías, dibuje una gráfica de barras y un diagrama circular.

O	O	N	J	C	F	B	B	F	O	J	O	O	M	O
F	F	O	O	N	O	N	J	F	J	B	O	C	J	O
J	J	F	N	O	B	M	O	J	M	O	B	O	F	J
O	O	B	N	C	O	O	O	M	B	F	J	O	F	N

3. Transductores de temperatura de cierto tipo se envían en lotes de 50. Se seleccionó una muestra de 60 lotes y se determinó que el número de transductores en cada lote que no cumple con las especificaciones de diseño y se obtuvieron los datos siguientes:

2	1	2	4	0	1	3	2	0	5	3	3	1	3	2	4	7	0	2	3
0	4	2	1	3	1	1	3	4	1	2	3	2	2	8	4	5	1	3	1
5	0	2	3	2	1	0	6	4	2	1	6	0	3	3	3	6	1	2	3

Organice los datos en una tabla de frecuencias con frecuencia absoluta, relativa, absoluta acumulada y relativa acumulada.

4. Los siguientes datos muestran el número de clientes en un centro de servicio de General Motors en una muestra de 30 días:

67	76	58	82	59	51	63	69	70	75
67	43	58	61	40	58	46	57	72	71
65	73	58	45	48	49	50	64	53	64

Construya una gráfica de tallo y hoja.

5. El presidente de Occan Airlines intenta hacer una estimación de cuánto se tardará el Departamento de Aeronáutica Civil (DAC) en decidir acerca de la solicitud de la compañía sobre una nueva ruta entre Charlotte y Nashville. Los asesores del presidente han organizado los siguientes tiempos de espera de las solicitudes formuladas durante el año anterior. Los datos se expresan

en días, desde la fecha de la solicitud hasta la respuesta del DAC.

34	40	23	28	31	40	25	33	47	32
44	34	38	31	38	42	26	35	27	31
29	40	31	30	34	31	38	35	37	33
24	44	37	39	32	36	34	36	41	39
29	22	28	44	51	31	44	28	47	31

a) Construya una tabla de distribución de frecuencias absoluta, relativa, absoluta acumulada y relativa acumulada.

b) Dibuje un histograma.

c) Dibuje el polígono de frecuencias.

d) Dibuje la ojiva.

6. Se tiene el siguiente conjunto de datos: 5, 14, 4, 8, 9, 10, 4, 7, 6, 8.

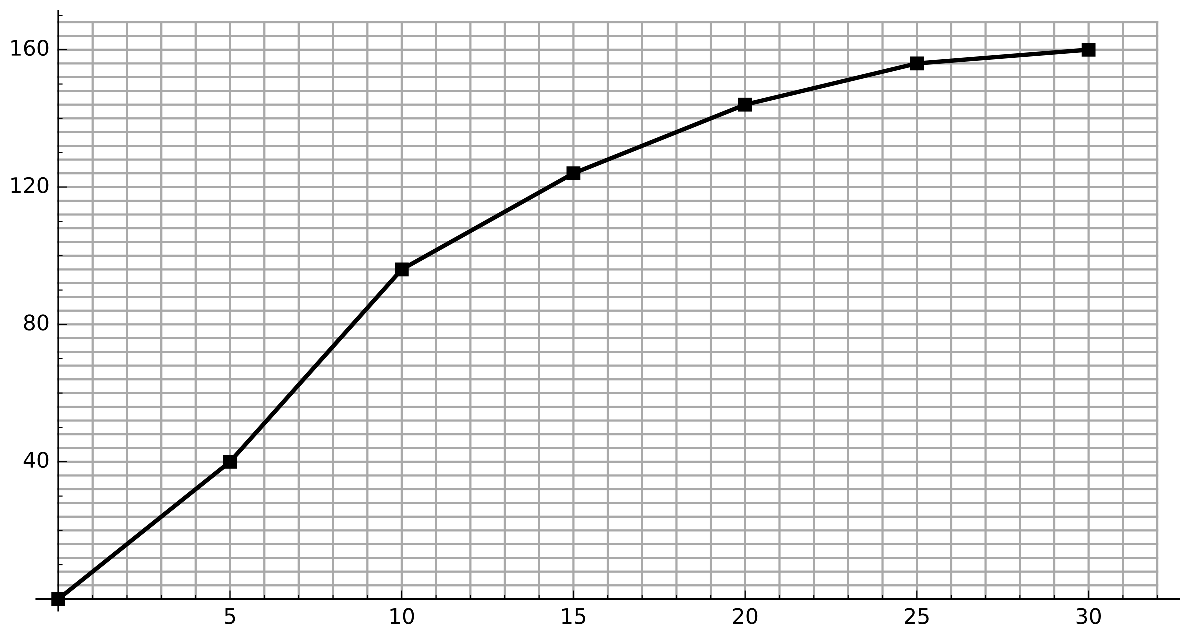
   a) Calcule la media, moda y mediana.

   b) Calcule el rango, varianza, desviación estándar y coeficiente de variación.

   c) Calcule los tres cuartiles (Q_1, Q_2, Q_3), D_5, P_25 y P_50.

   d) Realice el diagrama de caja y brazos.

7. La figura muestra la ojiva de la longitud de las llamadas telefónica (en minutos) realizadas desde mi casa durante los primeros 6 meses de este año.



a) Construya la correspondiente tabla de frecuencias con las columnas: intervalo, frecuencia absoluta y frecuencia acumulada.

b) Encuentre los cuartiles de la longitud de las llamadas.

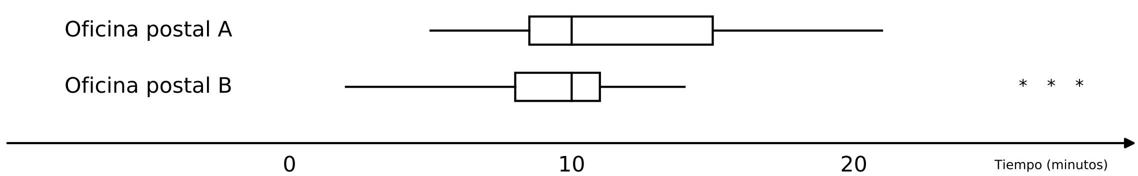
c) ¿Cuántas llamadas duraron menos de 17 minutos?

d) ¿Cuántas llamadas duraron entre 3 y 8 minutos?

e) El 20 % de las llamadas más largas duraron más de x minutos. ¿Cuál es el valor de x?

8. Una señora de mantenimiento de una escuela se acerca a la edad de su jubilación. Ella vive justo en medio de dos oficinas postales A y B, y tiene que decidir en cuál de las dos va a hacer los arreglos para recibir su pensión. Por algunos meses, deliberadamente, ha utilizado alternativamente las dos

oficinas postales cuando ha requerido de los servicios. En cada una de esas visitas ha registrado el tiempo que invierte desde que entra a la oficina hasta que es atendida. Las gráficas de caja siguientes muestran los tiempos de espera para las dos oficinas postales. El símbolo $*$ representa un valor atípico.



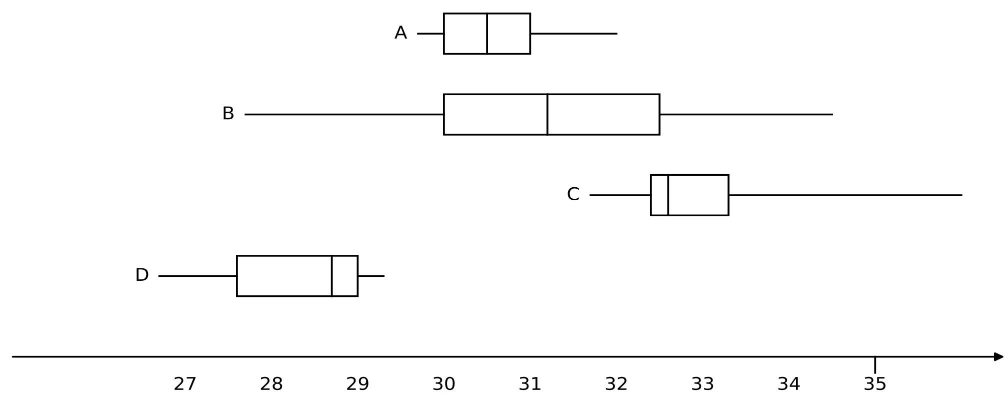
a) Compare las distribuciones de los tiempos de espera en las dos oficinas postales.

b) ¿Cuál de las dos oficinas postales le recomendaría utilizar a la señora si el valor atípico fuera debido a :

   1) falta de empleados en la oficina postal?

   2) que la compañía eléctrica haya cortado el suministro de energía de la oficina postal?

9. Un grupo de atletas corre frecuentemente cierto trayecto a campo traviesa para entrenar. Las siguientes gráficas representan los tiempos que les toma a los atletas A, B, C y D completar el trayecto.



a) Compare los tiempos de los atletas C y D. Describa la distribución de los tiempos de cada uno de los atletas (C y D)

b) ¿A cuál de los atletas A o B seleccionaría para ganar una carrera contra C. Explique la razón.

c) ¿A cuál de los atletas A o B seleccionaría para ganar una carrera contra D. Explique la razón.

d) Si A y B compitieran en un carrera, ¿Cuál de los dos cree que tendría mayor probabilidad de ganar? ¿Por qué?

10. De acuerdo con la Oficina de Censos en E. U., en 2001 el ingreso mediano por hogar era de \$44.517 para blancos y \$29.470 para negros, mientras que la media era de \$60.512 para blancos y \$39.248 para negros. ¿Esto sugiere que la distribución de los ingresos por hogar para cada raza es simétrico, sesgado a la derecha o sesgado a la izquierda? Explique.

11. Considere el conjunto de datos:

4.5	3.2	3.5	3.9	3.5	3.9
4.3	4.8	3.6	3.3	4.3	4.2
3.9	3.7	4.3	4.4	3.4	4.2
4.4	4.0	3.6	3.5	3.9	4.0

a) Construya un diagrama de tallo y hoja usando el dígito principal como tallo.

b) Construya un diagrama de tallo y hoja usando cada dígito principal dos veces. ¿Esta técnica mejora la presentación de los datos? Explique.

12. Una variable en estudio mide cuántos accidentes de auto serios tuvo un sujeto durante el año pasado. Explique por qué la media probablemente sería más útil que la mediana para presumir las respuestas de 60 sujetos.

13. Una Encuesta General Social actual preguntó a las mujeres entrevistadas: "¿Cuántas parejas sexuales has tenido en los últimos 12 meses?"De 365 entrevistadas, 102 dijeron 0 parejas, 233 dijeron 1 pareja, 18 dijeron 2 parejas, 9 dijeron 3 parejas, 2 dijeron 4 parejas y 1 dijo 5 parejas.

    a) Organice los datos en una tabla. Explique por qué la mediana es 1.

    b) Muestre que la media es 0.85.

    c) Supón que las 102 mujeres que respondieron 0 parejas hubieran contestado que 5 parejas. Muestre que la mediana seguiría siendo 1 y que la media aumentaría ¿Por qué?

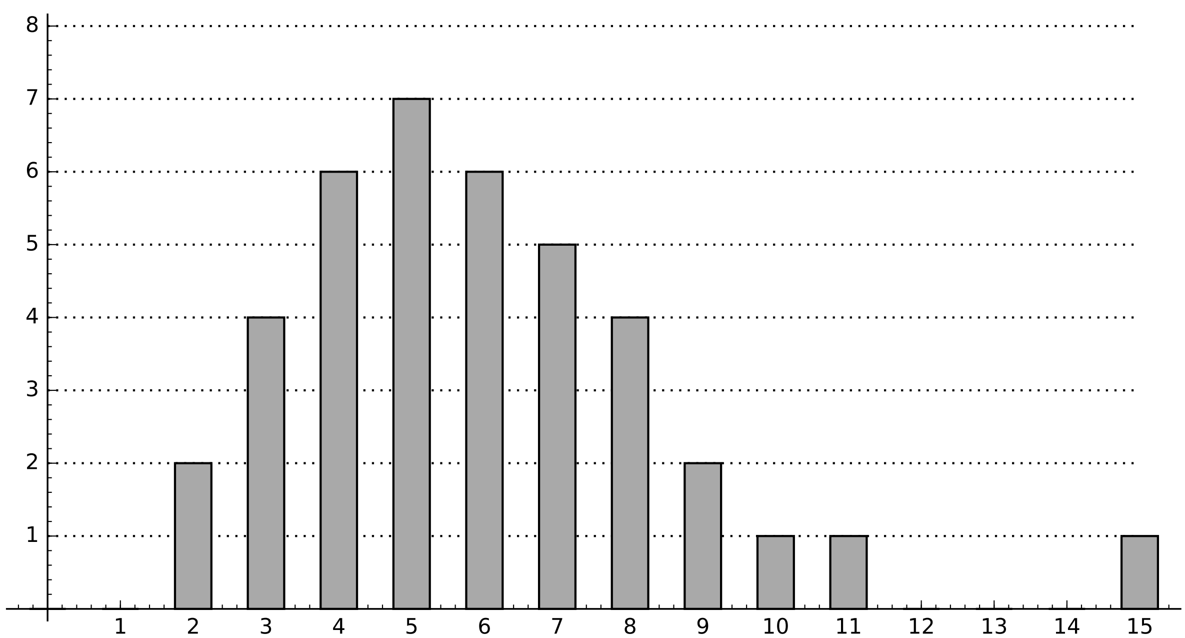
14. a) Calcule la media y la desviación estándar de los cuatro números 2, 3, 6 y 9.

    b) Dos números, a y b, se agregan a este conjunto de datos, tal que la media se incrementa en 1 y la varianza se incrementa en 2.5. Encuentre a y b.

15. Un conjunto de números tiene media 22 y desviación estándar de 6. Si se suma 3 a cada número del conjunto y luego cada número resultante se duplica, encuentre la media y la desviación estándar del nuevo conjunto de datos.

16. Para un conjunto de 10 datos con $\sum x = 290$ y $\sum x^2 = 8469$. Encuentre la media y la varianza.

17. A continuación se muestra el diagrama de frecuencias para un conjunto de datos.



a) Encuentre la mediana y la moda de los datos.

b) Dado que la media es 5.95 y la desviación estándar es 2.58, explique por qué el valor 15 debe ser tratado como un valor extremo (outlier).

c) Explique cómo tratarías ese valor extremo si el diagrama representara

    1) Las edades (en años cumplidos) de los niños de una fiesta.

    2) La suma de los resultados obtenidos cuando se lanza un par de dados.

d) Encuentre la media y la moda de los datos después de eliminar el valor extremo.

e) *Sin hacer ningún cálculo*, explique qué efecto, si es que hay, va a tener en la media y en la desviación estándar el haber quitado el valor extremo.

f) ¿El diagrama exhibe sesgo positivo, negativo o no tiene sesgo? ¿Cómo es afectado el sesgo al remover el valor extremo?

18. La comercializadora de flores Emmot Bulb Co. vende flores sorpresa con bulbos de flores. Las flores se venden según su peso; en consecuencia el número de bulbos en cada bolsa puede variar, dependiendo de las variedades incluidas. El número de bulbos que hay en cada bolsa de una muestra de 20 son:

21	35	37	56	47	36	23	26	33	37
25	33	32	47	34	26	37	37	43	45

a) ¿Cuáles son la media y la mediana del número de bulbos por bolsa? Con base en la información obtenida, ¿Qué puede concluir de la forma de la distribución del número de bulbos por bolsa?

b) Construya el histograma y el diagrama de caja para los datos.

19. Los siguientes datos corresponden a las calificaciones obtenidas en un examen: 86, 92, 100, 93, 89, 95, 79, 98, 68, 62, 71, 75, 88, 86, 93, 81, 100, 86, 96, 52.

a) Dibuje un diagrama de tallos y hojas para estos datos.

b) Calcule la moda, mediana, el rango, la desviación estándar y el rango intercuantil.

c) ¿Qué porcentaje de las calificaciones se encuentran alejadas una desviación estándar de la media? ¿y para dos desviaciones estándar?

d) Se observan *outliers*. Explique

20. Se colecta una muestra de comprobantes de un expediente de gastos, las cantidades pagadas son: 276.72, 194.17, 259.83, 249.45, 201.43, 237.66, 199.28, 211.49, 240.16, 261.10, 226.21.

    a) Dibuje un histograma para los datos, señale en la gráfica los siguientes elementos: la media, la mediana, rango intercuantil. Explica.

    b) Calcule la varianza y el coeficiente de variación.

21. Dada la siguiente tabla de distribución de frecuencias , encuentre:

Número (x)	1	2	3	4	5	6
Frecuencia f	5	9	16	18	20	7

    a) La media.

    b) La desviación estándar.

    c) El coeficiente de variación.

    d) Los cuartiles.

    e) Agrega en la tabla la frecuencia relativa.

Inferencia Estadística

El objetivo de la inferencia estadística es conocer o acercarse , lo más posible, a la ley de probabilidad que subyace a un fenómeno aleatorio. Para llegar a este objetivo no es necesario trabajar con todo el conjunto de objetos de interés, al cual se le conoce como población $P = \{a_1, a_2, ..., a_N\}$ basta con estudiar a un subconjunto representativo de ellos. Cada elemento de P tiene una característica $X(a_i)$, $i = \{1, ..., N\}$, la cual es motivo de estudio y depende del elemento tomado a_i. Una vez que se realizan $n \leq N$ observaciones del fenómeno en cuestión se obtienen $X_1, X_2, X_3, ..., X_n$ registros con cierta distribución empírica.

La inferencia estadística sólo se centra en situaciones en las cuales la característica, X, tiene una distribución empírica que coincide con alguna ley de probabilidad con función de densidad (o de probabilidad) $f_X(x) = f(x;\theta)$ y la incertidumbre radica en el valor del parámetro θ que, si bien es fijo, se desconoce (algunas veces se usa Θ si se tiene un conjunto de parámetros).

La función $f_X(x)$ pertenece a una familia de densidades $\digamma\{f_\theta \mid \theta \in H \subseteq \mathbb{R}^k\}$ dónde H es el espacio parametrico. El objetivo del capítulo es hallar la densidad $f^* \in \digamma$ cuya distribución ajuste mejor los datos.

La principal razón para no trabajar con toda la población es que la obtención de toda la información requerida, en general, da lugar a costos elevados tanto económicos como en tiempo. Así que se opta por estudiar una muestra representativa de la población. Se dice que una muestra es representativa si los elementos que la componen no presentan ninguna característica que los diferencie de los restantes . Las conclusiones sacadas de ella se pueden extender a la población total.

Sea X la variable aleatoria asociada a la característica de interés de una población con función de distribución $F(x)$. Cada repetición del experimento proporciona la observación X_i de la característica X, repitiendo n veces el experimento se obtienen $X_1, X_2, \cdots, X_n$ observaciones cuya distribución de probabilidad es idéntica a la de X.

Definición **2.0.1.** Una ***muestra aleatoria*** es cualquier conjunto de variables aleatorias $X_1, X_2, ..., X_n$ que son independientes y tienen las misma función de distribución que la de la población.

Una vez obtenida la muestra aleatoria se opera con los valores observados para inferir acerca de la población de interés, así toda la información que se pueda obtener del o los parámetros queda totalmente determinada por la muestra.

Ya se menciono que es deseable trabajar con una muestra característica de la población, sin embargo debido a que su distribución no es del todo conocida no es posible determinar si una muestra es representativa o no. Ante este problema la única solución inmediata es dejarle la tarea al azar. Los procedimientos posibles para obtener una muestra aleatoria son diversos y tienen sus bases en la teoría de muestreo que no se estudiara en este texto.

Para determinar la función de probabilidad que se ajusta al experimento es preciso determinar el valor del parámetro de del cual depende la función. Los parámetros poblaciones son las características numéricas de la población. El conocimiento del parámetro permite describir la distribución de probabilidad de la característica estudiada.

Ejemplo **2.0.1.** Suponga que la variable X en estudio es la duración de un cierto componente eléctrico que no envejece; es decir que, si el componente sobrevive en el instante t su estado es el mismo que en un principio y la distribución del tiempo que falta para su reemplazo sigue siendo la misma que al principio. Debido a que la distribución exponencial es la única con esta propiedad de ‶pérdida de memoria″, la distribución de X sera exponencial.

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda X} & X \geq 0 \\ \\ 0 & X \leq 0 \end{cases}$$

Todo el desconocimiento acerca de la distribución teórica f, queda entonces centrado en el valor del parámetro $\lambda \in (0,\infty)$, o, equivalente, en el valor de la duración media $\frac{1}{\lambda}$, pues recuerde que la esperanza de una variable exponencial es el reciproco de su parámetro.

El siguiente concepto será fundamental para estimar los parámetros poblacionales.

Definición **2.0.2.** Un ***estadístico*** o ***estadística*** $T(\mathbb{X})$ es cualquier función $T(\mathbb{X}) : \mathbb{R}^n \longrightarrow \mathbb{R}^r$, $r \in \mathbb{N}$, que depende de la muestra $\mathbb{X} = (X_1, X_2, ..., X_n) \subset \mathbb{R}^n$ y que no involucra a ningún parámetro desconocido.

Ejemplo **2.0.2.** Las siguientes funciones son estadísticos que dependen de la muestra $\mathbb{X} = (X_1, X_2, ..., X_n)$

1. $T_1(\mathbb{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i$
2. $T_2(\mathbb{X}) = \frac{X_1^2+...X_n^2}{n}$
3. $T_3(\mathbb{X}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Definición **2.0.3.** Un ***estimador*** es cualquier estadística $T(\mathbb{X})$ que ayude a proporcionar información del parámetro de interés, Θ o θ.

Si el problema consiste en estimar todo un vector de parámetros, es decir, $\Theta = (\theta_1, \theta_2, ..., \theta_k)$, entonces se utilizan k estadísticas $T_1(X), T_2(X), ..., T_k(X)$ para tal propósito.

Definición **2.0.4.** Observados los valores $X_1 = x_1, X_2 = x_2, ..., X_m = x_m$ el ***estimado*** de θ es aquel valor real k tal que $T(x_1, x_2, ..., x_m) = k$

Es decir una estimación consiste en establecer un valor concreto para un parámetro. Por ejemplo si se quiere conocer el valor del parámetro "media de una población", se puede utilizar la media de la muestra como estimador, en particular si se tienen las observaciones $\{4,4,4,6,8,10\}$, entonces el estimado es $\bar{x} = \dfrac{4+4+4+6+8+10}{6} = 6.$

2.1 Estimación puntual

La estimación puntual consiste en seleccionar un estadístico muestral el cual generará un único número que servirá como estimación del parámetro poblacional.

Existen varias formas de estimar un parámetro. En esta sección se estudiarán dos métodos para la estimación puntual: *método de momentos* y *método de máxima verosimilitud*.

2.2 Método de momentos

El método de momentos es uno de los primeros procedimientos de estimación que comienza con los trabajos de Karl Pearson a finales del siglo XIX. Tiene la ventaja de ser muy sencillo y casi siempre genera, al menos, un estimador.

La idea general del método de momentos consiste en igualar ciertas características muestrales con sus respectivos atributos poblacionales. Para lo cual se necesita la siguiente definición.

Definición **2.2.1.** Sea X una variable aleatoria con función de densidad $f_X(x)$. A $\mathbb{E}(X^k)$ se le conoce como el momento poblacional no centrado de orden k ($k \in \mathbb{N}$), esto es

$$\mathbb{E}\left(X^k\right) = \begin{cases} \displaystyle\sum_{i=1}^{n} x^k \cdot f_X(x_i) & \text{si X es discreta,} \\ \\ \displaystyle\int_{-\infty}^{\infty} x^k \cdot f_X(x) & \text{si X es continua.} \end{cases}$$

Mientras que el k-ésimo momento poblacional se define como

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \tag{2.1}$$

Este método se basa en la versión general de la ley fuerte de los grandes números, la cual afirma que si se tiene una muestra aleatoria de tamaño n y $g : \mathbb{R} \longrightarrow \mathbb{R}$ una función continua tal que $\mathbb{E}[g(X)] < \infty$, entonces

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{\sum_{i=1}^{n} g(X_i)}{n} = \mathbb{E}[g(X)] \right) = 1.$$

En particular si $g(x) = x^k$,lo anterior sugiere que conforme el tamaño de la muestra aumenta, con probabilidad 1, el k-ésimo momento muestral converge al k-ésimo momento poblacional.

Utilizando este hecho y la Fórmula 2.1. Dada una muestra aleatoria de tamaño n de $f(X,\theta)$. El estimador por momentos para θ, es cualquier estadística $\widehat{\theta} = T(X)$ que satisface la siguiente ecuación

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k = \mathbb{E}_{\widehat{\theta}}\left(X^k\right). \tag{2.2}$$

Ejemplo **2.2.1.** Se desea estimar la probabilidad de que salga *sol* en una moneda. El experimento consiste en contar el número de águilas que aparecen antes del segundo sol. Para dicho experimento se toma una muestra de tamaño 76, la cual se muestra en la tabla 2.1.

valor de x	$n°$ de observaciones
0	19
1	8
2	13
3	17
4	13
5	2
6	2
7	0
8	0
9	0
10	1
11	0
12	1

Tabla 2.1: Número de águilas

Nótese que, dada la naturaleza del experimento, se puede definir a X como la variable aleatoria que cuenta el número de fracasos antes del segundo éxito, entonces se puede sugerir que X tiene distribución binomial negativa de parámetros $(2, p)$, donde p es la probabilidad de que caiga *sol*. El problema consiste en estimar el valor de p. Para este fin se usará el método de momentos.

La esperanza (media poblacional) de una variable aleatoria con distribución binomial negativa es $\frac{r\left(1-\widehat{p}\right)}{\widehat{p}}$ y si $g(x) = x$ entonces $\mathbb{E}_{\widehat{\theta}}[X] = \frac{2(1-\hat{p})}{\hat{p}}$, para encontrar el valor de $\hat{p}$ se debe resolver la siguiente ecuación:

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{2\left(1-\widehat{p}\right)}{\widehat{p}},$$

de lo que se sigue,

$$
\begin{aligned}
\bar{x} &= \frac{r\left(1-\widehat{p}\right)}{\widehat{p}} \\
\bar{x}\widehat{p} &= 2 - 2\widehat{p} \\
2 &= \bar{x}\widehat{p} + 2\widehat{p} \\
2 &= \widehat{p}(\bar{x}+2) \\
\widehat{p} &= \frac{2}{\bar{x}+2}
\end{aligned}
$$

Por lo tanto, el estimador por momentos para p está dado por $\widehat{p} = \dfrac{2}{\bar{x}+2}$.

Para la muestra dada se tiene que $\bar{x} = 2.38$ y el estimado es $\widehat{p} = 0.4$.

Ejemplo **2.2.2.** En una investigación biológica se determina que el número de células de un organismo, después de un día de ser fecundado, sigue una distribución tal que su función de probabilidad está dada por:

$$
f(x,\theta) = \frac{1}{\Gamma\left(\frac{x}{2}\right)} \sqrt{\theta^{x-2}}\, e^{-2\theta}\, 2^{\frac{x-2}{2}}; \quad x \in \{2,4,6,\ldots\}
$$

Para encontrar el estimador por momentos para el parámetro θ considérese que la estimación se realiza con base a una muestra aleatoria $x_1, x_2, \ldots, x_n$ de $f(x;\theta)$. Sea $g(x) = x$ entonces,

$$
\frac{1}{n}\sum_{i=1}^{n} x_i = \mathbb{E}_{\widehat{\theta}}(x_i),
$$

es decir,

$$
\bar{X} = \mathbb{E}_{\widehat{\theta}}(x_i).
$$

Además, si $n \in \mathbb{N}$, entonces

$$\mathbb{E}(X) = \sum_{x \in 2n} \frac{x}{\Gamma\left(\frac{x}{2}\right)} \sqrt{\theta^{x-2}}\, e^{-2\theta}\, 2^{\frac{x-2}{2}}$$

Haciendo el cambio de variable $y = \dfrac{x-2}{2}$, entonces

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{y=0}^{\infty} \frac{2y+2}{\Gamma(y+1)} \sqrt{\theta^{2y}}\, e^{-2\theta}\, 2^{y} \\
&= \sum_{y=0}^{\infty} (2y+2)\, \frac{e^{-2\theta}(2\theta)^{y}}{y!} \\
&= e^{-2\theta} \sum_{y=0}^{\infty} (2y+2)\, \frac{(2\theta)^{y}}{y!} \\
&= 2e^{-2\theta} \left[\sum_{y=0}^{\infty} y \frac{(2\theta)^{y}}{y!} + \sum_{y=0}^{\infty} \frac{(2\theta)^{y}}{y!} \right] \\
&= 2e^{-2\theta} \left[\sum_{y=0}^{\infty} \frac{(2\theta)^{y-1}}{(y-1)!} (2\theta) + e^{2\theta} \right] \\
&= 2e^{-2\theta} \left[2\theta e^{2\theta} + e^{2\theta} \right] \\
&= 4\theta + 2.
\end{aligned}$$

Por lo tanto, $\theta = \dfrac{\bar{X} - 2}{4}$.

Cabe mencionar que para distintos valores de K en $g(x) = x^{k}$ se pueden obtener distintos estimadores.

Ejemplo **2.2.3.** Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria *Poisson*(θ); $\theta > 0$. Encontrar el estimador por momentos para θ.

Para este ejemplo se van a considerar las funciones $g_1(x) = x$ y $g_2(x) = x^2$ para obtener el estimador del parámetro θ. Hay que recordar que, si X se distribuye $Poisson(\theta)$, entonces $\mathbb{E}[X] = \theta$ (primer momento poblacional) y $Var[X] = \theta$, lo cual implica que el segundo momento poblacional esta dado por $\mathbb{E}\left[X^2\right] = \theta + \theta^2$.

Para $g_1(x) = x$ se genera la ecuación

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \mathbb{E}_{\widehat{\theta}}[X] = \widehat{\theta}$$

de la cual se obtiene que el estimador por momentos para θ es $\widehat{\theta} = \bar{x}$.

Por otra parte, si $g_2(x) = x^2$ se igualara el segundo momento muestral $\frac{1}{n}\sum_{i=1}^{n} x_i^2$ al segundo momento poblacional $\mathbb{E}_{\widehat{\theta}}\left[X^2\right]$, es decir,

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 = \widehat{\theta} + \widehat{\theta}^2.$$

Lo anterior genera la ecuación de segundo grado

$$\widehat{\theta}^2 + \widehat{\theta} - \frac{1}{n}\sum_{i=1}^{n} x_i^2 = 0$$

y como $\theta > 0$ el estimador es

$$\widehat{\theta} = -\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{n}\sum_{i=1}^{n} x_i^2}.$$

Ejemplo **2.2.4.** Sea X una variable aleatoria continua con distribución $U(-\theta, \theta)$, encontrar el estimador por momentos para θ.

Como $\mathbb{E}[X] = 0$ entonces la función $g(x) = x$ no resulta útil, así que se usará $g(x) = x^2$, por lo que se necesitará el segundo momento poblacional $\mathbb{E}\left[X^2\right] = \frac{(-\theta - \theta)^2}{12} = \frac{4\theta^2}{12} = \frac{\theta^2}{3}$ entonces

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{\theta^2}{3}$$

de lo que se obtiene el estimador

$$\widehat{\theta} = \sqrt{\frac{3}{n}\sum_{i=1}^{n} x_i^2}$$

Hasta el momento sólo se ha trabajado con un parámetro desconocido, sin embargo la estimación por momentos para familias multiparamétricas también es posible. Dada una muestra aleatoria de tamaño n proveniente de la distribución $f(X;\Theta)$ con $\Theta = (\theta_1, \theta_2, \ldots, \theta_k)$. El estimador $\widehat{\Theta}$ debe satisfacer el sistema de ecuaciones:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} x_i &= \mathbb{E}_{\widehat{\Theta}}[X] \\
\frac{1}{n}\sum_{i=1}^{n} x_i^2 &= \mathbb{E}_{\widehat{\Theta}}\left[X^2\right] \\
&\vdots \\
\frac{1}{n}\sum_{i=1}^{n} x_i^k &= \mathbb{E}_{\widehat{\Theta}}\left[X^k\right].
\end{aligned}$$

Ejemplo **2.2.5.** Suponga que $x_1, x_2, \ldots, x_n$ es un conjunto de variables aleatorias independientes e identicamente distribuidas con distribución normal de media μ y varianza σ^2, $N(\mu, \sigma^2)$. Se desea encontrar el estimador por momentos para $\Theta = (\mu, \sigma^2)$. Si se considera a las funciones $g_1(x) = x$, $g_2(x) = x^2$ y dado que $\mathbb{E}(X) = \mu$, $\mathbb{E}(X^2) = \sigma^2 + \mu^2$. El vector $\widehat{\Theta} = (\mu, \sigma^2)$, debe cumplir

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \mathbb{E}_{\widehat{\Theta}}[X],$$
$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 = \mathbb{E}_{\widehat{\Theta}}[X^2].$$

De lo que se genera el sistema de ecuaciones

$$\bar{x} = \widehat{\mu} \tag{2.3}$$
$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 = \widehat{\sigma}^2 + \widehat{\mu}^2 \tag{2.4}$$

Al sustituir $\widehat{\mu} = \bar{x}$ en 2.4 se obtiene que

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n-1}{n}S^2$$

La expresión $\frac{n-1}{n}S^2$ se denotará por S_*^2. Por lo tanto, el estimador de Θ estará dado por:

$$\widehat{\Theta} = (\widehat{\mu}, \widehat{\sigma}^2) = \left(\bar{x}, S_*^2\right).$$

Ejemplo **2.2.6.** Una fábrica produce botones cuyo diámetro varía aleatoriamente entre dos valores a y b. Suponiendo que este valor se ajusta a una variable aleatoria con distribución uniforme, estime a partir de la muestra 10.20, 10.22, 10.10, 10.14 los parámetros a y b, con a<b.

Como los momentos poblacionales están dados por $\mathbb{E}[x_i] = \dfrac{a+b}{2}$ y $\mathbb{E}\left[x_i^2\right] = \dfrac{a^2+ab+b^2}{3}$ los estimadores $\widehat{a}$ y $\widehat{b}$ de los parámetros a y b son aquellos que cumplen,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} x_i &= \mathbb{E}_{\widehat{\Theta}}[X] \\ \frac{1}{n}\sum_{i=1}^{n} x_i^2 &= \mathbb{E}_{\widehat{\Theta}}\left[X^2\right]\end{aligned}$$

Además de la muestra se obtiene que $\frac{1}{n}\sum_{i=1}^{n} x_i = 10.165, \frac{1}{n}\sum_{i=1}^{n} x_i^2 = 103.3295$. Así el sistema anterior queda expresado de la siguiente manera

$$10.165 = \frac{\widehat{a}+\widehat{b}}{2} \text{ y } 103.3295 = \frac{\widehat{a}^2+\widehat{a}\widehat{b}+\widehat{b}^2}{3}.$$

Por lo tanto, los estimadores son $\widehat{a} = 10.2476$ y $\widehat{b} = 10.0824$.

2.3 Estimación máximo verosímil

El método de máxima verosimilitud fue utilizado y popularizado por Fisher en 1922, aunque ya se había utilizado en casos particulares por otros autores como Friedrich Gauss, Simon Laplace y Francis Edgeworth.

Dada $\mathbb{X} = (X_1, X_2, ..., X_n)$ una muestra aleatoria de una población con función de densidad $f(x, \Theta)$, $\Theta \in \mathbb{R}^k$, se define la *función de verosimilitud* $L(\Theta) : \mathbb{R}^k \to \mathbb{R}^+$

como

$$L(\Theta \mid \mathbb{X}) = \prod_{i=1}^{n} f(x_i \mid \Theta),$$

se buscará maximizar a $L(\Theta)$ y su óptimo será el estimador del parámetro desconocido.

Definición **2.3.1.** Sea $x_1, x_2, ..., x_n$ muestra aleatoria de $f(x,\Theta)$; $\Theta \in \mathbf{H}$ y $L(\Theta)$ la función de verosimilitud de la muestra aleatoria dada. Se define al estimador máximo verosímil (EMV) denotado por $\widehat{\Theta_{mv}}$ como aquel que satisface

$$L(\widehat{\Theta_{mv}} \mid x_1, x_2, \ldots, x_n) = \mathbf{Sup}_{\Theta \in \mathbf{H}} \left\{ L(\Theta \mid x_1, x_2, \ldots, x_n) \right\}.$$

Intuitivamente el problema consiste en que, dada la muestra aleatoria, se desea encontrar Θ de tal forma que sea más verosímil haber observado esos valores $X_1 = x_1, ..., X_n = x_n$ y ese valor de Θ será el EMV.

En particular, si $L(x_1, x_2, \ldots, x_n; \Theta_1) > L(x_1, x_2, \ldots, x_n; \Theta_2)$, entonces es más factible que Θ valga Θ_1 a Θ_2 por lo que, en este sentido, Θ_1 será un mejor estimador para Θ.

Ejemplo **2.3.1.** Una determinada empresa quiere planificar su producción. Se calcula que el producto que ofrecen puede gustar a un porcentaje de la población que oscila entre el 40%y 50%de los habitantes de su ciudad, pero tras tomar una muestra aleatoria de 10 individuos se observa que sólo 3 tienen interés por él. Teniendo esto en cuenta, ¿Cuál de las dos proporciones contempladas se deberá tomar en consideración?

El problema se puede modelar a través de una variable aleatoria Bernoulli de parámetro (p) con p desconocido. La función de verosimilitud es

$$L(p) = \prod_{i=1}^{n} f(x_1, x_2, \ldots, x_{10}; p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{10} x_i} \left(1-p\right)^{\left(10 - \sum_{i=1}^{n} x_i\right)},$$

como sólo hay tres interesados, entonces,

$$L(p) = p^3(1-p)^7,$$

Si $p = .4$ y $p = .5$ los valores de la función de verosimilitud son, respectivamente, $L(.4) = .001741$ y $L(.5) = .000976563$. Como $L(.4) > L(.5)$ se concluye que al 40 % de la población le gusta el producto.

Si $L(\theta)$ es diferenciable con respecto a θ_i, $i = 1, 2, \ldots, k$, los candidatos a ser EMV son los valores $\Theta = (\theta_1, \theta_2, ..., \theta_k)$ que satisfacen la ecuación

$$\frac{\partial}{\partial \theta_i} L(\Theta) = 0 \tag{2.5}$$

Recuerde que si Θ es solución de 2.5 no es suficiente el resultado obtenido a partir del criterio de la primera o segunda derivada, para garantizar que es un máximo hay que evaluar en los extremos del dominio de $L(\Theta)$ pues ahí es donde se puede encontrar el óptimo.

Para facilitar el trato algebraico con $L(\Theta)$ se suele aplicar logaritmo natural a $L(\Theta)$ y esta nueva aplicación de denota por $l(\theta) = log(L(\Theta))$ y se conoce como función de *log-verosimilitud*. Nótese que como el logaritmo natural es estrictamente creciente entonces $l(\Theta)$ alcanzará su máximo en el mismo punto que $L(\Theta)$.

Ejemplo **2.3.2.** Sea $x_1, x_2, ..., x_n$ una muestra aleatoria que proviene de una variable aleatoria con distribución binomial de parámetros (k, p) con k conocido. Hallar el EMV para p.

La función de densidad para cada observación es:

$$\mathbb{P}(X = x_i) = \binom{k}{x_i} p^{x_i} (1-p)^{k-x_i}, \qquad 0 \leq x_i \leq k.$$

Como las observaciones son independientes, la función de verosimilitud para la muestra con respecto a p es,

$$\begin{aligned} L(p) &= \prod_{i=1}^{n} f(x_i, p) \\ &= \prod_{i=1}^{n} \binom{k}{x_i} p^{x_i} (1-p)^{k-x_i} \\ &= p^{\sum_{i=1}^{n} x_i} \left(1-p\right)^{\left(nk - \sum_{i=1}^{n} x_i\right)} \prod_{i=1}^{n} \binom{k}{x_i}. \end{aligned}$$

Entonces,

$$log(L(p)) = \left(\sum_{i=1}^{n} x_i\right) log(p) + \left(nk - \sum_{i=1}^{n} x_i\right) log(1-p) + \sum_{i=1}^{n} log\binom{k}{x_i}$$

La derivada de esta función es

$$\frac{d(log(p))}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{nk - \sum_{i=1}^{n} x_i}{1-p}.$$

De la ecuación $\frac{d(log(p))}{dp} = 0$, se obtiene que el candidato a máximo es $\widehat{p} = \frac{\sum_{i=1}^{n} x_i}{nk} = \frac{\bar{x}}{k}$.

Como

$$\frac{d^2(log(p))}{dp^2} = -\frac{\sum_{i=1}^{n} x_i}{p^2} - \frac{nk \sum_{i=1}^{n} x_i}{1-p} < 0.$$

Entonces $\widehat{p} = \frac{\sum_{i=1}^{n} x_i}{nk} = \frac{\bar{x}}{k}$ es máximo.

Para garantizar que el máximo es global se deben evaluar los extremos del espacio paramétrico, $\mathbf{H} = [0,1]$, en la función de verosimilitud, de lo que se obtiene $L(0) = L(1) = 0$. Por lo tanto el EMV para el parámetro p es $\widehat{p} = \frac{\bar{x}}{k}$.

Ejemplo **2.3.3.** Sea $x_1, x_2, ..., x_n$ una muestra aleatoria que proviene de una distribución Uniforme continua en el intervalo $(0,\theta)$ con $\theta > 0$, la función de máxima verosimilitud para el parámetro θ es:

$$L(\theta) = \prod_{i=1}^{n} f(x_i, \theta) = \frac{1}{\theta} \left(\mathbb{1}_{(0,\theta)}(x_i)\right) = \frac{1}{\theta^n} \prod_{i=1}^{n} \left(\mathbb{1}_{(0,\theta)}(x_i)\right).$$

Tomando el logaritmo natural

$$\log(L(x_1,\ldots,x_n;\theta)) = -n\log\theta,$$

derivando respecto a θ e igualando con cero se tiene

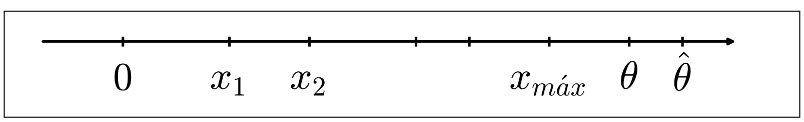
$$\frac{d}{d\theta}(-n\log\theta) = -\frac{n}{\theta} = 0.$$

la cual no tiene solución, nótese que $L(\theta)$ se maximiza cuando,

i) $\dfrac{1}{\theta^n}$ es máximo.

ii) $\displaystyle\prod_{i=1}^{n}\left(\mathbb{1}_{(0,\theta)}(x_i)\right) = 1.$

Como $\mathbb{1}_{(0,\theta)}(x_i) = 1$ $\forall i = 1,2,\ldots,n$ entonces $\widehat{\theta} \geq x_i$ $\forall i = 1,2,\ldots,n$ y la condición $i)$ se cumple si y sólo si θ es mínimo entonces se concluye que $\widehat{\theta} = \text{máx}(x_1, x_2, ..., x_n)$.



Ejemplo **2.3.4.** Sean $x_1, x_2, \ldots x_m$ muestra aleatoria $N(\mu,\sigma^2)$, ambos parámetros desconocidos.
Encontrar el EMV para $\theta = \left(\mu,\sigma^2\right) \in \mathbf{H} = \left\{\left(\mu,\sigma^2\right) \mid -\infty < \mu < \infty; 0 \leq \sigma^2 < \infty\right\}$.

La función $\mathbf{H} : \mathbb{R}^2 \to \mathbb{R}$, $\mathbf{H}$ tiene un máximo en (x_0, y_0) si

$$G = \begin{pmatrix} \dfrac{\partial^2\mathbf{H}}{\partial x^2} & \dfrac{\partial^2\mathbf{H}}{\partial x\partial y} \\ \dfrac{\partial^2\mathbf{H}}{\partial y\partial x} & \dfrac{\partial^2\mathbf{H}}{\partial y^2} \end{pmatrix}$$

i) $\left.\dfrac{\partial \mathbf{H}}{\partial x}\right|_{x_0} = 0$ y $\left.\dfrac{\partial \mathbf{H}}{\partial y}\right|_{y_0} = 0$

ii) $\left.\dfrac{\partial^2 \mathbf{H}}{\partial x^2}\right|_{(x_0,y_0)} < 0$

iii) $\det|G|_{(x_0,y_0)} > 0$

La función de verosimilitud es

$$L\left(\mu,\sigma^2\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{x_i-\mu}{2\sigma^2}\right)} \mathbb{1}_{(-\infty,\infty)}(x) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)} \prod_{i=1}^{n} \mathbb{1}_{(-\infty,\infty)}(x_i)$$

$$\Rightarrow l(\theta) = \log\left[L\left(\mu,\sigma^2\right)\right] = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 + \log\left(\prod_{i=1}^{n}\mathbb{1}_{(-\infty,\infty)}(x)\right).$$

Entonces las derivadas parciales de la función de log-verosimilitud con respecto a μ y σ^2 son

$$\frac{\partial l(\theta)}{\partial \mu} = \frac{1}{2\sigma^2}\sum_{i=1}^{n} 2\,(x_i-\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) \tag{2.6}$$

$$\frac{\partial l(\theta)}{\partial \sigma^2} = -\frac{n}{2}\left(\frac{2\pi}{2\pi\sigma^2}\right) + \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\,\frac{1}{(\sigma^2)^2} \qquad = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 \tag{2.7}$$

Igualando a cero 2.6 y 2.7

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0$$

$$-\frac{n}{2\sigma}^2 + \frac{1}{2\sigma}^4 \sum_{i=1}^{n}(x_i-\mu)^2 = 0.$$

De la primera, ya que $\sigma^2 > 0$ se obtiene $\sum_{i=1}^{n}(x_i - \mu) = \sum_{i=1}^{n} x_i - n\mu = 0$, por lo tanto

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \tag{2.8}$$

Sustituyendo 2.8 en 2.7

$$2\sigma^2\left[-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \bar{x})^2\right] = 0$$

$$\Rightarrow \left[-n + \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0\right]\sigma^2$$

$$\Rightarrow -n\sigma^2 + \sum_{i=1}^{n}(x_i - \bar{x})^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n}(x_i - \bar{x})^2 = n\sigma^2$$

$$\Rightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = s_*^2.$$

Por lo tanto $l\left(\mu, \sigma^2\right)$ tiene un punto crítico en $\left(\bar{x}, s_*^2\right)$.

Por demostrar que es máximo se debe verificar $ii)\frac{\partial^2 l\left(\mu,\sigma^2\right)}{\partial \mu^2} < 0$

$$\frac{\partial^2 l\left(\mu,\sigma^2\right)}{\partial \mu^2} = \frac{\partial}{\partial \mu}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)\right) = \frac{\partial}{\partial \mu}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n} x_i - \frac{n\mu}{\sigma^2}\right) = -\frac{n}{\sigma^2} < 0$$

$iii)$ Por demostrar $\det|G| > 0$

$$\left|\begin{pmatrix} \dfrac{\partial^2 l(\theta)}{\partial \mu^2} & \dfrac{\partial^2 l(\theta)}{\partial \mu \partial \sigma^2} \\ \dfrac{\partial^2 l(\theta)}{\partial \sigma^2 \partial \mu} & \dfrac{\partial^2 l(\theta)}{\partial (\sigma^2)^2} \end{pmatrix}\right|_{\mu=\bar{x},\sigma^2=s_*^2} > 0$$

Falta calcular las otras 3 parciales:

$$\frac{\partial^2 l(\theta)}{\partial \mu \partial \sigma^2} = \frac{\partial}{\partial \mu}\left[-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2\right] = \frac{1}{2\sigma^4}\sum_{i=1}^{n} 2\,(x_i-\mu)\,(-1) = -\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu).$$

$$\frac{\partial^2 l(\theta)}{\partial \sigma^2 \partial \mu} = \frac{\partial}{\partial \sigma^2}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)\right] = \sum_{i=1}^{n} x_i(-1)\left(\frac{1}{\sigma^4}\right) + \frac{n\mu}{\sigma^4}.$$

$$\frac{\partial^2 l(\theta)}{\partial (\sigma^2)^2} = \frac{\partial}{\partial \sigma^2}\left[-\frac{n}{2\sigma^2} + \sum_{i=1}^{n}(x_i-\mu)^2\left(\frac{1}{2\sigma^4}\right)\right] = \frac{n}{2\sigma^4} + \sum_{i=1}^{n}(x_i-\mu)^2\left(-\frac{1}{(\sigma^2)^3}\right) =$$

$$\frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Ahora,

$$\det|G| = \left|\begin{matrix} -\dfrac{n}{\sigma^2} & -\dfrac{1}{\sigma^4}\displaystyle\sum_{i=1}^{n}(x_i-\mu) \\ \displaystyle\sum_{i=1}^{n} x_i\left(-\frac{1}{\sigma^4}\right) + \frac{n\mu}{\sigma^4} & \dfrac{n}{2\sigma^4} - \displaystyle\sum_{i=1}^{n}(x_i-\mu)^2\left(\frac{1}{\sigma^6}\right) \end{matrix}\right|_{\mu=\bar{x},\sigma^2=s_*^2}$$

$$= -\frac{n}{s_*^2}\left[\frac{n}{2s_*^4} - \frac{\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2}{s_*^6}\right] + \left[\left(-\frac{1}{\sigma^4}\right)\left(\sum_{i=1}^{n}(x_i-\mu)\right)\right]^2 =$$

$$-\frac{n}{s_*^2}\left[\frac{n}{2s_*^4} - \frac{n}{s_*^6}\frac{\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2}{n}\right] = -\frac{n}{s_*^2}\left[\frac{n}{2s_*^4} - \frac{n}{s_*^4}\right] = \frac{n^2}{2s_*^6} > 0.$$

Por lo tanto el EMV de $\theta = \left(\mu, \sigma^2\right)$ es $\hat{\theta}_{MV} = \left(\bar{x}, s_*^2\right)$.

Ejemplo **2.3.5.** Sea $x_1, x_2, ..., x_n$ una muestra aleatoria de una población que se distribuye como una $\Gamma(p,a)$, con ambos parámetros desconocidos. Obtener los EMV.

La función de verosimilitud es:

$$
\begin{aligned}
L(x_1, x_2, \ldots, x_n; p, a) &= \prod_{i=1}^{n} f(x_i, p, a) \\
&= \prod_{i=1}^{n} \frac{a^p}{\Gamma(p)} x_i^{p-1} e^{-ax_i} \\
&= \frac{a^{np}}{(\Gamma(p))^n} e^{\left(-a \sum_{i=1}^{n} x_i\right)} \prod_{i=1}^{n} x_i^{p-1}.
\end{aligned}
$$

Tomando el logaritmo natural de la función de verosimilitud

$$\log(L) = np \log a - n \log \Gamma(p) + (p-1) \sum_{i=1}^{n} \log x_i - a \sum_{i=1}^{n} x_i.$$

Derivando respecto de p y de a e igualando a cero, se obtienen las ecuaciones de verosimilitud

$$\frac{\partial \log(L}{\partial p} = n \log a - \frac{n}{\Gamma(p)} \frac{\partial \Gamma(p)}{\partial p} + \sum_{i=1}^{n} x_i = 0$$

$$\frac{\partial \log(L}{\partial a} = \frac{np}{a} - \sum_{i=1}^{n} x_i = 0.$$

Obteniendo el parámetro $\widehat{a}$ de la segunda ecuación se tiene

$$\frac{np}{a} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \widehat{a} = \frac{np}{\sum_{i=1}^{n} x_i} = \frac{\widehat{p}}{\bar{x}}.$$

Sustituyendo en la primera ecuación

$$n \log \frac{p}{\bar{x}} - \frac{n}{\Gamma(p)} \frac{\partial \Gamma(p)}{\partial p} + \sum_{i=1}^{n} x_i = 0$$

o bien

$$n \log \frac{p}{\bar{x}} - n \frac{\Gamma'(p)}{\Gamma(p)} + \sum_{i=1}^{n} x_i = 0.$$

La solución de esta ecuación se debe obtener de manera aproximada por métodos numéricos, una vez obtenido el estimador para p, el estimador para a se obtiene de manera inmediata.

Si el espacio paramétrico **H** es discreto, el método de maximización usando derivadas no garantiza poder encontrar el valor máximo, debido a que el valor obtenido no esté en el espacio paramétrico.

Ejemplo **2.3.6.** Se lanza una moneda equilibrada k veces. La muestra $x_1, x_2, \ldots, x_n$ proviene de una variable aleatoria binomial de parámetros (k, p) que cuenta el número de soles en k lanzamientos. Se busca determinar el número de veces, k, que es lanzada la moneda.

La función de verosimilitud es

$$\begin{aligned} L(k) &= \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} \binom{k}{x_i} p^{x_i} (1-p)^{(k-x_i)} \\ &= p^{\sum\limits_{i=1}^{n} x_i} \left(1-p\right)^{\left(nk - \sum\limits_{i=1}^{n} x_i\right)} \prod_{i=1}^{n} \binom{k}{x_i} \\ &= \frac{p^{\sum\limits_{i=1}^{n} x_i} \left(1-p\right)^{nk}}{\left(1-p\right)^{\sum\limits_{i=1}^{n} x_i}} \prod_{i=1}^{n} \binom{k}{x_i}. \end{aligned}$$

A pesar de que la función de verosimilitud es igual a la del ejercicio 2.3.1, en este caso resulta complejo la maximización debido a que ahora el parámetro a estimar tiene como posibilidades un valor entero positivo. Una manera alternativa de obtener el máximo en esta situación es encontrar el valor de k para el cual la función pasa de creciente a decreciente para ello se analizará el cociente

$$\frac{L(k)}{L(k-1)} = \frac{\left(\dfrac{p^{\sum_{i=1}^{n} x_i}\left(1-p\right)^{nk}}{\left(1-p\right)^{\sum_{i=1}^{n} x_i}} \prod_{i=1}^{n} \binom{k}{x_i}\right)}{\left(\dfrac{p^{\sum_{i=1}^{n} x_i}\left(1-p\right)^{n(k-1)}}{\left(1-p\right)^{\sum_{i=1}^{n} x_i}} \prod_{i=1}^{n} \binom{k-1}{x_i}\right)} = \frac{p^{\sum_{i=1}^{n} x_i}\left(1-p\right)^{nk}\left(1-p\right)^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \binom{k}{x_i}}{p^{\sum_{i=1}^{n} x_i}\left(1-p\right)^{n(k-1)}\left(1-p\right)^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \binom{k-1}{x_i}}$$

$$= \frac{\left(1-p\right)^{nk} \prod_{i=1}^{n} \binom{k}{x_i}}{\left(1-p\right)^{n(k-1)} \prod_{i=1}^{n} \binom{k-1}{x_i}} = \frac{\left(1-p\right)^{n} \prod_{i=1}^{n} \dfrac{k!}{(k-x_i)!x_i!}}{\prod_{i=1}^{n} \dfrac{(k-1)!}{(k-1-x_i)!x_i!}} = \left(1-p\right)^{n} \prod_{i=1}^{n} \frac{k}{k-x_i} = \frac{\left(1-p\right)^{n} k^n}{\prod_{i=1}^{n}(k-x_i)}.$$

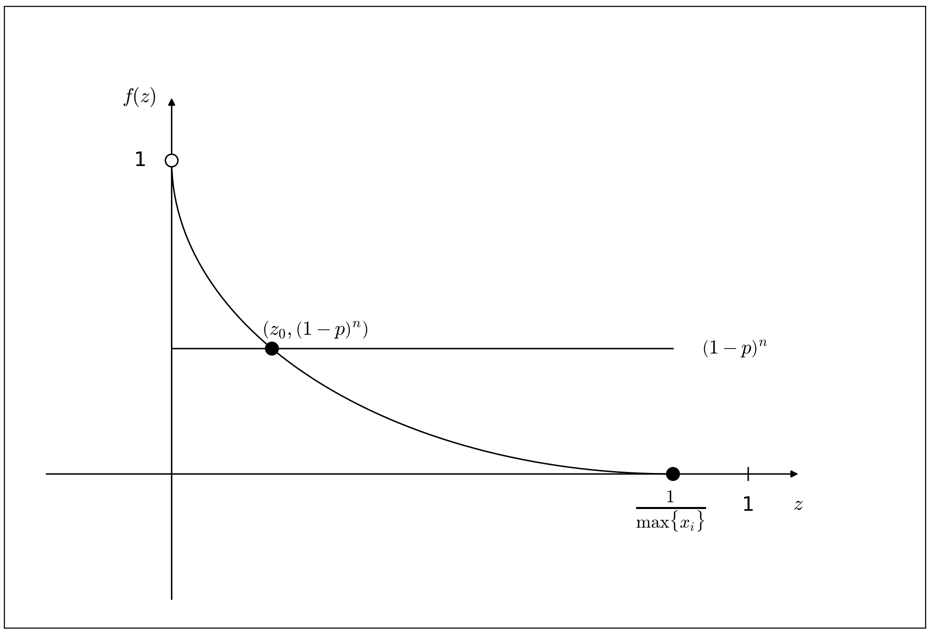
La función $L(k)$ es creciente cuando se cumple la desigualdad:

$$\frac{L(k)}{L(k-1)} = \frac{\left(1-p\right)^{n} k^n}{\prod_{i=1}^{n}(k-x_i)} \geq 1 \Leftrightarrow \left(1-p\right)^{n} \geq \prod_{i=1}^{n}\left(1-\frac{x_i}{k}\right).$$

Y es decreciente cuando

$$\frac{L(k)}{L(k-1)} = \frac{\left(1-p\right)^{n} k^n}{\prod_{i=1}^{n}(k-x_i)} \leq 1 \Leftrightarrow \left(1-p\right)^{n} \leq \prod_{i=1}^{n}\left(1-\frac{x_i}{k}\right).$$

Considérese la función continua $f(z) = \prod_{i=1}^{n}(1 - x_i z)$ con $z = \frac{1}{k}$. Como $0 < p < 1$ entonces $0 < \left(1-p\right)^n < 1$; $f(z)$ es una función decreciente, con dominio $\left(0, \frac{1}{\text{máx}\{x_i\}}\right)$, $\lim_{z \to 0^+} f(z) = 1$ y $\lim_{z \to \frac{1}{\text{máx}\{x_i\}}} f(z) = 0$, por lo tanto existe una única z_0 tal que $\left(1-p\right)^n = \prod_{i=1}^{n}(1 - x_i z_0)$.



De lo anterior se tiene que

$$\begin{cases} \dfrac{\left(1-p\right)^n}{\prod_{i=1}^{n}(1 - x_i z_0)} \leq 1 & \text{si } z \in (0, z_0). \\ \dfrac{\left(1-p\right)^n}{\prod_{i=1}^{n}(1 - x_i z_0)} \geq 1 & \text{si } z \notin (0, z_0). \end{cases}$$

Es decir

$$\begin{cases} \dfrac{\left(1-p\right)^n}{\displaystyle\prod_{i=1}^{n}\left(1-\dfrac{x_i}{k}\right)} \leq 1 & \text{si } k \in \left(\frac{1}{z_0}, \infty\right). \\ \dfrac{\left(1-p\right)^n}{\displaystyle\prod_{i=1}^{n}\left(1-\dfrac{x_i}{k}\right)} \geq 1 & \text{si } k \in \left(\text{máx}\{x_i\}, \frac{1}{z_0}\right). \end{cases}$$

Por lo que $L(k)$ es creciente si $k \in \left(\text{máx}\{x_i\}, \frac{1}{z_0}\right)$ y decreciente si $k \in \left(\frac{1}{z_0}, \infty\right)$, entonces existe una k para la cual $L(k)$ tiene un máximo y se alcanza en $\left\lfloor \frac{1}{z_0} \right\rfloor$, o bien $\left\lfloor \frac{1}{z_0} \right\rfloor + 1$.

Ejemplo **2.3.7.** Dada una muestra aleatoria $x_1, x_2, \ldots, x_n$ de una población con función de densidad

$$f(x;\theta) = \begin{cases} (1+\theta)x^{\theta} & 0 < x < 1 \\ 0 & \text{en otro caso.} \end{cases}$$

Calcule el estimador del parámetro θ por el método de momentos y el de máxima verosimilitud.

a) Por el método de momentos, para $g(x) = x$. Primero se encontrara el primer momento poblacional

$$\begin{aligned} &\mathbb{E}[X] \\ &\quad = \int_0^1 (1+\theta)x^{1+\theta}dx \\ &\quad = \frac{1+\theta}{2+\theta}. \end{aligned}$$

E igualando a su respectivo momento muestral se obtiene $\bar{x} = \dfrac{1+\widehat{\theta}}{2+\widehat{\theta}}$, lo cual implica

$$
\begin{aligned}
1+\widehat{\theta} &= 2\bar{x} + \bar{x}\widehat{\theta} \\
\widehat{\theta} - \bar{x}\widehat{\theta} &= 2\bar{x} - 1 \\
\widehat{\theta}(1-\bar{x}) &= 2\bar{x} - 1
\end{aligned}
$$

Por lo tanto, $\widehat{\theta} = \dfrac{1-2\bar{x}}{\bar{x}-1}$.

b) Por el método de máxima verosimilitud, el logaritmo natural de $L(\theta)$ de la función de densidad conjunta para la muestra es

$$
\begin{aligned}
\log(L) &= \log \prod_{i=1}^{n} f(x_i, \theta) \\
&= \log\left((1+\theta)^n \left(\prod_{i=1}^{n} x_i\right)^{\theta}\right) \\
&= n\log(1+\theta) + \theta \sum_{i=1}^{n} \log x_i
\end{aligned}
$$

Derivando respecto a θ

$$
\frac{d\log(L)}{d\theta} = \frac{n}{1+\theta} + \sum_{i=1}^{n} log(x_i),
$$

Igualando a cero

$$\frac{n}{1+\theta} + \sum_{i=1}^{n} log(x_i) = 0$$

$$\frac{n}{1+\widehat{\theta}} = -\sum_{i=1}^{n} log(x_i)$$

$$n = -\sum_{i=1}^{n} log(x_i) - \widehat{\theta}\sum_{i=1}^{n} log(x_i)$$

$$n + \sum_{i=1}^{n} log(x_i) = -\widehat{\theta}\sum_{i=1}^{n} log(x_i)$$

$$\widehat{\theta} = \frac{n + \sum_{i=1}^{n} log(x_i)}{-\sum_{i=1}^{n} log(x_i)}$$

por lo tanto, el candidato a EMV es

$$\theta = \frac{n}{\left(-\sum_{i=1}^{n} \log x_i\right)} - 1.$$

Por otro lado

$$\frac{d^2 \log(L)}{d\theta^2} = -\frac{n}{(1+\theta)^2} < 0.$$

Por lo que $\widehat{\theta} = -\left(\frac{n}{\left(\sum_{i=1}^{n} \log x_i\right)} + 1\right)$.

Es importante mencionar el hecho de que puede suceder que la función de verosimilitud tenga más de una solución. Sin embargo, en este texto no se trabajara con esos casos.

2.4 Criterios de evaluación de estimadores

En el capítulo anterior se estudiaron dos métodos para encontrar estimadores del parámetro de interés. En este capítulo se estudiaran algunas cualidades que se deben tomar en cuenta para la elección de un estimador. Entre las que se verán están insesgabilidad, eficiencia y consistencia, esta última una propiedad asintótica de los estimadores.

Insesgabilidad

Una característica deseable de los estimadores es que generen estimados lo más ‶parecidos″al valor real del parámetro. sin embargo, ¿Cuándo saber que la estimación está lo suficientemente cerca del verdadero valor?. Parece razonable pensar que, si se utiliza un estimador $T(\mathbb{X})$ varias veces en la misma situación y en distintas muestras tomadas de la misma población, al menos, en promedio $T(\mathbb{X})$ sea el valor que se busca estimar. De hecho, si la distribución de $T(\mathbb{X})$ se centra en el verdadero valor de la función del parámetro se dice que $T(\mathbb{X})$ es insesgado.

Definición **2.4.1.** Sea $\mathbb{X}$ una muestra aleatoria de $f(X;\Theta)$ con $T(\mathbb{X})$ el estimador puntual del parámetro Θ. Se dice que $T(\mathbb{X})$ es un *estimador insesgado* para Θ si y sólo si $\forall\Theta \in \mathbf{H}\ \ \mathbb{E}(T(\mathbb{X})) = \Theta$.

Ejemplo **2.4.1.** Sea $X_1, X_2, ..., X_n$ una muestra aleatoria que proviene de una distribución Bernoulli de parámetro p. Por demostrar que el estimador por momentos es insesgado.

Sea $g(x) = x$, como $\mathbb{E}[X] = p$, entonces, el estimador por momentos es

$$\bar{x} = \widehat{p} \tag{2.9}$$

Por otro lado, por 2.9 y propiedades de la esperanza,

$$\begin{aligned}
\mathbb{E}(\widehat{p}) &= \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(X_i) \\
&= \frac{1}{n} (np) = p.
\end{aligned}$$

Como $\mathbb{E}(\widehat{p}) = p$ entonces $\bar{x} = \widehat{p}$ es un estimador insesgado para p.

De hecho

Proposición 2.4.1. *Sea $X_1, X_2, \ldots, X_n$ muestra aleatoria de una distribución $f(X)$ con media μ. Entonces, la media muestral, $\bar{x}$, es un estimador insesgado para μ.*

Demostración. La media muestral se define como $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$. Luego,

$$\mathbb{E}(\bar{x}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(x_i) = \frac{1}{n}\, n\, \mu = \mu.$$

Por lo tanto $\bar{x}$ es un estimador insesgado para μ. □

Ejemplo 2.4.2. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población con distribución Uniforme $(0, \theta)$ con $\theta > 0$. Compruébese si los estimadores por momentos y máxima verosimilitud son o no insesgados y, de no serlo, proponer un estimador insesgado.

Como el primer momento poblacional de una variable aleatoria Uniforme $(0,\theta)$ es $\theta/2$ entonces el estimador por momentos, con $g(x)=x$, es $\widehat{\Theta}=2\bar{x}$ y se denotará por $T_1(\mathbb{X})=2\bar{x}$.

Por otra parte, en el Ejemplo 2.3.3 se probó que el EMV para esta distribución es $T_2(\mathbb{X})=X_{(n)}=\text{máx}\{X_1,X_2,\ldots,X_n\}$.

Para $T_1(\mathbb{X})=2\bar{x}$,

$$\mathbb{E}[2\bar{x}]=\mathbb{E}\left[2\,\frac{1}{n}\,\sum_{i=1}^{n}X_i\right]=\frac{2}{n}\,\sum_{i=1}^{n}\mathbb{E}[X_i]=\frac{2}{n}\,\sum_{i=1}^{n}\frac{\theta}{2}=\theta.$$

Por lo tanto, $T_1(\mathbb{X})=2\bar{x}$ es insesgado para θ.

Para $T_2(\mathbb{X})=X_{(n)}$

Nótese, primero, que como la función de densidad y distribución de X son $f_X\,(x)=\frac{1}{\theta}$ y $F_X(x)=\frac{x}{\theta}\,\mathbb{1}_{(0,\theta)}\,(x)$ respectivamente, entonces la función de densidad de $X_{(n)}$ está dada por

$$f_{X_{(n)}}(x)=n\,\left(F_X\,(x)\right)^{n-1}f_x\,(x)\mathbb{1}_{(0,\theta)}\,(x)$$

entonces

$$f_{X_{(n)}}(x)=\frac{n\,x^{n-1}}{\theta^n}\,\mathbb{1}_{(0,\theta)}(x).$$

Por lo cual,

$$\begin{aligned}\mathbb{E}\left[x_{(n)}\right]&=\int_0^\theta x\cdot f_{x_{(n)}}\,(x)\,d_x=\int_0^\theta x\cdot\frac{n\,x^{n-1}}{\theta^n}\,d_x\\&=\frac{n}{\theta^n}\,\int_0^\theta x^n\,d_x=\frac{n}{\theta^n}\cdot\frac{x^{n+1}}{n+1}\,\bigg|_0^\theta=\frac{n}{\theta^n}\left(\frac{\theta^{n+1}}{n+1}\right)=\frac{n\theta}{n+1}.\end{aligned}$$

Como $\mathbb{E}\left[x_{(n)}\right] = \dfrac{n\,\theta}{n+1} \neq \theta$ entonces $X_{(n)}$ no es insesgado. Despejando a θ de la última igualdad se tiene que $\dfrac{n+1}{n}\,\mathbb{E}\left[X_{(n)}\right] = \theta$ y por la linealidad del operador $E(\cdot)$ se observa que $\mathbb{E}\left[\dfrac{n+1}{n} X_{(n)}\right] = \theta$ lo cual permite considerar a la estadística $T_3(\mathbb{X}) = \dfrac{n+1}{n} X_{(n)}$ como un nuevo estimador insesgado.

Una nueva estadística, $T_3(\mathbb{X})$, función de $x_{(n)}$ que sí es insesgado para $q(\theta) = \theta$ es

$$T_3(\mathbb{X}) = \frac{n+1}{n}\,X_{(n)}.$$

Ejemplo **2.4.3.** Sea $X_1, X_2, \ldots, X_n$ muestra aleatoria de una población con distribución Uniforme $(0, \theta)$ con $\theta > 0$. Crear un estimador que sea función del estadístico de orden $X_{(1)} = \text{mín}\{X_1, X_2, \ldots, X_n\}$ que sea insesgado para θ.

Se sabe que la función de densidad para el estadístico de orden $X_(1)$ es $f_{x_{(1)}}(x) = n\ (1 - F_X(x))^{n-1}\ f_X(x)$.

Entonces, procediendo de manera análoga al ejercicio anterior

$$\mathbb{E}\left[x_{(1)}\right] = \int_0^{\theta} x\,n \left(1 - \frac{x}{\theta}\right)^{n-1} \frac{1}{\theta}\,d_x = \frac{n}{\theta^n}\int_0^{\theta} x\,(\theta - x)^{n-1}\,d_x.$$

Al realizar el cambio de variable: $u = \theta - x$ entonces $du = -dx$, $x = \theta - u$ y la integral anterior se transforma en

$$
\begin{aligned}
\frac{n}{\theta^n}\int_\theta^0 (\theta-u)\,(u)^{n-1}\,(-d_u) &= \frac{n}{\theta^n}\int_0^\theta (\theta\, u^{n-1}-u^n)\,d_u \\
&= \frac{n}{\theta^n}\left(\frac{\theta u^n}{n}-\frac{u^{n+1}}{n+1}\Bigg|_0^\theta\right) \\
&= \frac{n}{\theta^n}\left[\frac{\theta\,\theta^n}{n}-\frac{\theta n+1}{n+1}\right] \\
&= \frac{n\,\theta n+1}{n\,\theta^n}-\frac{n\,\theta n+1}{(n+1)\theta^n} \\
&= \theta-\frac{n\theta}{n+1} \\
&= \frac{\theta+n\,\theta-n\,\theta}{n+1} \\
&= \frac{\theta}{n+1}
\end{aligned}
$$

.

Por lo tanto, $\mathbb{E}\left[x_{(1)}\right]=\frac{\theta}{n+1}$. Con esto último se concluye que el estimador insesgado en función de $x_{(1)}$ es $T_4(\mathbb{X})=(n+1)x_{(1)}$.

Otro criterio que ayuda a medir la bondad de los estimadores para θ está basado en:

Definición **2.4.2.** (*Error Cuadrático Medio*) Sea $\mathbb{X}$ una muestra aleatoria de $f(X,\theta)$ y sea $T(\mathbb{X})$ el estimador de θ. Se define el *Error Cuadrático Medio (ECM)* de $T(\mathbb{X})$ para θ como

$$ECM(T(\mathbb{X}))=\mathbb{E}\,(T(\mathbb{X})-\theta)^2.$$

Y a la diferencia $b_t=\mathbb{E}(T)-\theta$ se le llama sesgo de T.

El ECM es la media de que tan lejos se encuentran los valores que se estiman para θ del valor del parámetro. De hecho, cualquier función no decreciente y con valores en $\mathbb{R}^+$ como la distancia absoluta $|\, T(\mathbb{X}) - \theta\,|$ sirve para medir la calidad de cualquier estimador, sin embargo, se deben notar dos ventajas importantes del ECM: es una función derivable en todo punto, tratable algebraicamente; y se puede escribir como función de la varianza y el sesgo del estimador, es decir,

$$ECM(T(\mathbb{X})) = Var(T(\mathbb{X}) + b_t^2,$$

En efecto,

$$\begin{aligned}
ECM_q(T(\mathbb{X})) &= \mathbb{E}(\Big((T(\mathbb{X}) - q(\theta))^2\Big) \\
&= \mathbb{E}\Big(T^2(\mathbb{X}) - 2\;T(\mathbb{X})\;q(\theta) + q^2(\theta)\Big) \\
&= \mathbb{E}\Big(T^2(\mathbb{X})\Big) - 2\;q(\theta)\;\mathbb{E}\,(T(\mathbb{X})) + q^2(\theta) \\
&= \mathbb{E}\Big(T^2(\mathbb{X})\Big) - \mathbb{E}^2\,(T(\mathbb{X})) + \mathbb{E}^2\,(T(\mathbb{X})) - 2\;q(\theta)\;\mathbb{E}\,(T(\mathbb{X})) + q^2(\theta) \\
&= Var\,(T(\mathbb{X})) + b_t^2
\end{aligned}$$

Con esto note que si el estimador es insesgado el ECM coincide con su varianza.

El ECM es una medida de dispersión y precisión ya que, intuitivamente, el sesgo es, en promedio, que ‶tanto″se acerca el estimador al valor del parámetro verdadero y la varianza es un indicador de cuanto se‶pueden″alejar los valores del estimador de su media. Considerado lo anterior cabe esperar que cuanto menor sea el ECM mejor será el estimador.

Ejemplo **2.4.4.**

En los Ejemplos 2.4.2 y 2.4.3 se calcularon cuatro estimadores para el parámetro θ de una función de distribución Uniforme $(0,\theta)$, tres de ellos son insesgados y uno es sesgado. A continuación se calcula el ECM de dichos estimadores, con el objetivo de mostrar que para elegir al mejor estimador hay que considerar varias características.

Para el estimador $T_1(\mathbb{X}) = 2\bar{x}$

$$\begin{aligned} ECM_\theta(T_1(\mathbb{X})) = Var(2\bar{x}) = 4Var\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) &= \frac{4}{n^2}\sum_{i=1}^{n} Var(x_i) \\ = \frac{4}{n^2}\sum_{i=1}^{n}\frac{(\theta-0)^2}{12} &= \frac{4}{n^2}\cdot\frac{n\theta^2}{12}. \end{aligned}$$

Por lo tanto, $ECM_\theta(T_1(\mathbb{X})) = \dfrac{\theta^2}{3n}$.

Para $T_2(\mathbb{X}) = x_{(n)}$ primero se calculará el sesgo,

$$b_t^2 = \left(\mathbb{E}\left[x_{(n)}\right]-\theta\right)^2 = \left(\frac{n\theta}{n+1}-\theta\right)^2 = \theta^2\left(\frac{n-n-1}{n+1}\right)^2 = \frac{\theta^2}{(n+1)^2}.$$

Luego para obtener la varianza de $X_{(n)}$ se necesita,

$$\begin{aligned} \mathbb{E}\left[x_{(n)}^2\right] = \int_0^\theta x^2 f_{x_{(n)}}(x)d_x &= \int_0^\theta x^2\cdot n\left(\frac{x}{\theta}\right)^{n-1}\cdot\frac{1}{\theta}d_x \\ = \frac{n}{\theta^n}\int_0^\theta x^{n+1}d_x &= \frac{n}{\theta^n}\left.\frac{x^{n+2}}{n+2}\right|_0^\theta = \frac{n\theta^2}{n+2}. \end{aligned}$$

Por otro lado, $Var(x_{(n)}) = \mathbb{E}\left[x_{(n)}^2\right] - \mathbb{E}^2\left[x_{(n)}\right]$, entonces

$$Var(x_{(n)}) = \frac{n\theta^2}{n+2} - \left(\frac{n\theta}{n+1}\right)^2 = n\theta^2\left(\frac{1}{n+2} - \frac{n}{(n+1)^2}\right) = \frac{n\theta^2}{(n+2)(n+1)^2}.$$

Y así,

$$
\begin{aligned}
ECM_\theta(T_2(x_{(n)})) &= \frac{n\theta^2}{(n+2)(n+1)^2} + \frac{\theta^2}{(n+1)^2} \\
&= \frac{\theta^2}{(n+1)^2}\left(\frac{n}{n+2}+1\right) = \frac{2\theta^2(n+1)}{(n+1)^2(n+2)} = \frac{2\theta^2}{(n+1)(n+2)}
\end{aligned}
$$

Por lo tanto, $ECM_\theta\left(T_2(X_{(n)})\right) = \dfrac{2\theta^2}{(n+1)(n+2)}$.

Para $T_3(\mathbb{X}) = \dfrac{n+1}{n}X_{(n)}$ que fue insesgado, se tiene

$$
\begin{aligned}
ECM_\theta(T_3(\mathbb{X})) &= Var\left(\frac{n+1}{n}X_{(n)}\right) = \frac{(n+1)^2}{n^2}Var(X_{(n)}) \\
&= \frac{(n+1)^2}{n^2}\frac{n\theta^2}{(n+2)(n+1)^2} \\
&= \frac{\theta^2}{n(n+2)}.
\end{aligned}
$$

Por lo tanto $ECM_\theta(T_3(\mathbb{X})) = \dfrac{\theta^2}{n(n+2)}$.

Por último, para $T_4(\mathbb{X}) = (n+1)X_{(1)}$, primero hay que calcular $\mathbb{E}\left[X_{(1)}^2\right]$

$$
\begin{aligned}
\mathbb{E}\left[x_{(1)}^2\right] &= \int_0^\theta x^2 n\left(1-\frac{x}{\theta}\right)^{n-1}\cdot\frac{1}{\theta}dx \\
&= \frac{n}{\theta^n}\int_0^\theta x^2(\theta-x)^{n-1}dx.
\end{aligned}
$$

Si $u = x^2$ y $dv = (\theta-x)^{n-1}dx$, entonces usando integración por partes lo anterior equivale a:

$$
\begin{aligned}
\frac{n}{\theta^n}\int_0^\theta x^2(\theta-x)^{n-1}d_x &= \frac{n}{\theta^n}\left(\left.\frac{-x^2(\theta-x)^n}{n}\right|_0^\theta + \frac{2}{n}\int_0^\theta x(\theta-x)^n dx\right) \\
&= \frac{2}{\theta^n}\int_0^\theta x(\theta-x)^n dx.
\end{aligned}
$$

Nuevamente usando integración por partes con $u = x$ y $dv = (\theta - x)^n dx$;

$$\begin{aligned}
\frac{2}{\theta^n}\int_0^\theta x(\theta-x)^n d_x &= \frac{2}{\theta^n}\left(\left.\frac{-x(\theta-x)^{n+1}}{n+1}\right|_0^\theta + \frac{1}{n+1}\int_0^\theta (\theta-x)^{n+1} d_x\right)\\
&= \frac{2}{\theta^n}\left(\frac{1}{n+1}\int_0^\theta (\theta-x)^{n+1} d_x\right)\\
&= \frac{2}{\theta^n(n+1)}\left(\left.-\frac{(\theta-x)^{n+2}}{n+2}\right|_0^\theta\right)\\
&= \frac{2\theta^2}{(n+1)(n+2)}.
\end{aligned}$$

Por último,

$$\begin{aligned}
ECM_\theta\left(T_4(\mathbb{X})\right) &= Var((n+1)x_{(1)}) = (n+1)^2 Var(x_{(1)})\\
&= (n+1)^2\left(\frac{2\theta^2}{(n+1)(n+2)} - \frac{\theta^2}{(n+1)^2}\right)\\
&= \theta^2\left(\frac{2(n+1)}{n+2} - 1\right) = \frac{2n+2-(n+2)}{n+2}\theta^2 = \frac{n\theta^2}{n+2}
\end{aligned}$$

Por lo tanto $ECM_\theta(T_4(\mathbb{X})) = \dfrac{n\theta^2}{n+2}$.

Resumiendo la información anterior

$T(\mathbb{X})$	ECM	n=1	n=2	n=3	n=4	$n \to \infty$	Insesgado
$T_1(\mathbb{X}) = 2\bar{x}$	$\frac{\theta^2}{3n}$	$\frac{\theta^2}{3}$	$\frac{\theta^2}{6}$	$\frac{\theta^2}{9}$	$\frac{\theta^2}{12}$	0	si
$T_2(\mathbb{X}) = x_{(n)}$	$\frac{2\theta^2}{(n+1)(n+2)}$	$\frac{\theta^2}{3}$	$\frac{\theta^2}{6}$	$\frac{\theta^2}{10}$	$\frac{\theta^2}{15}$	0	no
$T_3(\mathbb{X}) = \frac{n+1}{n}x_{(n)}$	$\frac{\theta^2}{n(n+2)}$	$\frac{\theta^2}{3}$	$\frac{\theta^2}{8}$	$\frac{\theta^2}{15}$	$\frac{\theta^2}{24}$	0	si
$T_4(\mathbb{X}) = (n+1)x_{(1)}$	$\frac{n\theta^2}{n+2}$	$\frac{\theta^2}{3}$	$\frac{\theta^2}{2}$	$\frac{3\theta^2}{5}$	$\frac{2\theta^2}{6}$	θ^2	si

Si sólo se toma en cuenta el error cuadrático medio el de mayor ECM es $T_4(\mathbb{X})$ tanto para muestras grandes como para muestras pequeñas y $\forall n \in \mathbb{N}$ sucede que

$$\frac{\theta^2}{n(n+2)} \leq \frac{2\theta^2}{(n+1)(n+2)} \leq \frac{\theta^2}{3n}$$

entonces $ECM(T_3) \leq ECM(T_2) \leq ECM(T_1) \leq ECM(T_4)$ y por lo tanto los estadísticos T_1, T_2 y T_3 son buenos estimadores pues conforme mayor sea el tamaño de la muestra, el ECM tiende a cero. A pesar de que T_2 es sesgado se prefiere sobre T_4 pues, a veces, es preciso permitir un poco de sesgo para disminuir la varianza del estimador. Para los estimadores con menor ECM se prefieren aquellos que sean insesgados, es decir T_1 y T_3 y para seleccionar entre estos dos hay que observar que para muestras pequeñas T_3 es la mejor opción.

Esto ejemplifica que no basta con que el estimador cumpla con una cualidad, hay que buscar aquel con la mayoría de las propiedades deseadas algunas de las cuales se irán mencionando conforme avance el texto.

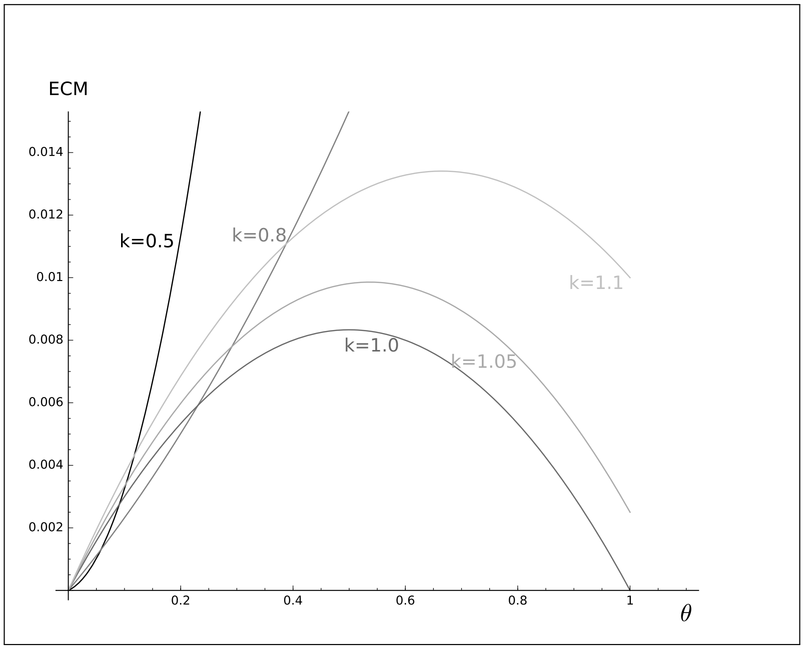
Ejemplo **2.4.5.** Una máquina produce piezas, cada una de las cuales, independientemente, puede ser defectuosa con probabilidad θ. A fin de estimar θ se define

$$X_i = \begin{cases} 1 & \text{si la i-ésima pieza es defectuosa} \\ 0 & \text{e.o.c .} \end{cases}$$

Se desea comparar el estimador $T_1 = \bar{x}$ con $T_k = k\bar{x}$, para $k > 0$. Como X_i se distribuye como una variable aleatoria Bernoulli (θ) el ECM del estimador T_k es

$$\begin{aligned} ECM_{T_k}(\theta) &= Var(T_k) + \Big(\mathbb{E}(T_k) - \theta\Big)^2 \\ &= k^2 Var(\bar{x}) + \Big(k\mathbb{E}(\bar{x}) - \theta\Big)^2 \\ &= k^2\left[\frac{\theta(1-\theta)}{n}\right] + \Big[\theta(1-k)\Big]^2 \\ &= \frac{k^2\theta}{n} - \frac{k^2\theta^2}{n} + \theta^2(1-k)^2 \\ &= \theta^2\left[(1-k)^2 - \frac{k^2}{n}\right] + \frac{k^2\theta}{n}. \end{aligned}$$

De lo que se observa que para k fija el $ECM_{T_k}(\theta)$ es una parábola. Para ejemplificar, fijemos $n = 30$ y dependiendo del valor de k su variación se ve de la siguiente forma.



> Con la gráfica se puede observar que el estimador T_1 tiene menor error cuadrático medio comparado con todos los T_k, $k > 1$; pero entre los estimadores T_k con $0 < k < 1$ no se puede afirmar lo mismo ya que, como se aprecia en la figura, el ECM dependerá del valor de θ. Dicho de manera formal note que el valor de k que minimiza al ECM es $k = \dfrac{n\theta}{1 + \theta(n-1)}$ el cual depende de θ, entonces el estimador T_k que minimiza el ECM es distinto para cada valor del parámetro.

En el ejemplo anterior se puede observar que el ECM no es una medida absoluta de la "calidad"de los estimadores pues no suele ocurrir que uno solo minimice el error de estimación para todos los valores de θ simultáneamente. Obtener el error cuadrático medio para determinar la bondad de un estimador si bien reduce la búsqueda dentro de una clase más pequeña de estimadores no es suficiente así que se deben pedir que el estimador cumpla con propiedades adicionales.

Estimador insesgado uniforme de mínima varianza

Al comprar todos los posibles estimadores de un parámetro puede resultar poco viable debido a la cantidad de elementos a cotejar.

Una primera forma de restringir la clase de estimadores que se van a comparar es pedir que sean insesgados y de esos aquellos que tengan la menor dispersión alrededor del parámetro . Una propuesta interesante es la hecha por Cramer y Rao que consiste en acotar inferiormente la varianza. Esta sección tiene por objetivo encontrar al mejor estimador para θ considerando los criterios ya mencionados.

Se denota la clase de estimadores insesgados con segundo momento finito para θ como

$$C = \left\{ T(\mathbb{X}) \mid \mathbb{E}[T(\mathbb{X})] = \theta, \ \mathbb{E}_\theta \left[T^2 \right] < \infty \right\}.$$

Definición **2.4.3.** Sea $f(X, \theta)$ una familia paramétrica. Se dice que $T'(\mathbb{X}) \in C$ es el estimador insesgado de varianza mínima uniforme (en inglés suele abreviarse como UMVUE) para θ si y sólo si

$$Var(T'(\mathbb{X})) \leq Var(T(\mathbb{X}))$$

para todo $\theta \in \mathbb{H}, n \in \mathbb{N}$ y $T(\mathbb{X}) \in C$.

Nota: A lo largo de este texto se referirá al ‶estimador insesgado de varianza mínima uniforme″con sus siglas en inglés.

Encontrar el UMVUE no es tarea fácil, sin embargo se disponen de herramientas para tal empresa. Algunas de ellas necesitan ciertas consideraciones, con respecto al parámetro, para poder aplicarlas. A dichas hipótesis se las llama condiciones de regularidad y son las siguientes:

1. θ toma valores en un intervalo abierto de $\mathbb{R}$.

2. El conjunto $S = \{(x_1, x_2, \dots, x_n) \in \mathbb{X} \mid f_\theta(x_1, x_2, \dots, x_n) > 0\}$ no depende de θ y para todo θ existe $\dfrac{\partial}{\partial\theta} f_\theta(x_1, x_2, \dots, x_n)$. Al conjunto S se le llama soporte de la distribución f_θ.

3. Para la clase C, de los estimadores insesgados, se debe cumplir que

$$\frac{d}{d\theta}\int_{\mathbb{X}} T(x_1,\dots,x_n) f_\theta(x_1,\dots,x_n)dx_1\dots dx_n = \int_{\mathbb{X}} T(x_1,\dots,x_n)\frac{\partial}{\partial\theta} f_\theta(x_1,\dots,x_n)dx_1\dots dx_n.$$

   Y que esta última expresión sea válida también en el caso $T(x_1,\dots,x_n) = 1$, aunque este no sea un estimador insesgado.

Verificar las condiciones en general no resulta fácil, sin embargo en el texto se dará por hecho que las distribuciones usuales: binomial, exponencial, gamma, poisson, normal y muchas otras las cumplen.

Proposición **2.4.2.** *La condición 3. de regularidad es equivalente a* $\mathbb{E}\left[\dfrac{\partial}{\partial\theta}\log f_\theta(\mathbb{X})\right] = 0$.

Demostración. En efecto,

$$\frac{d}{d\theta}\int\dots\int f_\theta(\mathbb{X})dx_1dx_2\dots dx_n = \frac{d}{d\theta}1 = 0$$

y la tercera condición equivale a

$$
\begin{aligned}
0 &= \frac{d}{d\theta}\int \cdots \int f_\theta(\mathbb{X}) dx_1 dx_2 \ldots dx_n \\
&= \int \cdots \int \frac{\partial}{\partial\theta} f_\theta(\mathbb{X}) dx_1 dx_2 \ldots dx_n \\
&= \int \cdots \int \frac{\frac{\partial}{\partial\theta} f_\theta(\mathbb{X})}{f_\theta(\mathbb{X})} f_\theta(\mathbb{X}) dx_1 dx_2 \ldots dx_n \\
&= \int \cdots \int \left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right) f_\theta(\mathbb{X}) dx_1 dx_2 \ldots dx_n \\
&= \mathbb{E}\left[\frac{\partial}{\partial\theta}\log f_\theta(\mathbb{X})\right].
\end{aligned}
$$

□

Definición **2.4.4.** A la varianza de la variable aleatoria $\frac{\partial}{\partial\theta}\log f_\theta(\mathbb{X})$ se le llama *Información de Fisher* y se denota por

$$I_{\mathbb{X}}(\theta) := Var_\theta\left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right) = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_\theta\left(X_1, X_2, \ldots, X_n\right)\right)^2\right].$$

Proposición **2.4.3.** *Si $f_\theta(\mathbb{X})$ cumple las condiciones de regularidad. y las observaciones son de una muestra aleatoria simple, entonces $I_{\mathbb{X}}(\theta) = nI_1(\theta)$; con $I_1(\theta)$ es la información del modelo dada una observación.*

Demostración**.** Como la muestra es un conjunto de variables independientes, entonces la función de densidad conjunta es el producto de las densidades

$$f_\theta(x_1, x_2, \ldots, x_n) = f_\theta(x_1) f_\theta(x_2) \ldots f_\theta(x_n)$$

por lo tanto

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f_\theta(\mathbb{X})\right)^2\right] &= \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f_\theta(x_i)\right)^2\right]\\
&= \mathbb{E}\left[\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f_\theta(x_i)\right)^2\right]\\
&= \sum_{i=1}^{n}\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f_\theta(x_i)\right)^2\right] + \sum_{i\neq j}\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log(f_\theta(x_i))\right)\left(\frac{\partial}{\partial\theta}\log(f_\theta(x_j))\right)\right]\\
&= n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f_\theta(x)\right)^2\right] + \sum_{i\neq j}\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log(f_\theta(x_i))\right)\right]\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log(f_\theta(x_j))\right)\right]\\
&= n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f_\theta(x_i)\right)^2\right].
\end{aligned}
$$

Es decir, la información que aporta una muestra de tamaño n es n veces la información que aporta una muestra de tamaño uno. □

Teorema **2.4.1.** *(Teorema de Cramer y Rao) Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de $f_\theta(X)$. Supóngase que $f_\theta(X)$ satisface las condiciones de regularidad, además sea $T(\mathbb{X})$ cualquier estadística insesgada para $q(\theta)$. Entonces,*

$$Var(T(X)) \geq CICR.$$

Donde, $CICR = \dfrac{[q'(\theta)]^2}{I_{\mathbb{X}}(\theta)}$ *con* $I_{\mathbb{X}}(\theta) = n\,\mathbb{E}\left[\left(\dfrac{\partial}{\partial\theta}\log(f_\theta(X))\right)^2\right]$

Dándose la igualdad si y sólo si, existe una función $K(n,\theta)$ tal que

$$\sum_{i=1}^{n}\frac{d}{d_\theta}\log(f_\theta(X_i)) = K\,(n,\theta)\,[T^*(\mathbb{X}) - q(\theta)]. \tag{2.10}$$

En tal caso $T^*(\mathbb{X})$ es el UMVUE para $q(\theta)$.

Demostración. Primero hay que notar que

$$\mathbb{E}\left[\Big(T(X)-q(\theta)\Big)\left(\frac{\partial}{\partial\theta}\log(f(X,\theta))\right)\right]=q'(\theta)$$

$$\begin{aligned}
E\left[\Big(T(\mathbb{X})-q(\theta)\Big)\left(\frac{\partial}{\partial\theta}\log(f_\theta(X))\right)\right] &= \int\cdots\int\Big(T(\mathbb{X})-q(\theta)\Big)\left(\frac{\partial}{\partial\theta}\log(f_\theta(X))\right)\cdot f_\theta(\mathbb{X})d_{\mathbb{X}}\\
&= \int\cdots\int\Big(T(\mathbb{X})-q(\theta)\Big)\left(\frac{\partial}{\partial\theta}f_\theta(\mathbb{X})\right)\cdot f_\theta(\mathbb{X})d_{\mathbb{X}}\\
&= \int\cdots\int\Big(T(\mathbb{X})-q(\theta)\Big)\left(\frac{\partial}{\partial\theta}\log(f_\theta(X))\right)d_{\mathbb{X}}\\
&= \int\cdots\int T(\mathbb{X})\left(\frac{\partial}{\partial\theta}\log(f_\theta(X))\right)d_{\mathbb{X}}-\int\cdots\int q(\theta)\left(\frac{\partial}{\partial\theta}\log(f_\theta(X))\right)d_{\mathbb{X}}\\
&= \frac{\partial}{\partial\theta}\int\cdots\int T(\mathbb{X})f_\theta(\mathbb{X})d_{\mathbb{X}}-q(\theta)\frac{\partial}{\partial\theta}\int\cdots\int f_\theta(\mathbb{X})d_{\mathbb{X}}\\
&= \frac{d}{d_\theta}q(\theta)-q(\theta)\frac{\partial}{\partial\theta}1
\end{aligned}$$

Entonces,

$$\mathbb{E}\left[\Big(T(X)-q(\theta)\Big)\left(\frac{\partial}{\partial\theta}\log(f_\theta(\mathbb{X}))\right)\right]=q'(\theta) \qquad (2.11)$$

Luego, aplicando la desigualdad de Cauchy-Schwartz a $E\left[\Big(T(X)-q(\theta)\Big)\left(\frac{\partial}{\partial\theta}\log(f_\theta(\mathbb{X}))\right)\right]$ se obtiene:

$$\mathbb{E}^2\left[\Big(T(X)-q(\theta)\Big)\left(\frac{\partial}{\partial\theta}\log(f_\theta(\mathbb{X}))\right)\right]\leq\mathbb{E}^2\left[\Big(T(X)-q(\theta)\Big)^2\right]\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f_\theta(\mathbb{X})\right)^2\right]$$

además $\mathbb{E}^2\left[\Big(T(X)-q(\theta)\Big)^2\right]=Var(T(\mathbb{X}))$. Entonces por la Ecuación 2.11

se obtiene;

$$\Big(q'(\theta)\Big)^2 \leq Var\Big(T(\mathbb{X})\Big)\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f_\theta(\mathbb{X})\right)^2\right],$$

por lo tanto,

$$Var[T(\mathbb{X})] \geq \frac{[q'(\theta)]^2}{\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f_\theta(\mathbb{X})\right)^2\right]},$$

es decir,

$$Var(T(\mathbb{X})) \geq CICR.$$

Para finalizar la prueba, resta probar que la igualdad se cumple si y sólo si existe $K = K(n,\theta)$ tal que;

$$\sum_{i=1}^{n}\frac{d}{d_\theta}\log f_\theta(x_i) = K(n;\theta)\Big[T^*\Big(\mathbb{X} - q(\theta)\Big)\Big].$$

En efecto, hay que notar que en la desigualdad de Cauchy-Schwartz

$$\mathbb{E}^2[XY] \leq \mathbb{E}[X^2]\,\mathbb{E}[Y^2],$$

la igualdad se da si y sólo si, $Y = aX$ con $a \in \mathbb{R}$, $a \neq 0$ si y sólo si

$$\frac{\partial}{\partial\theta}[\log f_\theta(\mathbb{X})] = K\,[T(\mathbb{X} - q(\theta))]$$

pero

$$\frac{\partial}{\partial\theta}[lnf_\theta(\mathbb{X})] = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f_\theta(x_i)$$

Por lo tanto, existe una constante K que depende del tamaño de la muestra y del parámetro θ tal que,

$$\sum_{i=1}^{n}\frac{d}{d_\theta}\log f_\theta(x_i) = K(n;\theta)\Big[T^*\Big(\mathbb{X} - q(\theta)\Big)\Big].$$

□

El teorema anterior garantiza que si $f_\theta(\mathbb{X})$ satisface las condiciones de regularidad entonces la varianza de cualquier estimador insesgado para θ no puede ser menor que la Cota Inferior de Cramer y Rao y esto, a su vez, implica que cualquier estimador insesgado para θ cuya varianza coincida con la CICR es un UMVUE para $q(\theta)$. Sin embargo, cabe mencionar que el hecho de que no existan estimadores insesgados cuya varianza coincida con la CICR no implica que no exista el UMVUE.

Ejemplo **2.4.6.** sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria con distribución exponencial de parámetro θ. Encontrar la CICR para $\frac{1}{\theta}$.

Como primer paso se calcula la información de Fisher para una observación:

La función de densidad es $f_\theta(X) = \theta e^{-\theta x}$

$$\log f_\theta(X) = \log\theta - \theta x$$

$$\Rightarrow \frac{\partial}{\partial\theta}\log f_\theta(X) = \frac{1}{\theta} - x$$

$$\Rightarrow \left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right)^2 = \frac{1}{\theta^2} - \frac{2X}{\theta} + X^2$$

$$\Rightarrow \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right)^2\right] = \frac{1}{\theta^2} - \frac{2\mathbb{E}[X]}{\theta} + \mathbb{E}[X^2]$$

Como $\mathbb{E}[X] = \dfrac{1}{\theta}$ y $\mathbb{E}[X^2] = var[X] + \mathbb{E}^2[X] = \dfrac{1}{\theta^2} + \dfrac{1}{\theta^2} = \dfrac{2}{\theta^2}$,

$$\Rightarrow \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right)^2\right] = \frac{1}{\theta^2} - \frac{2}{\theta^2} + \frac{2}{\theta^2} = \frac{1}{\theta^2}$$

Hasta el momento se obtuvo la información de Fisher por unidad muestral, por la Proposición 2.4.3

$$I_{\mathbb{X}}(\theta) = n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right)^2\right] = \frac{n}{\theta^2}$$

Por otra parte $[q'(\theta)]^2 = \dfrac{d}{d\theta}\dfrac{1}{\theta} = -\dfrac{1}{\theta^2}$ por lo que la cota inferior de Cramer y Rao es

$$CICR = \frac{1/\theta^2}{n/\theta^2} = \frac{1}{n}$$

Con esto se concluye que cualquier estimador insesgado para $q(\theta) = \frac{1}{\theta}$ siempre tendrá varianza mayor o igual que $\frac{1}{n}$.

Bajo las condiciones de regularidad, si se alcanza la CICR del Teorema 2.4.1, entonces

$$\frac{\partial}{\partial\theta}\log f_\theta(X) = K(n,\theta)\,[T^*(\mathbb{X}) - q(\theta)]$$

integrando respecto de θ y aplicando la función exponencial se obtiene

$$f_\theta(X) = e^{T(x)\int K(n,\theta)d\theta + \int K(n,\theta)q(\theta)d\theta + s(x)}.$$

Donde $s(x)$ es una constante de integración que puede depender de x. Las distribuciones que se pueden escribir de esa forma se dice que pertenecen a la familia de exponenciales, esta familia incluye muchas distribuciones comunes entre ellas están la distribución Normal, Exponencial, Gamma, Bernoulli, Poisson, etc. En algunos casos se debe fijar algún parámetro de la misma, por ejemplo las distribuciones binomial y multinomial pertenecen a la familia de exponenciales si se fija el número de pruebas. Una función de distribución de la familia de exponenciales, $f_\theta(X)$, cumple las condiciones de regularidad cuando $\int T(x)K(n,\theta)d\theta$ es derivable respecto a θ y su derivada nunca es nula [1].

Ejemplo **2.4.7.** Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria Poisson(θ) con $\theta > 0$.

a) Hallar la CICR para el estimador insesgado de $q(\theta) = \sqrt{\theta}$

b) Verificar si existe o no una función $q(\theta)$ para la cual exista una estadística cuya varianza alcanza la CICR.

Al pertenecer a la familia de exponenciale cumple las propiedades de regularidad, su información de Fisher es

$$
\begin{aligned}
I_{\mathbb{X}}(\theta) &= n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log\left(f_\theta(X)\right)\right)^2\right] \\
&= n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log\frac{e^{-\theta}\cdot\theta^X}{X!}\right)^2\right] \\
&= n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\left(-\theta + X\log(\theta) - \log(X)\right)\right)^2\right] \\
&= n\mathbb{E}\left[\left(-1 + \frac{X}{\theta}\right)^2\right] \\
&= n\mathbb{E}\left[1 - \frac{2X}{\theta} + \frac{X^2}{\theta^2}\right] \\
&= n\left[1 - \frac{2}{\theta}\mathbb{E}[X] + \frac{1}{\theta^2}[X^2]\right] \\
&= n\left[1 - \frac{2}{\theta}\cdot\theta + \frac{1}{\theta^2}(\theta+\theta^2)\right] \\
&= \frac{n}{\theta}
\end{aligned}
$$

Y como $[\frac{d}{d\theta}\sqrt{\theta}]^2 = \frac{1}{4\theta}$ Entonces, $CICR = \dfrac{\left(\dfrac{1}{4\theta}\right)}{\left(\dfrac{1}{\theta}\right)} = \dfrac{1}{4n}$.

Para el inciso b).

$$
\begin{aligned}
\sum_{i=1}^{n} \frac{d}{d_\theta} \log f_\theta(X_i) &= \sum_{i=1}^{n} \left(-1 + \frac{X_i}{\theta}\right) \\
&= -n + \frac{1}{\theta} \sum_{i=1}^{n} x_i \\
&= \frac{1}{\theta} \left[\sum_{i=1}^{n} X_i - n\theta \right] \\
&= \frac{n}{\theta} [\bar{x} - \theta].
\end{aligned}
$$

Entonces existe una función que depende del parámetro y del tamaño de la muestra que cumple con la relación 2.10 y por el Teorema 2.4.1 se concluye que $\bar{x}$ es el **UMVUE** para $q(\theta) = \theta$.

La CICR establece un límite inferior para la varianza del estimador buscado, sin embargo existen UMVUE que no alcanzan tal cota, por lo que se define

Definición **2.4.5.** Sea $T(\mathbb{X})$ un estimador para $q(\theta)$. Se define la eficiencia de $T(\mathbb{X})$ como:

$$
ef(T) = \frac{CICR}{Var(T(\mathbb{X}))}
$$

con $0 < ef(T) \leq 1$.

Esto permite comparar dos estimadores entre sí, en el sentido que un estimador T es más eficiente que S si $Var(T) < Var(S)$. En el caso que se cumplan las condiciones de regularidad, si existe el UMVUE este es el estimador más eficiente. Además el comparar la varianza de un estimador con el mínimo que puede alcanzar es un buen criterio para decidir si vale la pena cambiar un estimador T por otro de menor varianza o si la proximidad entre $Var(T)$ y su mínimo hace poco rentable el esfuerzo. Por ejemplo los estimadores obtenidos por máxima verosimilitud son asintóticamente eficientes, es decir cuando n tiende a infinito, la varianza tiende a la CICR.

Consistencia de Estimadores

Ya se mencionó en la sección anterior que se busca que los valores del estimador sean muy parecidos al verdadero valor del parámetro. La propiedad de consistencia hace referencia al comportamiento asintótico de un estimador, cuando el tamaño de la muestra n tiende a infinito.

Definición **2.4.6.** Si $T_n(\mathbb{X}) = T_n(X_1, X_2, \ldots, X_n)$ es una sucesión de estimadores asociada a los tamaños muestrales n y $q(\theta)$ una función de θ. Se denomina a $T_n(\mathbb{X})$ un estimador consistente simple para $q(\theta)$ si converge en probabilidad a $q(\theta)$, es decir si para todo $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\Big(q(\theta) - \epsilon \leqslant T_n(\mathbb{X}) \leqslant q(\theta) + \epsilon\Big) = 1.$$

El concepto de consistencia se puede restringir a versiones más estrictas si se sustituye la convergencia en probabilidad por convergencia casi segura o en media cuadrática (consistencia en ECM):

$$\mathbb{P}\Big(T_n \neq q(\theta)\Big) = 0 \quad \text{y} \quad \lim_{n\to\infty} \mathbb{E}\Big((T_n(\mathbb{X}) - q(\theta))^2\Big) = \lim_{n\to\infty} ECM_q(T_n(\mathbb{X})) = 0$$

.

Como se puede notar un estimador $T(\mathbb{X})$ es consistente en ECM si, conforme el tamaño de la muestra crece, entonces la varianza y el sesgo de $T(\mathbb{X})$ tienden a *cero*. Por lo que es deseable trabajar con esta versión de consistencia, en particular:

Teorema 2.4.2. *Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de $f(X,\theta)$. Si $T_n(\mathbb{X})$ es un estimador consistente en error cuadrático medio entonces $T_n(\mathbb{X})$ es un estimador consistente simple.*

Demostración. Supóngase que $T_n(\mathbb{X})$ es tal que

$$\lim_{n\to\infty} \mathbb{E}\Big((T_n(\mathbb{X}) - q(\theta))^2\Big) = 0.$$

$$\begin{aligned}\mathbb{P}\Big(q(\theta) - \epsilon \leqslant T_n(\mathbb{X}) \leqslant q(\theta) + \epsilon\Big) &= \mathbb{P}\Big(-\epsilon \leqslant T_n(\mathbb{X}) - q(\theta) \leqslant \epsilon\Big) \\ &= \mathbb{P}\Big(\mid T_n(\mathbb{X}) - q(\theta) \mid\Big) \\ &\geq 1 - \frac{\mathbb{E}\Big(\Big[T_n(\mathbb{X}) - q(\theta)\Big]^2\Big)}{\epsilon^2}.\end{aligned}$$

Cabe mencionar que la primera desigualdad fue por Chebysev, tomando límites

$$\begin{aligned}1 &\geq \lim_{n\to\infty} \mathbb{P}\Big(q(\theta) - \epsilon \leqslant T_n(\mathbb{X}) \leqslant q(\theta) + \epsilon\Big) \\ &\geq \lim_{n\to\infty}\left(1 - \frac{\mathbb{E}\Big((T_n(\mathbb{X}) - q(\theta))^2\Big)}{\epsilon^2}\right)\end{aligned}$$

Entonces $\lim_{n\to\infty} \mathbb{P}\Big(q(\theta) - \epsilon \leqslant T_n(\mathbb{X}) \leqslant q(\theta) + \epsilon\Big) = 1$. Por lo tanto, $T_n(\mathbb{X})$ es un estimador consistente simple. □

La consistencia de un estimador por tratarse de una propiedad asintótica que no depende de las cualidades del estimador para ningún n fijo es un requisito que no presta mucha ayuda a la hora de seleccionar el mejor estimador, sin embargo, en general, se considera una propiedad indispensable para cualquier método de estimación sobre todo cuando un estimador no es insesgado se le exige que al menos sea consistente.

En particular, los estimadores de máxima verosimilitud cumplen con ser consistentes. Para demostrar esto se denotará θ el parámetro a estimar, θ_0 el valor verdadero del parámetro y θ_{EMV} el estimador de máxima verosimilitud y se pedirán las siguientes condiciones de regularidad:

- El espacio paramétrico debe estar acotado y cerrado, es decir Ω es compacto.
- El valor del parámetro real θ_0 de ser un punto interior del espacio paramétrico.
- La función de verosimilitud debe ser continua para θ.
- La función de verosimilitud es dos veces diferenciable en un entorno de θ_0.
- La diferenciación y la integración son intercambiables.

Teorema **2.4.3.** *El EMV es consistente.*

Demostración**.** Sea $(X_1, X_2, \ldots, X_n)$ una muestra i.i.d., entonces:

$$\frac{1}{n}\ell(X_1, X_2, \ldots, X_n, \theta) = \frac{1}{n}\sum_{i=1}^{n}\ell(X_i, \theta)$$

Si se define la función esperanza del logaritmo de verosimilitud:

$$\mathcal{L}(\theta) = \mathbb{E}_{\theta_0} \log f_\theta(X) = \int_{-\infty}^{\infty} \log f_\theta(X) f_{\theta_0}(X) dx$$

Si se considera la diferencia de la función anterior evaluada en algún valor θ y el verdadero valor del parámetro θ_0 se tiene

$$\mathcal{L}(\theta) - \mathcal{L}(\theta_0) = \mathbb{E}_{\theta_0}\left[\log f_\theta(X) - \log f_{\theta_0}(X)\right] = \mathbb{E}_{\theta_0}\left[\log \frac{f_\theta(X)}{f_{\theta_0}(X)}\right].$$

Aplicando la desigualdad de Jensen

$$\mathbb{E}_{\theta_0}\left[\log \frac{f_\theta(X)}{f_{\theta_0}(X)}\right] \leq \log \mathbb{E}_{\theta_0}\left[\frac{f_\theta(X)}{f_{\theta_0}(X)}\right] = \log \int_{-\infty}^{\infty} \frac{f_\theta(X)}{f_{\theta_0}(X)} f_{\theta_0}(X) dx = \log \int_{-\infty}^{\infty} f_\theta(X) dx = 0,$$

por lo que $\mathcal{L}(\theta) \leq \mathcal{L}(\theta_0)$, es decir, si existe θ_i tal que $\mathcal{L}(\theta_i) \geq \mathcal{L}(\theta)$ para todo $\theta \in \Omega$, entonces $\theta_i \leq \theta_0$, por lo tanto θ_0 es el máximo de $\mathcal{L}(\theta)$

Por la ley fuerte de los grandes números

$$\frac{1}{n}\sum_{i=1}^{n} \ell(X_i, \theta) \overset{a.s.}{\to} \mathbb{E}_{\theta_0} \log f_\theta(X).$$

Como el espacio Ω es compacto, el EMV cumple

$$\theta_{EMV} = \sup_{\theta \in \Omega} \frac{1}{n} \sum_{i=1}^{n} \ell(X_i, \theta) \overset{a.s.}{\to} \sup_{\theta \in \Omega} \mathbb{E}_{\theta_0} \log f_\theta(x) = \theta_0.$$

Por lo tanto

$$\theta_{EMV} \overset{a.s.}{\to} \theta_0$$

$\square$

2.5 Suficiencia

Como ya se vio en capítulos anteriores que, dado un experimento aleatorio, se puede tomar una muestra $x_1, x_2, ..., x_n$ con la que es posible inferir el valor del parámetro desconocido. Sin embargo, para tamaños muy grandes de la muestra, los procesos de estimación e interpretación pueden llegar a ser complicados por lo que surge el interés en encontrar métodos viables de estimación, en el sentido que la cantidad de datos con los que se trabaje sea mínima y al mismo tiempo sea suficiente para dar información de toda población. En este capítulo se estudian algunas herramientas para seleccionar el mejor estadístico: aquel que resume al máximo la información obtenida de la muestra y que logra representar a toda la población.

Estadísticas suficientes

Toda inferencia realizada acerca del parámetro Θ se tiene que obtener única y exclusivamente de la muestra aleatoria, además la función $T(\mathbb{X})$ no es única y al usar cualquier estadístico $T(\mathbb{X})$ se puede perder información.

En 1992 R. A. Fisher introdujo el concepto de *Estadística suficiente*, que es un estadístico que resume toda la información que puede proporcionar la muestra respecto a Θ, es decir, una vez conocido el valor de una estadística suficiente $S(\mathbb{X})$ la función de densidad conjunta de la muestra $x_1, x_2, \ldots, x_n$ ya no podrá aportar información adicional del parámetro Θ, de manera formal:

Definición **2.5.1.** Sea $x_1, x_2, \ldots, x_n$ muestra aleatoria de $f_\Theta(x)$ y sea $S(\mathbb{X})$ una estadística, se dice que $S(\mathbb{X})$ es una estadística suficiente si y sólo si $\mathbb{P}[\mathbb{X} = x \mid T(\mathbb{X}) = T(x)]$ no depende de Θ para todo vector $\mathbb{X} = (X_1, \ldots, X_n)$ en el espacio muestral, $\mathbb{H}$, de $\mathbb{X} = (X_1, ..., X_n)$.

Es claro que la estadística $T(X) = X$ es suficiente, a esta estadística se le llama *estadística suficiente trivial*, naturalmente se buscan las estadísticas que eliminen la mayor cantidad de información superflua y que sean de la menor dimensión, de este modo se reduce la clase de los estimadores que actúan únicamente en función de T, además se debe utilizar toda la información relevante, respecto a θ, contenida en la muestra y que no haya otro estadístico que pueda proporcionar más información adicional sobre el parámetro poblacional.

Ejemplo **2.5.1.** Una urna contiene N bolas numeradas de $N+1$ a $2N$; se extraen n bolas con reemplazamiento para estimar el valor de N, quedando la muestra $(X_1, X_2, \ldots, X_n)$. La función de probabilidad de la muestra es

$$\mathbb{P}_N\left\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\right\} = \frac{1}{N^n}$$

denotando $X_{(1)} = \min\limits_{1 \le i \le n}\left\{X_i\right\}$ y $X_{(n)} = \max\limits_{1 \le i \le n}\left\{X_i\right\}$, entonces

$$\begin{aligned}\mathbb{P}_N\left\{X_{(1)} = y\right\} &= \mathbb{P}_N\left\{X_{(1)} \ge y\right\} - \mathbb{P}_N\left\{X_{(1)} \ge y+1\right\} \\ &= \left(\frac{2N - y + 1}{N}\right)^n - \left(\frac{2N - y}{N}\right)^n\end{aligned}$$

para $y = N+1, N+2, \ldots, 2N$. Análogamente

$$\begin{aligned}\mathbb{P}_N\left\{X_{(n)} = z\right\} &= \mathbb{P}_N\left\{X_{(n)} \leq z\right\} - \mathbb{P}_N\left\{X_{(n)} \leq z-1\right\}\\ &= \left(\frac{z-N}{N}\right)^n - \left(\frac{z-N-1}{N}\right)^n\end{aligned}$$

para $y = N+1, N+2, \ldots, 2N$. La distribución conjunta de $X_{(1)}$ y $X_{(n)}$, es:

$$\begin{aligned}\mathbb{P}_N\left\{X_{(1)} = y, X_{(n)} = z\right\} &= \mathbb{P}_N\left\{X_{(1)} \geq y, X_{(n)} \leq z\right\} - \mathbb{P}_N\left\{X_{(1)} \geq y+1, X_{(n)} \leq z\right\}\\ &\quad - \mathbb{P}_N\left\{X_{(1)} \geq y, X_{(n)} \leq z-1\right\} + \mathbb{P}_N\left\{X_{(1)} \geq y+1, X_{(n)} \leq z-1\right\}\\ &= \left(\frac{z-y+1}{N}\right)^n - 2\left(\frac{z-y}{N}\right)^n + \left(\frac{z-y-1}{N}\right)^n\end{aligned}$$

para $y < z$; $y, z \in \left\{N+1, N+2, \ldots, 2N\right\}$ $\left(\text{ y } \frac{1}{N^n} \text{ si } y = z\right)$. Como

$$\mathbb{P}_N\left\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\right\} = \frac{1}{N^n},$$

siempre que $y \leq x_1, \ldots, x_n \leq z$; $y, z \in \{N+1, N+2, \ldots, 2N\}$ y haya algún x_i que coincida con y y algún x_j que coincida con z, se obtiene

$$\begin{aligned}\mathbb{P}_N\left\{X_1 = x_1, \ldots, X_n = x_n \mid X_{(1)} = y, X_{(n)} = z\right\} &= \frac{\frac{1}{N^n}}{\left(\frac{z-y+1}{N}\right)^n - 2\left(\frac{z-y}{N}\right)^n + \left(\frac{z-y-1}{N}\right)^n}\\ &= \frac{1}{(z-y+1)^n - 2(z-y)^n + (z-y-1)^n}\end{aligned}$$

para cualesquiera $x_1, x_2, \ldots, x_n \in y, y+1, \ldots, z$, con algún $x_i = y$ y algún $x_j = z$. Como el resultado no depende de N, entonces $T = \left(X_{(1)}, X_{(n)}\right)$ es suficiente. En cambio

$$\mathbb{P}_N\left\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid X_{(1)} = y\right\} = \frac{1}{(2N-y+1)^n - (2N-y)^n}$$

siempre que $x_1, x_2, \ldots, x_n \in y, y+1, \ldots, 2N$ y exista $x_i = y$; por lo que $X_{(1)}$ no es suficiente, de manera análoga $X_{(n)}$ tampoco es suficiente.

Ejemplo **2.5.2.** Se desea conocer el número de alumnos fumadores dentro de un salón de clases. Para ello se tomará una muestra aleatoria de n alumnos. Definimos a la variable aleatoria X como

$$X_i = \begin{cases} 0 & \text{no fumador c.p } 1-\theta, \\ 1 & \text{fumador c.p. } \theta. \end{cases}$$

En este caso el vector $\mathbb{X}$ esta dado por $\mathbb{X} = (X_1, \ldots, X_n)$ y X_i son i.i.d Bernoulli de parámetro θ. Por otro lado el espacio de todos los posibles resultados, $\mathbb{H}$, de $\mathbb{X}$ está dado por

$$\mathbb{H} = \Big\{(x_1, \ldots, x_n) : x_i \in \{0, 1\}\Big\}.$$

Parece intuitivo que para determinar θ no es necesario saber las posiciones de los alumnos fumadores y no fumadores, basta saber cuál es el total de fumadores, de este modo se define la estadística $S(\mathbb{X}) = \sum_{i=1}^{n} X_i$, ésta reduce la información del espacio euclideano $\mathbb{H}$; ya que, $A = \{0, 1, \ldots, n\}$, el conjunto de posibles valores para $S(\mathbb{X})$ es de menor cardinalidad y dimensión que $\mathbb{H}$.

Ahora se probará que $S(\mathbb{X})$ no sólo reduce la información si no que es una estadística suficiente, es decir $\mathbb{P}\Big(X_1 = x_1, \ldots, X_n = x_n \mid S(\mathbb{X}) = s\Big)$ no depende de Θ para toda muestra $\mathbb{X} \in \mathbb{H}$.

En efecto, sea $s \in \mathbb{H}$.

$$
\begin{aligned}
\mathbb{P}\Big(X_1 = x_1, \ldots, X_n = x_n \mid S(\mathbb{X}) = s\Big) &= \frac{\mathbb{P}\Big(X_1 = x_1, \ldots, X_n = x_n, S(\mathbb{X}) = s\Big)}{\mathbb{P}\Big(S(\mathbb{X}) = s\Big)} \\
&= \frac{\mathbb{P}\Big(X_1 = x_1, \ldots, X_n = x_n\Big)\mathbb{P}\Big(S(\mathbb{X}) = s \mid X_1 = x_1, \ldots, X_n = x_n\Big)}{\mathbb{P}\Big(S(\mathbb{X}) = s\Big)} \\
&= \frac{\theta^{x_1}(1-\theta)^{1-x_1} \ldots \theta^{x_n}(1-\theta)^{1-x_n}\mathbb{P}\Big(S(\mathbb{X}) = s \mid X_1 = x_1, \ldots, X_n = x_n\Big)}{\binom{n}{s} p^t (1-p)^{n-s}} \\
&= \frac{\theta^{s}(1-\theta)^{1-s}\mathbb{P}\Big(S(\mathbb{X}) = s \mid X_1 = x_1, \ldots, X_n = x_n\Big)}{\binom{n}{s} \theta^t (1-\theta)^{n-s}} \\
&= \begin{cases} \dfrac{1}{\binom{n}{s}} & \text{si } \displaystyle\sum_{i=1}^{n} x_i = t \\ 0 & \text{si } \displaystyle\sum_{i=1}^{n} x_i \neq t \end{cases}
\end{aligned}
$$

Por lo tanto, la función de densidad condicional $\mathbb{P}\Big(X_1 = x_1, \ldots, X_n = x_n \mid S(\mathbb{X}) = s\Big)$ no depende del parámetro θ y la estadística $S(\mathbb{X}) = \displaystyle\sum_{i=1}^{n} x_i$ es una estadística suficiente.

En la práctica la Definición 2.5.1 no es muy útil porque se necesita tener previamente el candidato a estadístico suficiente, para después comprobar si lo es, además no nos dice como se puede encontrar un estadístico o estimador suficiente. Un teorema que permite obtener estadísticos suficientes es:

Teorema **2.5.1.** *(Teorema de factorización de Fisher-Neyman) Sea* $\mathbb{X} = \Big(X_1, \ldots, X_n\Big)$ *una muestra aleatoria simple de una población con función de densidad* $f_{\mathbb{X}}(\mathbb{X}, \theta)$*, entonces el estadístico* $S(\mathbb{X})$ *es suficiente si y sólo si existen dos funciones* $g\Big(S(\mathbb{X}), \theta\Big)$ *y*

$h(\mathbb{X})$ tal que $g\Big(S(\mathbb{X}),\theta\Big)$ sólo depende de θ y $h(\mathbb{X})$ no depende de θ y se cumple que $f_{\mathbb{X}}(\mathbb{X},\theta) = g\Big(S(\mathbb{X}),\theta\Big)h(\mathbb{X})$.

Demostración. Aquí se hace la demostración del caso discreto, la demostración general se puede encontrar en [3]. En este caso se puede escribir $f_{\mathbb{X}}(\mathbb{X},\theta) = \mathbb{P}_\theta\Big(X_1 = x_1,\ldots,X_n = x_n\Big)$

$\Rightarrow$) Supóngase que $S(\mathbb{X})$ es una estadística suficiente para θ, entonces la distribución condicionada

$$\mathbb{P}_\theta\Big(X_1 = x_1,\ldots,X_n = x_n \mid S(x_1,\ldots,x_n) = t\Big) = \frac{\mathbb{P}_\theta\Big(X_1 = x_1,\ldots,X_n = x_n, S(x_1,\ldots,x_n) = t\Big)}{\mathbb{P}_\theta\Big(S(x_1,\ldots,x_n) = t\Big)}$$

es independiente del parámetro θ, así se puede escribir

$$\begin{aligned}\mathbb{P}_\theta\Big(X_1 = x_1,\ldots,X_n = x_n\Big) &= \mathbb{P}_\theta\Big(X_1 = x_1,\ldots,X_n = x_n, S(x_1,\ldots,x_n) = t\Big)\\ &= \mathbb{P}_\theta\Big(S(x_1,\ldots,x_n) = t\Big)\mathbb{P}\Big(X_1 = x_1,\ldots,X_n = x_n \mid S(x_1,\ldots,x_n) = t\Big)\end{aligned}$$

siempre y cuando la probabilidad condicionada

$$\mathbb{P}\Big(X_1 = x_1,\ldots,X_n = x_n \mid S(x_1,\ldots,x_n) = t\Big)$$

esté bien definida. Haciendo

$$h(\mathbb{X}) = \mathbb{P}\Big(X_1 = x_1,\ldots,X_n = x_n \mid S(x_1,\ldots,x_n) = t\Big)$$

que no depende de θ, y

$$g\Big(S(\mathbb{X}),\theta\Big) = \mathbb{P}_\theta\Big(X_1 = x_1,\ldots,X_n = x_n, S(x_1,\ldots,x_n) = t\Big)$$

se tiene que

$$\mathbb{P}_\theta\Big(X_1 = x_1,\ldots,X_n = x_n\Big) = g\Big(S(\mathbb{X}),\theta\Big)h(\mathbb{X}).$$

$\Leftarrow$) Supóngase que existen $g\Big(S(\mathbb{X}),\theta\Big)$ y $h(\mathbb{X})$ tal que

$$\mathbb{P}_\theta\Big(X_1 = x_1,\dots,X_n = x_n\Big) = g\Big(S(\mathbb{X}),\theta\Big)h(\mathbb{X}),$$

entonces

$$\mathbb{P}_\theta\Big(X_1 = x_1,\dots,X_n = x_n \mid S(x_1,\dots,x_n) = t\Big) = \begin{cases} 0 & \text{si } S(\mathbb{X}) \neq t \\ \dfrac{\mathbb{P}_\theta\Big(X_1 = x_1,\dots,X_n = x_n, S(x_1,\dots,x_n) = t\Big)}{\mathbb{P}_\theta\Big(S(x_1,\dots,x_n) = t\Big)} & \text{si } S(\mathbb{X}) = t \end{cases}$$

$$= \begin{cases} 0 & \text{si } S(\mathbb{X}) \neq t \\ \dfrac{\mathbb{P}_\theta\Big(X_1 = x_1,\dots,X_n = x_n\Big)}{\mathbb{P}_\theta\Big(S(x_1,\dots,x_n) = t\Big)} & \text{si } S(\mathbb{X}) = t \end{cases}$$

Si $S(x_1,\dots,x_n) \neq t$, entonces la probabilidad condicionada

$$\mathbb{P}_\theta\Big(X_1 = x_1,\dots,X_n = x_n \mid S(x_1,\dots,x_n) = t\Big) = 0$$

no depende del parámetro θ. Si $S(x_1,\dots,x_n) = t$ entonces teniendo en cuenta que se verifica el criterio de factorización se puede escribir

$$\begin{aligned} \mathbb{P}_\theta\Big(S(x_1,\dots,x_n) = t\Big) &= \sum_{S(x_1,\dots,x_n)=t} \mathbb{P}_\theta\Big(X_1 = x_1,\dots,X_n = x_n\Big) \\ &= \sum_{S(x_1,\dots,x_n)=t} g\Big(S(\mathbb{X}),\theta\Big)h(\mathbb{X}) \\ &= g\Big(S(\mathbb{X}),\theta\Big) \sum_{S(x_1,\dots,x_n)=t} h(\mathbb{X}) \end{aligned}$$

luego

$$\begin{aligned}
\mathbb{P}_\theta\Big(X_1 = x_1,\ldots,X_n = x_n, S(x_1,\ldots,x_n) = t\Big) &= \frac{\mathbb{P}_\theta\Big(X_1 = x_1,\ldots,X_n = x_n\Big)}{\mathbb{P}_\theta\Big(S(x_1,\ldots,x_n) = t\Big)} \\
&= \frac{g\Big(S(\mathbb{X}),\theta\Big)h(\mathbb{X})}{g\Big(S(\mathbb{X}),\theta\Big)\displaystyle\sum_{S(x_1,\ldots,x_n)=t} h(\mathbb{X})} \\
&= \frac{h(\mathbb{X})}{\displaystyle\sum_{S(x_1,\ldots,x_n)=t} h(\mathbb{X})}
\end{aligned}$$

que no depende de θ, y por definición se tiene que $S(\mathbb{X})$ es suficiente.

□

Ejemplo **2.5.3.** Sea $x_1, x_2, ..., x_n$ muestra aleatoria i.i.d con distribución $N(\mu,\sigma^2)$. La función de densidad de la muestra $\mathbb{X}$ es

$$\begin{aligned}
f(\mathbb{X},\theta) &= \prod_{i=1}^{n} \left(2\,\pi\,\sigma^2\right)^{-1/2}\ exp\left(-(x_i-\mu)^2\Big/\left(2\,\sigma^2\right)\right) \\
&= \left(2\,\pi\,\sigma^2\right)^{-n/2} exp\left(-\sum_{i=1}^{n}(x_i-\mu)^2\Bigg/\left(2\,\sigma^2\right)\right) \\
&= \left(2\,\pi\,\sigma^2\right)^{-n/2} exp\left(-\sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\mu)^2\Bigg/\left(2\,\sigma^2\right)\right) \\
&= \left(2\,\pi\,\sigma^2\right)^{-n/2} exp\left(\left(-\sum_{i=1}^{n}(x_i-\bar{x})^2 - 2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\bar{x}) - \sum_{i=1}^{n}(\bar{x}-\mu)^2\right)\Bigg/\left(2\,\sigma^2\right)\right)
\end{aligned}$$

$$= \left(2\,\pi\,\sigma^2\right)^{-n/2} exp\left(\left(-\sum_{i=1}^{n}(x_i-\bar{x})^2 - 2(\bar{x}-\mu)\left(\sum_{i=1}^{n}x_i - \sum_{i=1}^{n}\bar{x}\right) - \sum_{i=1}^{n}(\bar{x}-\mu)^2\right)\Big/\left(2\,\sigma^2\right)\right)$$

$$= \left(2\,\pi\,\sigma^2\right)^{-n/2} exp\left(\left(-\sum_{i=1}^{n}(x_i-\bar{x})^2 - 2(\bar{x}-\mu)\,(n\bar{x}-n\bar{x}) - n(\bar{x}-\mu)^2\right)\Big/\left(2\,\sigma^2\right)\right)$$

$$= \left(2\,\pi\,\sigma^2\right)^{-n/2} exp\left(-\left(\sum_{i=1}^{n}(x_i-\bar{x})^2 + n\,(\bar{x}-\mu)^2\right)\Big/\left(2\,\sigma^2\right)\right)$$

$$= (2\,\pi\,\sigma^2)^{-n/2} exp\left(-\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{(2\,\sigma^2)}\right) exp\left(-\frac{n(\bar{x}-\mu)^2}{(2\,\sigma^2)}\right)$$

Si μ es una constante conocida, por el Teorema 2.5.1, se tiene que $\sum_{i=1}^{n}(x_i-\bar{x})^2$ es un estadístico suficiente (para σ^2). Además

$$f(\mathbb{X},\theta) = (2\,\pi\,\sigma^2)^{-n/2} exp\left(-\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{(2\,\sigma^2)}\right) exp\left(-\frac{n(\bar{x}-\mu)^2}{(2\,\sigma^2)}\right)$$

$$= (2\,\pi\,\sigma^2)^{-n/2} exp\left(-\frac{ns^2}{2\sigma^2}\right) exp\left(-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right).$$

El hecho de que $f(\mathbb{X},\theta)$ se puede expresar como el producto de la función $h(\mathbb{X}) = (2\,\pi\,\sigma^2)^{-n/2} exp\left(-\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{(2\,\sigma^2)}\right)$, la cual no depende del parámetro μ con $g(\bar{X},\mu) = exp\left(-\frac{n(\bar{x}-\mu)^2}{(2\,\sigma^2)}\right)$ implica, por el Teorema 2.5.1, que si σ^2 es conocida, entonces $\bar{X}$ es un estadístico suficiente (para μ).

Estadística Suficiente Minimal

En la sección anterior se vio que las estadísticas suficientes no son únicas lo cual lleva a preguntarse: ¿es posible encontrar la mejor estadística suficiente? y la respuesta es sí, siempre es posible hallar la estadística que reduce al *máximo* la información, la que se define:

Definición **2.5.2.** Se dice que la estadística $S(\mathbb{X})$ es *suficiente minimal* si y sólo si es función de cualquier otra estadística suficiente $T(\mathbb{X})$.

Es decir que si T es una estadística suficiente minimal y $T(x) = T(y)$ entonces $S(x) = S(y)$. Para entender mejor este concepto obsérvese que todo estadístico determina una partición del espacio muestral y, recíprocamente, dada una partición del espacio muestral, asignando un valor a cada clase se tiene definido un estadístico. En este sentido, una estadística S es suficiente minimal si para cualquier otra estadística T y todo conjunto $B_t = \{x : Tx = t\}$ formado en la partición inducida por $T(x)$, existe $A_s = \{x : Sx = s\}$ en la partición inducida por $S(x)$ tal que $B_t \subseteq A_s$. Por lo que el estadístico suficiente minimal es único salvo transformaciones biunívocas, ya que no se específica el valor común del estimador a cada conjunto de la partición, sólo indica que deber el mismo para cada elemento del conjunto y diferente para elementos de conjuntos distintos de la partición.

En general, no es sencillo encontrar una estadística suficiente minimal usando la Definición 2.5.2. El siguiente teorema proporciona una herramienta más sencilla para éste propósito.

Teorema **2.5.2.** *Sea $X_1, X_2, ..., X_n$ muestra aleatoria de una población cuya distribución pertenece a $\mathbb{F} = \{F_\theta \mid \theta \in \Theta\}$ y se representa por $f_\theta(x_1, x_2, \ldots, x_n)$ la función de probabili-*

dad o la función de densidad de la muestra, según si la población es discreta o continua. T es un estadístico minimal suficiente si y sólo si

$$\begin{cases} T(x_1,\dots,x_n) = T(x'_1,\dots,x'_n) & \textit{si } \dfrac{f_\theta(x_1,\dots,x_n)}{f_\theta(x'_1,\dots,x'_n)} \textit{ no depende de } \theta, \\ \\ T(x_1,\dots,x_n) \neq T(x'_1,\dots,x'_n) & \textit{si } \dfrac{f_\theta(x_1,\dots,x_n)}{f_\theta(x'_1,\dots,x'_n)} \textit{ depende de } \theta. \end{cases}$$

Demostración. Sólo se probará este resultado para poblaciones discretas. Si se supone que T cumple las condiciones del teorema, entonces si $T(x'_1,x'_2,\dots,x'_n) = t$,

$$\mathbb{P}_\theta\,(X_1 = x'_1, X_2 = x'_2,\dots,X_n = x'_n \mid T = t) = \frac{\mathbb{P}_\theta\left(X_1 = x'_1, X_2 = x'_2,\dots,X_n = x'_n\right)}{\mathbb{P}_\theta(T = t)}$$

$$= \frac{f_\theta(x'_1,x'_2,\dots,x'_n)}{\displaystyle\sum_{A_t} f_\theta(x_1.x_2,\dots,x_n)} = \left(\sum_{A_t} \frac{f_\theta(x_1,x_2,\dots,x_n)}{f_\theta(x'_1,x'_2,\dots,x'_n)}\right)^{-1}$$

y, como $\dfrac{f_\theta(x_1,x_2,\dots,x_n)}{f_\theta(x'_1,x'_2,\dots,x'_n)}$ es independiente de θ en A_t, resulta que

$$\mathbb{P}_\theta\,(X_1 = x'_1, X_2 = x'_2,\dots,X_n = x'_n \mid T = t)$$

no depende de θ y T es suficiente.

Si T' es otro estadístico suficiente y $T'(x_1.x_2,\dots,x_n) = T'(x'_1.x'_2,\dots,x'_n) = t'$,

$$\mathbb{P}_\theta\,(X_1 = x_1, X_2 = x_2,\dots,X_n = x_n \mid T' = t') = \frac{f_\theta(x_1,x_2,\dots,x_n)}{\mathbb{P}_\theta(T' = t')}$$

será independiente de θ, lo mismo que

$$\mathbb{P}_\theta\,(X_1 = x'_1, X_2 = x'_2,\dots,X_n = x'_n \mid T' = t') = \frac{f_\theta(x'_1,x'_2,\dots,x'_n)}{\mathbb{P}_\theta(T' = t')}.$$

Dividiendo término a término resulta que $\dfrac{f_\theta(x_1,x_2,\ldots,x_n)}{f_\theta(x_1',x_2',\ldots,x_n')}$ no depende de θ; por lo tanto $T(x_1,x_2,\ldots,x_n) = T'(x_1',x_2',\ldots,x_n')$, entonces T es función de T'. Por lo que T es minimal suficiente.

Por otro lado, supóngase que T' es minimal suficiente,

$$\begin{cases} f_\theta(x_1,x_2,\ldots,x_n) = g_\theta(T'(x_1,x_2,\ldots,x_n))h(x_1,x_2,\ldots,x_n), \\ \\ f_\theta(x_1',x_2',\ldots,x_n') = g_\theta(T'(x_1',x_2',\ldots,x_n'))h(x_1',x_2',\ldots,x_n'). \end{cases}$$

Así, si $\dfrac{f_\theta(x_1,x_2,\ldots,x_n)}{f_\theta(x_1',x_2',\ldots,x_n')}$ depende de θ, también dependerá de θ el cociente $\dfrac{g_\theta(T'(x_1,x_2,\ldots,x_n))}{g_\theta(T'(x_1',x_2',\ldots,x_n'))}$; por lo que $T'(x_1,\ldots,x_n) \neq T'(x_1',\ldots,x_n')$. Además, si T es un estadístico que cumple las condiciones del teorema, tiene que ser $T' = \varphi(T)$. Luego si $\dfrac{f_\theta(x_1,x_2,\ldots,x_n)}{f_\theta(x_1',x_2',\ldots,x_n')}$ no depende de θ, será $T(x_1,\ldots,x_n) = T(x_1',\ldots,x_n')$ y $T'(x_1,\ldots,x_n) = T'(x_1',\ldots,x_n')$. □

Ejemplo **2.5.4.** Para una muestra de tamaño n de una población de Bernoulli(θ), $\sum_{i=1}^{n} X_i$ es un estadística minimal suficiente, ya que

$$\frac{f(x_1,x_2,\ldots,x_n)}{f(y_1,y_2,\ldots,y_n)} = \left(\frac{\theta}{1-\theta}\right)^{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i}$$

y este cociente es independiente de θ si y sólo si $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$.

Ejemplo **2.5.5.** Sea $X_1,X_2,...,X_n$ muestra aleatoria $U(0,\theta)$ $\theta > 0$. Sean $\mathbb{X}$ e $\mathbb{Y}$ dos elementos distintos de $\mathbb{H}$. El cociente de las densidades de éstos es:

$$
\begin{aligned}
\frac{f(\mathbb{X};\theta)}{f(\mathbb{Y},\theta)} &= \frac{(1/\theta^n)\prod_{i=1}^{n}\mathbb{1}_{(0,\theta)}(x_i)}{(1/\theta^n)\prod_{i=1}^{n}\mathbb{1}_{(0,\theta)}(y_i)} \\
&= \frac{\mathbb{1}_{(0,\theta)}(x_n)}{\mathbb{1}_{(0,\theta)}(y_n).}
\end{aligned}
$$

donde $x_n = \text{máx}(X_1, X_2, ..., X_n)$ y $y_n = \text{máx}(Y_1, Y_2, ..., Y_n)$.

Se puede observar que si $x_n = y_n$ entonces,

$$\frac{f(\mathbb{X};\theta)}{f(\mathbb{Y},\theta)} = 1$$

lo cual, claramente, no depende de θ y por lo tanto $x_n = \text{máx}(X_1, X_2, ..., X_n)$ es una estadística suficiente minimal.

Estadísticas Completas

En la sección anterior se vio que la suficiencia resume la información de la muestra relativa a un parámetro θ de manera eficiente y sin perder información. Si a la suficiencia añadimos la propiedad de *completitud* se obtienen mejores estimadores.

Definición **2.5.3.** Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de $f(X,\theta)$. Sea $T(\mathbb{X})$ una estadística. La familia de densidades de $T(\mathbb{X})$ es completa si cada vez que $\mathbb{E}[g(T(\mathbb{X}))] = 0$ implica que $\mathbb{P}[g(T(\mathbb{X})) = 0] = 1$.

Es decir, $T(\mathbb{X})$ es completa si el único estimador insesgado para 0 es aquel que toma el valor 0 con probabilidad 1.

Ejemplo **2.5.6.** Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria que proviene de una distribución *Bernoulli*(θ). Se quiere probar que $T(\mathbb{X}) = \sum_{i=1}^{n} X_i$ es completa. Para ello se supondrá que $\mathbb{E}[g(t)] = 0$. El objetivo es probar que $g(t) = 0$ con probabilidad 1 para todo $\theta \in (0,1)$.

Como X_1 es una v.a Bernoulli de parámetro θ, entonces $\sum_{i=1}^{n} X_i$ tiene distribución Binomial (n,p). Entonces,

$$
\begin{aligned}
0 = \mathbb{E}[g(t)] &= \sum_{t=0}^{n} g(t) \binom{n}{t} \theta^t (1-\theta)^{n-t} \\
&= (1-\theta)^n \sum_{t=0}^{n} g(t) \binom{n}{t} \frac{\theta^t}{(1-\theta)^t} \\
&= \sum_{t=0}^{n} g(t) \binom{n}{t} \frac{\theta^t}{(1-\theta)^t} \\
&= \sum_{t=0}^{n} g(t) \binom{n}{t} \left(\frac{\theta}{1-\theta}\right)^t .
\end{aligned}
$$

Como $0 < \dfrac{\theta}{1-\theta} < \infty$, el polinomio es igual a cero si y sólo si $g(t) = 0$ para todo $t = 0, \ldots, n$.

Por lo tanto, $T(\mathbb{X}) = \sum_{i=1}^{n} X_i$ es una estadística completa.

Ejemplo **2.5.7.** Sea $X_1, X_2, \ldots, X_n$ muestra aleatoria Uniforme $(0,\theta)$. $Y_n = \text{máx}(X_1, X_2, \ldots, X_n)$ es una estadística completa.

En efecto, como en el ejercicio anterior si se supone que $\mathbb{E}[g(Y_n)] = 0$.

$$
\begin{aligned}
0 &= \mathbb{E}[g(t)] \\
&= \int_0^n g(y_n)\, f_{Y_n}(y_n)\, dy_n \\
&= \int_0^n g(y_n)\, n \left(\frac{y_n}{\theta}\right)^{n-1} \frac{1}{\theta}\, dy_n \\
&= \int_0^n \frac{n}{\theta^n}\, g(y_n)\, y_n^{n-1}\, dy_n \\
0 &= \int_0^n g(y_n)\, y_n^{n-1}\, dy_n.
\end{aligned}
$$

Derivando de ambos lados con respecto a θ se obtiene que: $g(\theta)\theta^{n-1}$. Pero $\theta > 0$, por lo tanto $g(\theta) = 0$. Entonces $Y_n = \text{máx}(X_1, X_2, \ldots, X_n)$ es una estadística completa.

Ejemplo **2.5.8.** Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria Binomial(n, θ). ¿Será $T(\mathbb{X}) = X_1 - X_2$ una estadística completa?

Véase que $T(\mathbb{X}) = X_1 - X_2$ se distribuye de la siguiente forma:

Sea $g(t) = t$

$$\mathbb{E}[g(T)] = (-1)[(1-\theta)\,\theta] + (0)\left[\theta^2 + (1-\theta)^2\right] + (1)[\theta\,(1-\theta)] = 0.$$

Sin embargo, $g(T) \neq 0$ con probabilidad positiva. Por lo tanto, $T(\mathbb{X}) = X_2 - X_1$ no es una estadística completa.

Se puede calcular el UMVUE cuando se restringe a los estadísticos que sean función del estadístico suficiente, como demuestra el siguiente teorema [7]:

Teorema **2.5.3.** *(Rao-Blackwell) Sea S un estadístico suficiente de $q(\theta)$ y $T \in C_q$, entonces $\mathbb{E}[T \mid S] \in C_q$ y se cumple que*

$$V_\theta(\mathbb{E}[T \mid S]) \leq V_\theta[T]$$

para todo θ. Además la igualdad se alcanza si y sólo si $\mathbb{E}[T \mid S]$ coincide $\mathbb{P}_\theta$-casi seguro con T respecto a todo θ.

Este teorema asegura que, dado un estimador insesgado y un estadístico suficiente, éste se puede utilizar para hallar otro estimador insesgado de menor varianza, como consecuencia de este teorema el UMVUE, si existe, es único.

Proposición **2.5.1.** *Si T es UMVUE para $q(\theta)$, es único en probabilidad salvo un conjunto de medida nula.*

Demostración. Sea T es UMVUE para $q(\theta)$ y considérese que existe otro estimador S insesgado con la misma varianza mínima. Entonces $V_\theta(T) = V_\theta(S)$ para todo θ. El estimador $\dfrac{T+S}{2}$ también es insesgado y su varianza es

$$V_\theta\left(\frac{T+S}{2}\right) = \frac{1}{4}V_\theta(T) + \frac{1}{4}V_\theta(S) + \frac{1}{2}Cov(T,S) \leq$$

$$\frac{1}{4}V_\theta(T) + \frac{1}{4}V_\theta(S) + \frac{1}{2}\sqrt{V_\theta(T)V_\theta(S)} = \left(\frac{1}{4} + \frac{1}{4} + \frac{1}{2}\right)V_\theta(T) = V_\theta(T).$$

La desigualdad no puede ser estricta porque V_θ es el UMVUE, sin embargo la igualdad se da si y sólo si $Corr(T,S) = 1$, es decir $T = a(\theta)S + b(\theta)$, para algunas funciones $a(\theta)$ y $b(\theta)$. En ese caso, $\mathbb{E}[T] = a(\theta)\mathbb{E}[S] + b(\theta)$.

Por otra parte

$$V_\theta(T) = \sqrt{V_\theta(T)V_\theta(S)} = Cov(T,S) = Cov\,(a(\theta)\,S + b\,(\theta),S) =$$

$$cov\,(a\,(\theta)\,S,S) = a\,(\theta)\,cov\,(S,S) = a\,(\theta)\,V_\theta(T)$$

por lo que $a\,(\theta) = 1$, $b\,(\theta) = 0$ y $T = S$. □

Cuando se analiza la demostración del Teorema 2.5.3, ésta se reduce a ver que $Var_\theta[T] = Var(E[T \mid S]) + \mathbb{E}_\theta(Var\mathbb{E}[T \mid S])$; donde el segundo sumando del lado derecho es no negativo y por lo tanto la varianza del estimador $\mathbb{E}[T \mid S]$ es siempre menor o igual que la varianza del estimador T.

Aun si es posible reducir la varianza no se puede asegurar que se alcanza el UMVUE. Sin embargo al añadir la completitud se tiene

Teorema 2.5.4. *(Lehmann-Scheffe)Dada la familia de distribuciones de probabilidad de la muestra* $\mathbb{P} = \{f(x_1, x_2, \ldots, x_n \mid \theta) : \theta \in \Theta\}$*, si para estimar* $q(\theta)$ *la clase* C_q *es no vacía y existe un estadístico* $S(X_1, X_2, \ldots, X_n)$ *suficiente y completo, entonces* $\mathbb{E}[T \mid S]$*, para cualquier* T *perteneciente a* C_q*, es el UMVUE para* $q(\theta)$*.*

Demostración. Como $\mathbb{E}[T \mid S]$ es una función, $f(S)$, del estadístico S, si existiese otro estimador $f'(S)$, centrado para $g(\theta)$, sería

$$\mathbb{E}_\theta[f(S) - f'(S)] = 0 \text{ para cualquier } \theta \in \Theta.$$

En virtud de la completitud de S, tendría que ser $f'(S) = f(S)$, $\mathbb{P}_\theta$-casi seguro para todo $\theta \in \Theta$.

Puesto que el UMVUE de $g(\theta)$ ha de ser función de S, tiene que coincidir con $\mathbb{E}[T \mid S]$ $\mathbb{P}_\theta$-casi seguro para cada $\theta \in \Theta$. □

Esto asegura que, si existe un estadístico completo y suficiente, el estadístico función de él e insesgado para la función a estimar, es el estimador óptimo en la clase de los estimadores insesgados. Dicho de otra forma:

Corolario **2.5.2.** *Si $T(X)$ es un estadístico suficiente y completo, cualquier función suya que tenga esperanza finita es el UMVUE de su esperanza.*

Cuando existe un estadístico suficiente y completo, la cuestión de la existencia de UMVUE queda completamente cerrada: las funciones del parámetro, para las que exista algún estimador insesgado, tendrán UMVUE; las demás, obviamente, no lo tendrán. Sin embargo, se debe recalcar que si no existen estadísticos suficientes completos puede o no haber funciones del parámetro que posean UMVUE.

Ejemplo **2.5.9.** Determinar con una muestra aleatoria simple de tamaño n de una población de $Bernoulli(\theta)$ el UMVUE para $q(\theta) = \theta(1-\theta)$.

De los Ejemplos 2.5.2 y 2.5.6 se sabe que la estadística $S = \sum_{i=1}^{n} X_i$ es un estadístico suficiente y completo, además

$$\mathbb{E}_\theta\left[\sum_{i=1}^{n} X_i\right] = n\theta$$

implica que $\sum_{i=1}^{n} X_i$ es función del estadístico suficiente y completo, luego es el óptimo para su media, θ. Así, para $q(\theta) = \mathbb{P}\{X_1 = 1, X_2 = 0 \mid \theta\}$, todo se reduce a encontrar $\mathbb{E}[T \mid S]$ con

$$T(X_1, X_n, \ldots, X_n) = \begin{cases} 1 & \text{si } X_1 = 1 \text{ y } X_2 = 0 \\ 0 & \text{en otro caso.} \end{cases}$$

Como

$$\mathbb{E}\left[T\left|\sum_{i=1}^{n}X_i=s\right.\right]=\frac{\mathbb{P}\left(X_1=1,X_2=0|\theta\right)\mathbb{P}\left(\sum_{i=3}^{n}X_i=s-1\middle|\theta\right)}{\mathbb{P}\left\{\sum_{i=1}^{n}X_i=s\middle|\theta\right\}}$$

se sigue, utilizando que $\sum_{i=3}^{n}X_i$ se distribuye según una *Binomial(n-2,θ)*, que

$$\mathbb{E}\left[T\left|\sum_{i=1}^{n}X_i=s\right.\right]=\frac{s(n-s)}{n(n-1)}.$$

Con lo que

$$\frac{\sum_{i=1}^{n}X_i\left(n-\sum_{i=1}^{n}X_i\right)}{n(n-1)}$$

es el UMVUE para $q(\theta)=\theta(1-\theta)$.

Ejemplo **2.5.10.** Sea $\left(X_1,\ldots,X_n\right)$ una muestra aleatoria simple de una distribución exponencial de parámetro θ, el estadístico $S=\sum_{i=1}^{n}X_i$ es completo. En efecto, como su distribución es $\gamma(n,\theta)$, entonces

$$\mathbb{E}_\theta\left[f(S)\right]=\int_0^\infty f(x)\frac{\theta^n x^{n-1}e^{-\theta x}}{(n-1)!}dx.$$

Por lo tanto, $\mathbb{E}_\theta\left[f(S)\right]=0$ para cada $\theta>0$, significa que la transformada de Laplace:

$$\int_0^\infty f(x)x^{n-1}e^{-\theta x}dx,$$

de la función $f(x)x^{n-1}$ se anula. Como la transformada de Laplace es única, ello es posible sólo si $f(x)x^{n-1}=0$ o bien $f(x)=0$ para todo $x>0$.

Por otro lado

$$f_\theta(x_1,\dots,x_n) = \theta^n e^{-\theta(x_1+\cdots+x_n)} = g_\theta\left(\sum_{i=1}^n x_i\right),$$

haciendo $h(x_1,\dots,x_n)=1$, por el Teorema 2.5.1 se tiene que $S=\sum_{i=1}^n X_i$ es suficiente.

Ya que $\mathbb{E}\left[\sum_{i=1}^n X_i\right] = \dfrac{n}{\theta}$ se sigue que $\bar{X} = \dfrac{1}{n}\sum_{i=1}^n X_i$ es el UMVUE para la media poblacional $\dfrac{1}{\theta}$, puesto que no puede haber otro estimador insesgado de $\dfrac{1}{\theta}$, función de S.

2.6 Ejercicios

1. Suponga que X es una variable aleatoria discreta con la siguiente distribución, donde $0 \leq \theta \leq 1$

X	0	1	2	3
P(X)	$\frac{2\theta}{3}$	$\frac{\theta}{3}$	$\frac{2(1-\theta)}{3}$	$\frac{(1-\theta)}{3}$

La siguiente muestra tiene la misma distribución que X

$$3 \quad 0 \quad 2 \quad 1 \quad 3 \quad 2 \quad 1 \quad 0 \quad 2 \quad 1$$

Encuentre el estimador máximo verosímil para θ.

2. Se estima cada uno de 150 artículos recién fabricados y se anota el número de rayones por artículo (se supone que los artículos están libres de rayones) y se obtienen los siguientes datos:

Número de rayones por artículo	0	1	2	3	4	5	6	7
Frecuencia observada	18	37	42	30	13	7	2	1

Sea X el número de rayones en un artículo seleccionado al azar y suponga que X tiene una distribución *Poisson* con parámetro μ:

a) Encuentre un estimador insesgado para μ (utilice el método de máxima verosimilitud) y calcule el valor del estimador.

b) ¿Cuál es la desviación estándar del estimador? Calcule el error estándar estimado.

c) Encuentre el error cuadrático medio del estimador.

3. Sea X_1, X_2 ,X_3 ,..., X_n una muestra aleatoria con densidad $f(x;\theta) = \theta x^{-2}$

donde $\theta > 0$.

a) Encuentre un estimador por máxima verosimilitud para θ.

b) Encuentre el sesgo para el estimador obtenido.

c) Calcule θ a partir de las siguientes 10 observaciones de esfuerzo vibratorio de una espada de turbina en condiciones específicas: 16.88, 10.23, 4.59, 6.66, 13.68, 14.23, 19.87, 9.40, 6.51, 10.95.

d) Determine el estimador de máxima verosimilitud de la medina de la distribución del esfuerzo de vibración. (Sugerencia exprese primero la mediana en función de θ.)

4. Sea $X_1, X_2, X_3, \ldots, X_n$ una muestra aleatoria. Para cada una de las siguientes distribuciones:

$$f(x;\theta) = \theta x^{\theta-1} \text{ donde } \theta > 0;\ 0 < x < 1.$$

$$f(x;\theta) = \frac{\log\theta}{\theta - 1}\theta^x \text{ donde } \theta > 0;\ 0 < x < 1.$$

a) Encuentre una estadística suficiente y complete si es que existe.

b) ¿Existe una función de $\theta, g(\theta)$, para la cual exista un estimador insesgado cuya varianza sea igual a la cota inferior de Crammer y Rao? Si es así diga cual es. Si no, muestre por qué no existe.

5. La distribución gamma

$$f(x;\alpha,\beta) = \frac{\beta\alpha x^{\alpha-1}e^{-\beta x}}{\gamma(\alpha)};\ x > 0.$$

Tiene media $\dfrac{\alpha}{\beta}$ y varianza $\dfrac{\alpha}{\beta^2}$.

a) Si α es conocida, encuentra el estimador de máxima verosimilitud de β.

b) Encuentre el EMV de $\frac{1}{\beta}$. ¿Cuál es su media y su varianza? ¿Este estimador alcanza la Cota inferior de Crammer y Rao (CICR)?

c) Verifique que el EMV de $\frac{1}{\beta}$ obtenido en b) es suficiente. ¿El estimador de β también es suficiente?

6. En el siguiente ejercicio encuentre los estimadores que se piden utilizando la definición de estimación y lo visto en estadística descriptiva.

   a) Se selecciona una muestra aleatoria de 10 casas en un área en particular, cada una de las cuales se calienta con gas natural y se determina la cantidad de gas utilizada por cada casa durante el mes de enero. Las observaciones resultantes son 103, 156, 118, 89, 125, 147, 122, 109, 138, 99. Sea μ el consumo de gas promedio durante enero de todas las casas de área. Calcule una estimación de μ.

   b) Suponga que hay 10000 casas en esta área que utilizan gas natural para calefacción. Sea τ la cantidad total de gas consumido por todas estas casas durante enero. Calcule τ con los datos del inciso a). ¿Qué estimador utilizaste para calcular tu estimación?

   c) Use los datos del inciso a) para estimar p, la proporción de todas las casas que usaron por lo menos 100 terms.

   d) Proporcione una estimación puntual de la mediana de la población usada basada en la muestra del inciso a).¿Qué estimador utilizo?

7. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria $N(0,\theta)$ con $0 < \theta < \infty$. Muestre por

medio de alguna definición, que $T(\mathbb{X}) = \sum_{i=1}^{n} x_i^2$ es una estadística suficiente para θ.¿Es T una estadística completa?. De el **UMVUE** para $q(\theta) = \theta$.

8. Encuentre el estimador máximo verosímil para α en la densidad,

$$f(x;\alpha) = \frac{2(\alpha - x)}{\alpha^2}; \quad 0 < x < \alpha$$

para muestras aleatorias de tamaño 2. Encuentre además el estimador por momentos para α. ¿Cuál es el estimador máximo verosímil para la media poblacional?

9. Una muestra aleatoria de los montos de 10 reclamaciones en una compañía aseguradora, la cual provienen de una distribución $Gamma(r, \lambda)$, se da a continuación:

$$1,500 \;\; 6,000 \;\; 3,500 \;\; 3,800 \;\; 1,800 \;\; 5,500 \;\; 4,800 \;\; 4,200 \;\; 3,900 \;\; 3,000$$

Estime $\theta = (\mu, \lambda)$ por el método de momentos.

10. Una muestra aleatoria de 5 sueldos, en dolares, de personas que trabajan dentro de una empresa norteamericana se muestran a continuación:

$$500 \;\; 1,000 \;\; 1,500 \;\; 2,500 \;\; 4,500$$

Suponiendo que los sueldos dentro de la compañía siguen una distribución $LogNormal(\mu, \sigma^2)$,

a) . Estime $\theta = (\mu, \sigma^2)$ por máxima verosimilitud.

b) . Estime la probabilidad de que el sueldo de una persona seleccionada aleatoriamente exceda los 4500 dolares.

11. Un número desconocido N, de animales habitan en cierta población. Para obtener cierta información acerca del tamaño de la población, los ecologistas llevan a cabo el siguiente experimento:Primero capturan cierto número de animales, K, los cuales marcan de alguna manera,y los regresan a su lugar de origen. Después, sueltan a los animales permitiendo que se dispersen en dicha región y una nueva captura de tamaño n es realizada. Sea x_1 la variable aleatoria que denota el número de animales marcados en la segunda captura(Notemos que el tamaño de la muestra en este caso es 1). Asumiendo que el tamaño de la población entre las dos capturas permanece igual y que cualquier animal tiene la misma probabilidad de ser capturado,

   a) Determine la distribución de $X_1, f(x_1; N)$.

   b) Encuentre el máximo verosímil de N.(Debemos notar que , en este caso,N es un número entero positivo.

   c) Supongamos que la primera captura fue de 50 animales, los cuales fueron marcados y regresados a la población. Si la segunda captura consiste de 40 animales, de los cuales 4 están marcados. Determine el estimado del número total de animales en la región.

12. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población con función de densidad $Normal(\theta, 1)$, donde se sabe que μ debe ser no negativo; esto es, $\Theta = \{\mu \mid \mu \geq 0\}$.

   a) Halle el **EMV** para μ.

   b) Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población con función de densidad $Blli(\theta)$, para $0 \leq \theta \leq 1/2$; esto es, $\Theta = \{\theta \mid 0 \leq \theta \leq 1/2\}$. Halle el estimador máximo verosímil para θ.

13. Encontre el estimado, por máxima verosimilitud , de la función de densidad

$$f(x;\theta) = \frac{1}{2}e^{-|x-\theta|}; \quad -\infty < x < \infty; -\infty < \theta < \infty$$

para las siguientes muestras aleatorias:

a) $x_1 = -3$, $x_2 = 4$ y $x_3 = 6$

b) $x_1 = -4$, $x_2 = 0$, $x_3 = 1$ y $x_4 = 5$

14. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población con función de distribución dada por:

$$F(x;\theta_1,\theta_2) = 1 - \left(\frac{\theta_1}{x}\right)^{\theta_2}; \quad \theta_1 \leq x; \; \theta_1, \theta_2 \in (0,\infty)$$

Encuentre los estimadores máximo verosímil de θ_1 y θ_2.

15. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de la distribución Gaussiana Inversa cuya función de densidad está dada por:

$$f(x;\mu,\lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} exp\left[-\lambda(x-\mu)^2/2\mu^2 x\right], \quad x > 0.$$

Muestre que el estimador máximo verosímil de $\theta = (\mu, \lambda)$ es:

16. Sea $X_1, X_2, \ldots, X_n$ m.a de alguna densidad con media μ y varianza σ^2 finita.

a) Muestre que $\sum_{i=1}^{n} a_i X_i$ es un estimador insesgado de μ para cualquier conjunto de constantes conocidas $a_1, a_2, \cdots a_n$, tales que $\sum_{i=1}^{n} a_1 = 1$.

b) Si $\sum_{i=1}^{n} a_1 = 1$, muestre que $Var\left[\sum_{i=1}^{n} a_i X_i\right]$ es minimizado cuando $a_i = 1/n$, para toda $i = 1, 2, \cdots, n$.

$$HINT\text{: Muestre que } \sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n} (a_i - 1/n)^2 + 1/ncuando \sum_{i=1}^{n} a_i = 1.$$

17. Sea X_1, X_2 una muestra aleatoria de tamaño 2 de la distribución normal con media θ y varianza 1. Considere los siguientes estimadores de θ:

$$T_1(X_1, X_2) = \frac{1}{3}(2X_1 + X_2)$$

$$T_2(X_1, X_2) = \frac{1}{4}(X_1 + 3X_2)$$

$$T_3(X_1, X_2) = \frac{1}{2}(X_1 + X_2)$$

Muestre que T_i es insesgado para $i = 1, 2, 3$.

18. Supongamos que diariamente se cometen k asaltos dentro de una comunidad. Sea p la probabilidad de levantar el acta en la delegación correspondiente. Ambos parámetros (k y p) se desconocen y se desean estimar con base a una muestra aleatoria $X_1, X_2, \ldots, X_n$.

    a) Genere una muestra aleatoria de tamaño 1095 de variables aleatorias $Bin(k = 50, p = 0.4)$ que representará el número de asaltos reportados por día en los últimos 3 años(recordemos que estos parámetros en realidad no se conocen).

    b) Use la muestra aleatoria anterior y los resultados estudiados en el texto para encontrar el estimado por momentos para $\theta = (k, p)$.

19. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de la distribución $f(x, \theta)$.

a) Utilice el teorema de factorización para probar que el estimador máximo verosímil es función de una estadística suficiente.

b) Si $S(\mathbb{X})$ es una estadística suficiente entonces cualquier función inyectiva de $S(\mathbb{X})$ es una estadística suficiente.

20. Considere una muestra aleatoria de tamaño n de la densidad

$$f(x,\theta) = \frac{1}{2}\exp^{-|x-\theta|} \qquad ; x,\theta \in (-\infty,\infty)$$

Encuentre una estadística suficiente para θ.

Vea que

$$f(\mathbb{X},\theta) = \frac{1}{2}\exp^{-|x-\theta|} \prod_{i=1}^{n} \mathbb{1}_{(-\infty,\infty)} x_i$$

21. Si tenemos una muestra aleatoria de tamaño dos de

$$f(\mathbb{X},\theta) = \theta^{-1}\exp^{-x/\theta} \;\; \mathbb{1}_{(0,\theta)}(x)$$

Considere a $Y_1 = X_1 + X_2$ estadística suficiente. Pruebe que $Y_2 = X_2$ es un estimador insesgado para θ con varianza θ^2.

Encuentre $\phi(y_1) = \mathbb{E}[y_2/y_1]$ y determine la varianza y compare con la de y_2.

¿La función $\phi(y_1)$ es el **UMVUE** para θ?.

22. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de la distribución $f(\mathbb{X},\theta) = \frac{1}{\theta} \; \mathbb{1}_{(0,\theta)}(x)$

a) Calcule $\dfrac{1}{n\mathbb{E}[\frac{d}{d_\theta} lnf(x;\theta)^2]}$

b) En el inciso anterior se calculo la definición de la **CICR**,para la varianza de estimadores insesgados para $q(\theta) = \theta$. Ahora encuentre un estimador insesgado para $q(\theta) = \theta$ que sea función de y_n. Llámelo $G(y_n)$

c) Encuentre la varianza de $G(y_n)$ y compárela con la cantidad encontrada en el inciso *a)*. ¿contradice esto el teorema de la CICR?. Explique.

d) Encuentre el UMVUE para $q(\theta) = \theta$

23. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño $n \geq 2$ de una distribución con función de densidad $f(\mathbb{X}, \theta) = \theta \exp^{-\theta x} \;\; \mathbb{1}_{(0,\infty)}(x)$ y $\theta > 0$. Encuentre el mejor estimador insesgado y de menor varianza uniforme para θ.

24. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de la densidad:

$$f(x; \mu, \theta) = \theta^{-1} \exp^{-\frac{x-\mu}{\theta}} \;\; \mathbb{1}_{(\mu,\infty)}(x);\ \theta > 0,\ \mu \in \mathbb{R}$$

suponiendo θ conocida.

a) Encuentre el estimador máximo verosímil de μ

b) Encuentre el estimador de $q(\mu) = \mathbb{P}[X > x]$

25. La variable aleatoria X tiene la función de densidad definida por:

$$f(x; \alpha, \lambda) = \alpha \lambda^{\alpha} (\lambda + x)^{-\alpha - 1};\ \ x, \alpha, \lambda > 0$$

Se sabe que $\lambda = 1,000$. Se tienen las siguientes cinco observaciones: 43,145,233,396,775.

a) Determine por el método de momentos el estimador de α;

b) Encuentre el estimado de α.

26. Considerando una muestra aleatoria de tamaño n, contestar:

¿El estimador máximo verosímil (EMV) para λ en la f.d.p $f(x; \lambda) = \lambda^{-1} \exp^{-\frac{x}{\lambda}}$ con $x, \lambda > 0$, es insesgado para λ? Si no lo es proponga un estimador , función del EMV, que sí lo sea.

27. Sea X_1, X_2, X_3 una muestra aleatoria de una población con media μ y varianza σ^2.

   a) Sean $\mu_1 = \dfrac{2X_1 + 3X_3}{3}$, $\mu_2 = X_1 + X_2$ y $\mu_3 = \dfrac{X_1 + 3X_2 + X_3}{3}$ estimadores para la media μ. Verifique si alguno de ellos es un estimador insesgado.

   b) Encuentre el error cuadrático medio para cada uno.

   c) Basándose en μ_1, μ_2 y μ_3 encuentre un estimador insesgado para μ. Determine cuál es el estimador más eficiente para μ.

   d) Encuentre un estimador suficiente para μ.

   e) Determine la CICR para μ_1 y para el estimador encontrado en el inciso anterior. Comente.

28. (para este ejercicio inciso a, hay que usar lo de la familia exponencial) Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de la distribución $N(0, \dfrac{1}{\theta})$, $\theta > 0$.

   a) Utilice el teorema de Lehmann-Scheffé para demostrar que el UMVUE para θ es $T^*(\mathbb{X}) = \dfrac{n-2}{\sum_{i=1}^{n} X_i^2}$

   b) Determine la cota inferior de Cramer y Rao para estimadores insesgados de θ.

29. Muestre que cada una de las siguientes familias no son completas para $\theta > 0$

   a) $f(x;\theta) = \dfrac{1}{2\theta} \ \mathbb{1}_{(-\theta,\theta)}(x)$

   b) $f(x,\theta) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$

30. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño n de la densidad $Beta(\theta, 1)$.

a) Demuestre que $\prod_{i=1}^{n} X_i$ es una estadística suficiente.

b) ¿La estadística $T(\mathbb{X}) = -\sum_{i=1}^{n} ln(X_i)$ es suficiente para θ?

c) Encuentre el UMVUE para $q(\theta) = \theta^{-1}$ y para $q(\theta) = \theta$

31. Sea X la proporción de tiempo asignado que un estudiante seleccionado al azar pasa resolviendo cierta prueba de aptitud. Suponga que la función de densidad de probabilidad de X es

$$f(X;\theta) = \begin{cases} (\theta+1)x^{\theta} & si \quad 0 \leqslant x \leqslant 1 \\ \\ 0 & e.o.c \end{cases}$$

donde $-1 < \theta$. Una muestra aleatoria de diez estudiantes produce los datos

$$x_1 = 0.92, x_2 = 0.79, x_3 = 0.90, x_4 = 0.65, x_5 = 0.86,$$

$$x_6 = 0.47, x_7 = 0.73, x_8 = 0.97, x_9 = 0.94, x_{10} = 0.77$$

a) Use el método de momentos para obtener un estimador de θ y luego calcula la estimación para esos datos.

b) Obtenga el estimador de Máxima Verosimilitud de θ y luego calcule la estimación para los datos dados.

32. Para las siguientes familias, suponga que se tiene una muestra aleatoria de tamaño n . ¿Existe una función de θ, para la cual exista un estimador insesgado, cuya varianza alcance la CICR? de no ser así, muestre porque no.

a) $f(x;\theta) = \theta(1-x)^{\theta-1}; \;\; 0 < x < 1;$

b) $f(x;\theta) = \dfrac{log(\theta)}{\theta - 1}\theta^x;\ \ 0 < x < 1,\ \theta > 1;$

c) $f(x;\theta) = \dfrac{\theta}{x^2};\ \ x > \theta,\ \theta > 0;$

33. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución $N(0,1)$.

    a) Encuentre la cota inferior de Cramer y Rao para estimadores insesgados de $\theta, \theta^2 y \mathbb{P}[x > 0]$.

    b) ¿Existe un estimador insesgado para θ^2?. Si es así encuéntrelo.

    c) ¿Existe un estimador insesgado para $\mathbb{P} = [x > 0]$?. Si es así, encuéntrelo.

    d) Demostrar que el **UMVUE** para $q(\theta) = \theta^2$ es $\bar{X}^2 - 1/n$.

34. Muestre que la mínima estadística de orden Y_1, es una estadística suficiente para una muestra aleatoria de tamaño n de una población que tiene función de densidad $f(x;\theta) = e^{-}(x - \theta)$, con $\theta < x < \infty$, donde $\theta \in \mathbb{R}$. Se puede probar que dicha estadística es minimal y completa. Encuentre la única función de esta estadística que sea el mejor estimador para θ(**UMVUE**).

35. Existen dos instrumentos disponibles para medir una distancia particular. La variable aleatoria X representa la medida con el primer instrumento, y la variable aleatoria Y la medida con el segundo instrumento. Asuma que X y Y son independientes con $\mathbb{E}(X) = 0.8m$, $Var(X) = m^2$, y $\mathbb{E}(Y) = m$, $Var(Y) = 1.5m^2$, donde m es la distancia verdadera. Considere estimadores de m de la forma $Z = aX + bY$. Determine los valores de a y b que hagan de Z un **UMVUE** para m.

Apéndice

FAMILIA DE PARAMÉTRICAS DISCRETAS

Nombre	$P_X(x)$	$\mathbb{E}[x]$	$\mathrm{Var}\big(X\big)$	$\mathrm{M}_X(t)$	Modelo
Uniforme $X \sim U\{n\}$ $x = 1, 2, 3, \ldots, n$ $n \in \mathbb{N}$	$\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\sum_{x=0}^{n} \frac{1}{n} e^{xt}$	Eventos equiprobables
Bernoulli $X \sim Blli(p)$ $x = 0, 1$	$p^x(1-p)^{1-x}$	p	$p(1-p)$	$(1-p) + pe^t$	Un ensayo Bernoulli
Binomial $X \sim Bin(n, p)$ $x = 0, 1, 2 \ldots, n$	$\binom{n}{x} p^x (1-p)^{n-x}$	np	$np(1-p)$	$\big((1-p) + pe^t\big)^n$	n ensayos Bernoulli
Poisson $X \sim Poi(\lambda)$ $x = 0, 1, 2 \ldots$ $\lambda > 0$	$\frac{\lambda^x e^{-\lambda}}{x!}$	λ	λ	$e^{\lambda\left(e^t - 1\right)}$	Eventos en un intervalo de tiempo o espacio
Geométrica $X \sim Geo(p)$ $x = 1, 2, \ldots$	$p(1-p)^{x-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$	Número de ensayos hasta el primer éxito
Geométrica $X \sim Geo(p)$ $x = 0, 1, 2, \ldots$	$p(1-p)^x$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$	$\frac{p}{1-(1-p)e^t}$	Número de ensayos hasta el k-ésimo éxito
Binomial Negativa $X \sim BinNeg(k, p)$ $x = k, k+1, k+2, \ldots$	$\binom{x-1}{k-1} p^k (1-p)^{x-k}$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$	$\left(\frac{pe^t}{1-(1-p)e^t}\right)^k$	Número de ensayos hasta el k-ésimo éxito
Binomial Negativa $X \sim BinNeg(k, p)$ $x = 0, 1, 2, \ldots$	$\binom{x+k-1}{k-1} p^k (1-p)^x$	$\frac{k(1-p)}{p}$	$\frac{k(1-p)}{p^2}$	$\left(\frac{p}{1-(1-p)e^t}\right)^k$	Número de fracasos antes del k-ésimo éxito
Hipergeométrica $X \sim Hip(N, m, n)$ $x = 0, 2, 3, \ldots, n*$	$\frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$	$\frac{nm}{N}$	$\frac{nm}{N}\left(\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}\right)$		Número de elementos característicos dentro de la muestra
$*\text{máx}\{0, n-(N-m)\} \le x \le \text{mín}\{n, m\}$					

FAMILIA DE PARAMÉTRICAS CONTINUAS

Nombre	$F_X(x)$	$f_X(x)$	$\mathbb{E}[X]$	$\text{Var}(X)$	$M_X(t)$	Modelo
Uniforme $X \sim U(a,b)$ $x \in [a,b]$	$\frac{x-a}{b-a}$	$\frac{1}{b-a}$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$	Un número entre a y b
Exponencial $X \sim exp(\lambda)$ $x \in [0,\infty)$ $\lambda > 0$	$1-e^{-\lambda x}$	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda - t}$	(Magnitud) tiempo transcurrido hasta la primer ocurrencia de un evento tipo Poisson
Gamma $X \sim \Gamma(n,\lambda)$ $s \in [0,\infty)$ $\lambda > 0$	$1-e^{-\lambda x}\sum_{k=0}^{n-1}\frac{(\lambda x)^k}{k!}$	$\frac{1}{\Gamma(n)}\lambda e^{-\lambda x}(\lambda x)^{n-1}$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$	$\left(\frac{\lambda}{\lambda - t}\right)^n$	(Magnitud) tiempo transcurrido hasta la k-ésima ocurrencia de un evento tipo Poisson
Normal $X \sim N(\mu,\sigma^2)$ $x \in (-\infty,\infty)$ $\mu \in \mathbb{R}; \sigma > 0$	...	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^x}{2\sigma^2}}$	μ	σ^2	$e^{\mu t + \frac{\sigma^2 t^2}{2}}$	Una de las distribuciones más importantes de la teoría de la probabilidad
Cauchy $X \sim Cauchy(\theta)$ $x \in (-\infty,\infty)$ $\theta \in \mathbb{R}$	...	$\frac{1}{\pi(1+(x-\theta)^2)}$	∞	...	...	Distintas aplicaciones en la física
Beta $X \sim B(a,b)$ $x \in [0,1]$ $a,b \in \mathbb{R}$	...	$\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$		...
Weibull $X \sim W(\alpha,\beta,\nu)$ $x \in (\nu,\infty)$	$1-e^{-\left(\frac{x-\nu}{\alpha}\right)^\beta}$	$\frac{\beta}{\alpha}\left(\frac{x-\nu}{\alpha}\right)^{\beta-1}e^{-\left(\frac{x-\nu}{\alpha}\right)^\beta}$	...	...	...	Tiempo de vida de un artefacto con varios componentes

$\Gamma(a) = \int_0^\infty e^{-x}x^{-a-1}dx$	$\Gamma(a) = (a-1)\Gamma(a-1)$	$\forall n \in \mathbb{N}, \Gamma(n) = (n-1)!$
$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$		$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

Bibliografía

[1] Boente Gracielas ,Yohai Víctor , *Notas de estadística*.

[2] Casas Sánchez José, *Inferencia estadística para economía y administración de empresas*. Centro de estudios Ramón Areces, S. A.

[3] Casella George, L.Berger Roger ,*Statistical Inference*, 2nd ed., Edit.Duxbury, E.U.A 2002.

[4] Gómez Villegas Miguel Ángel , *Inferencia estadística*, Facultad de Ciencias Matemáticas de la Universidad Complutense de Madrid, Edit. Diaz de Santos

[5] Luis Rincón, *Curso Elemental de Probabilidad y Estadística*, Departamento de Matemáticas, Facultad de Ciencias, 2007.

[6] Roussas George G.,*Introduction to Probability and Statistical Inference*, Edit. Academic Press, E.U.A 2003.

[7] Vélez Ibarrola Ricardo , García Pérez Alfonso , *Principios de inferencia estadística*, Universidad nacional de educación a distancia, Edit. Grafo S.A., Madrid 2012.