Inferencia Estadística (IE)

Camilo José Torres-Jiménez

2023-11-27

Tabla de contenidos

Acerca de

EN CONSTRUCCION.

El presente es parte del material que se utilizará para las clases de la asignatura **Inferencia Estadística (IE)**. Dicha asignatura está dirigida a estudiantes de la Carrera de Estadística de la Facultad de Ciencias de la Universidad Nacional de Colombia, Sede Bogotá.

Histórico de este material

Este material se ha venido preparando, construyendo y transformando a través del tiempo. A continuación se mencionan los diferentes elementos que se han utilizado para producirlo.

Inicio	Fin	Entrada	Salida
2023-II		*Quarto book*	`html`

Acerca del autor (profesor de la asignatura)

Formación:

- **Matemático**. Línea de profundización: *Informática*. Área del trabajo de grado: *Computación en paralelo*.
- **Magister en Ciencias Estadística**, Área general y específica de la tesis: *Procesos estocásticos*; *procesos de ramificación*.

Experiencia laborar (no docente):

- Seis (6) años como **programador**, **administrador de bases de datos (DBA)**, **analista de datos** y **líder de proyecto** en el marco de convenios entre el *Observatorio Colombiano de Ciencia y Tecnología* y *Colciencias* (*Minciencias*).
- Dos (2) años como **gerente** y **representante legal** de una corporación de ciencia y tecnología sin animo de lucro dedicada a proyectos de desarrollo de software, de soporte operativo a conceptual y de análisis e investigación en torno a los datos del Sistema Nacional de Ciencia, Tecnología e Innovación Colombiano.
- Dos (2) años como **Profesional I** en la *Rectoría* de la *Pontificia Universidad Javeriana*, encargado de apoyar el desarrollo y seguimiento de la planeación y el

desarrollo de proyectos relacionados con las estadísticas institucionales.
- **Consultor y analista estadístico** para diferentes instituciones, en temas de educación superior.

Experiencia docente:

- Un semestre como **docente de cátedra** en las *universidades Central y Javeriana*. Cursos a cargo: Álgebra Lineal, Estadística II y Probabilidad y Estadística; dirigidos a: *Ingeniería y Ciencias Económicas*.
- Un (1) año como **docente de planta** en la *Universidad Santo Tomas*. Cursos a cargo: Estadística I, Estadística II e Inferencia Estadística; dirigidos a: *Ciencias Económicas y Estadística*.
- Seis (6) años como **auxiliar docente** en la *Universidad Nacional de Colombia, Sede Bogotá*. Cursos a cargo: Probabilidad y Estadística Fundamental, Estadística Descriptiva Multivariada, Programación en Lenguajes Estadísticos, Estadística I, Probabilidad Fundamental, Bioestadística Fundamental; dirigidos a: *Ciencias, Ciencias Económicas e Ingeniería*.
- Dos (2) años como **docente de planta** en la *Universidad Nacional de Colombia, Sede Bogotá*. Cursos a cargo: Probabilidad y Estadística Fundamental, Programación en Lenguajes Estadísticos, Estadística Social Fundamental; dirigidos a: *Estadística, Ciencias Económicas, Ingeniería y Ciencias Humanas, Políticas y Sociales*.

EN CONSTRUCCION.

Parte I

Probabilidad (repaso)

Variables aleatorias

EN CONSTRUCCION.

En esta sección se hará una revisión de algunos temas relacionados con variables aleatorias como lo son: la medida de probabilidad asociada a una variable aleatoria, las funciones de distribución acumulativas, las funciones de masa de probabilidad y de densidad, el valor esperado, la varianza, la desigualdad de Chebyshev y los cuantiles de una variable aleatoria.

Recordemos, ¿qué objeto matemático es una medida de probabilidad $P[\cdot]$?

Tip

$$
\begin{aligned}
P : \wp(\Omega) &\longrightarrow [0,1] \\
A &\longrightarrow P[A] = p
\end{aligned}
$$

$$A \subseteq \Omega \qquad 0 \leq p \leq 1$$

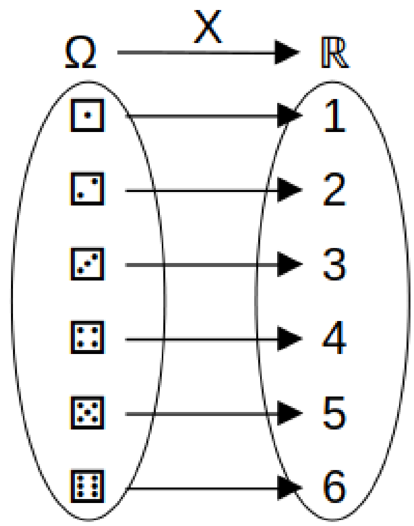
Definición

Sea X una variable asociada al **"Número de puntos en la cara superior de un dado común de seis caras luego de lanzarlo una vez"** o equivalentemente al **"Resultado al lanzar, una vez, un dado común y corriente de seis caras"**.

Evidentemente, los valores que puede llegar a tomar X son $x = 1, 2, 3, 4, 5, 6$. Además, se puede ver que la "variable" X resulta ser el objeto matemático representado por el siguiente gráfico:

Definición 0.1 (Variable aleatoria)**.** Una **variable aleatoria** es una función que asigna un número real a cada uno de los elementos del espacio muestral.

$$
\begin{aligned}
X : \Omega &\longrightarrow \mathbb{R} \\
\omega &\longrightarrow X(w) = x \in \mathbb{R}
\end{aligned}
$$



⚠ En otras palabras, una **variable aleatoria** es una representación o asociación numérica (en los números reales) para los resultados de un experimento aleatorio.

Ejercicio 0.1. Asocie una variable aleatoria a los siguientes experimentos aleatorios:

- Sacar una carta de un mazo (baraja).
- Resultado de un partido ("ganar", "perder", "empatar").

Medida de probabilidad asociada

Es lógico querer una manera de obtener probabilidades para un conjunto B de potenciales valores x en los reales, que podría tomar la variable aleatoria X. Es decir quisiera una medida de probabilidad P_X tal que:

$$\begin{aligned} P_X : \mathcal{B}(\mathbb{R}) &\longrightarrow [0,1] \\ B &\longrightarrow P_X[B] = p \in [0,1] \end{aligned}$$

⚠ ¿Cómo podríamos usar lo que ya hemos visto para lograr lo que queremos o necesitamos?

Ya tenemos una X y una P tal que,

$$
\begin{tikzcd}
\Omega \arrow[r, "X"] \arrow[d, "P"'] & \mathbb{R} \\
{[0, 1]} &
\end{tikzcd}
$$

Además, recordemos que,

- Imagen: Sea $A \subset \Omega$, la imagen de A por X es $X(A) = \{x \in \mathbb{R} : X(\omega) = x, \text{ para algún } \omega \in A\}$
- Preimagen: Sea $B \subset \mathbb{R}$, la preimagen de B por X es $X^{-1}(B) = \{\omega \in \Omega : x = X(\omega), \text{ para algún } x \in B\}$

Entonces, se puede definir la probabilidad de que la variable aleatoria X tome un subconjunto de valores en los reales B de la siguiente manera:

$$
\begin{aligned}
P_X[B] &= P_X[\{x : x \in B\}] \\
&= P[\{\omega \in \Omega : X(w) = x, \text{ para algún } x \in B\}] \\
&= P[X^{-1}(B)]
\end{aligned}
$$

> ⚠ Es decir, obtenemos P_X haciendo la composición de X^{-1} y P.
> Además, P_X cumple todos los axiomas, y por ende todas las propiedades, de una medida de probabilidad. Lo único que cambia con respecto a lo anteriormente visto acerca de probabilidades es que siempre estaremos trabajando con conjuntos en los números reales, en vez de estar trabajando con conjuntos de espacios muestrales distintos que dependen del experimento aleatorio.

Abusando un poco de la notación, siempre utilizaremos $P[X \in B]$ para referirnos a $P_X[B]$.

Lo cual implica que,

- Si $B = \{a\}$ entonces

$$\begin{aligned} P_X[\{a\}] &= P\Big[X^{-1}(\{a\})\Big] \\ &= P[X \in \{a\}] \\ &= P[X = a] \end{aligned}$$

- Si $B = (a, b]$ entonces

$$\begin{aligned} P_X[(a,b]] &= P\Big[X^{-1}((a,b])\Big] \\ &= P[X \in (a,b]] \\ &= P[a < X \leq b] \end{aligned}$$

En los siguientes ejercicios, reescriba la probabilidad solicitada en términos de la variable aleatoria usando la notación indicada, y si es posible obtenga el valor para dicha probabilidad.

Ejercicio 0.2. Sea $X :=$ "Número que aparece en la cara superior de un dado después de ser lanzado una vez".

$$P[\text{"Sacar más de 4"}] = ¿?$$

Ejercicio 0.3. Sea $Y :=$ "El resultado de un partido codificado de la siguiente manera: Ganar$=1$, Empatar$=0$, Perder$=-1$".

$$P[\text{"No perder"}] = ¿?$$

Ejercicio 0.4. Sea $Z :=$ "El número de caras que aparecen en el lanzamiento de una moneda normal tres veces".

$$P[\text{"Sacar al menos dos sellos"}] = ¿?$$

Función de distribución acumulativa

La función de distribución acumulativa F_X de una variable aleatoria X se define como:

$$\begin{aligned} F_X(x) &:= P[X \in (-\infty, x]] \\ &= P[X \leq x] \end{aligned}$$

donde $x \in \mathbb{R}$.

Para que una función (F_X) sea una función de distribución acumulativa, de una variable aleatoria X, debe cumplir lo siguiente:

1. F_X es no decreciente (Si $x_1 \leq x_2$ entonces $F_X(x_1) \leq F_X(x_2)$).
2. F_X es continua a derecha $\left(F_X(x^+) = \lim_{h \to 0} F_X(x+h) = F_X(x)\right)$.
3. F_X tiende a cero cuando x tiende a menos infinito $\left(\lim_{x \to -\infty} F_X(x) = 0\right)$.
4. F_X tiende a uno cuando x tiende a infinito $\left(\lim_{x \to \infty} F_X(x) = 1\right)$.

Ejercicio 0.5. Determine y haga el gráfico de la función de distribución acumulativa asociada a la siguiente variable aleatoria:

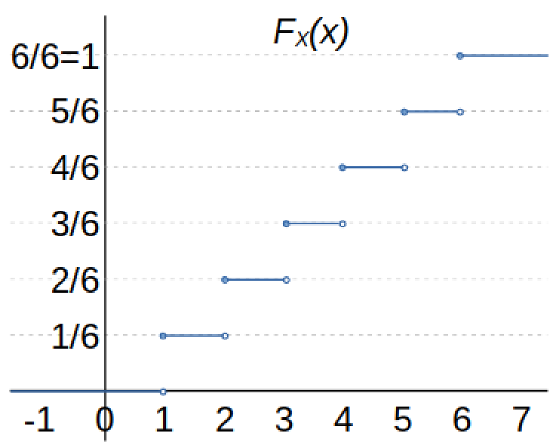
$X :=$ "Resultado al lanzar, una vez, un dado común y corriente".

Solución

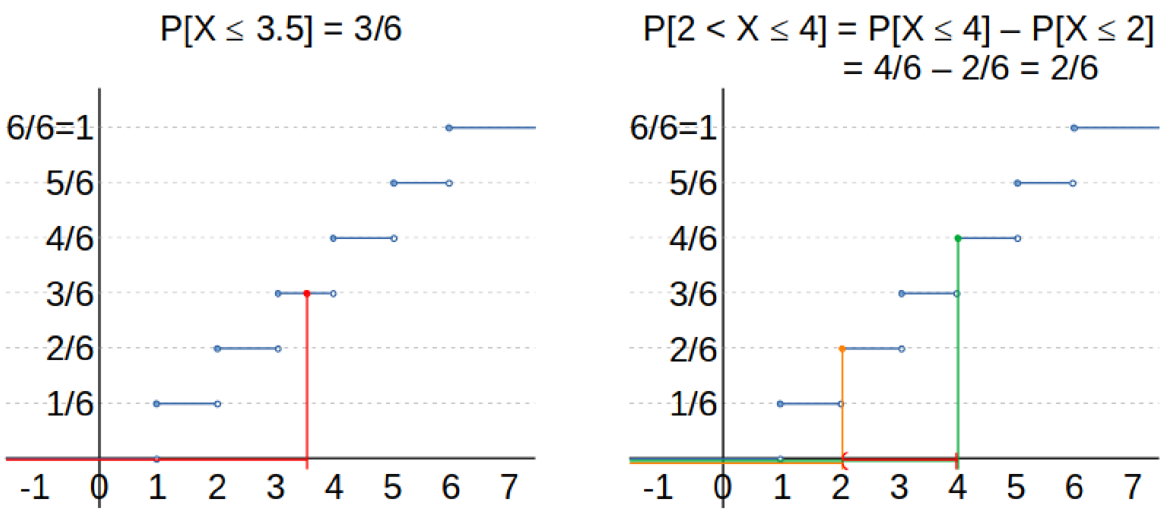
Valores que toma la variable aleatoria X:
$x = 1, 2, 3, 4, 5, 6.$
La función de distribución acumulativa asociada a la variable aleatoria X es,

$$F_X(x) = \begin{cases} \frac{0}{6} = 0 & \text{si } x < 1 \\ \frac{1}{6} & \text{si } 1 \leq x < 2 \\ \frac{2}{6} & \text{si } 2 \leq x < 3 \\ \frac{3}{6} & \text{si } 3 \leq x < 4 \\ \frac{4}{6} & \text{si } 4 \leq x < 5 \\ \frac{5}{6} & \text{si } 5 \leq x < 6 \\ \frac{6}{6} = 1 & \text{si } 6 \leq x \end{cases}$$

y su gráfico sería,



⚠ ¿Cómo hallar cualquier probabilidad deseada, a partir de la función de distribución acumulativa?



Si se conoce la función de distribución acumulativa de una variable aleatoria, entonces,

$$
\begin{aligned}
P[X \leq b] &= F_X(b) \\
\\
P[X < b] &= \lim_{h \to 0} F_X(b-h) \\
&= F_X(b^-) \\
\\
P[a < X \leq b] &= P[X \leq b] - P[X \leq a] \\
&= F_X(b) - F_X(a) \\
\\
P[a \leq X < b] &= P[X < b] - P[X < a] \\
&= F_X(b^-) - F_X(a^-) \\
\\
P[a < X < b] &= P[X < b] - P[X \leq a] \\
&= F_X(b^-) - F_X(a) \\
\\
P[a \leq X \leq b] &= P[X \leq b] - P[X < a] \\
&= F_X(b) - F_X(a^-) \\
\\
P[X = a] &= P[a \leq X \leq a] \\
&= P[X \leq a] - P[X < a] \\
&= F_X(a) - F_X(a^-)
\end{aligned}
$$

Las variables aleatorias se clasifican en:

- Discretas: Cuando F_X es una función escalonada. ($X(\Omega)$ es un conjunto finito o numerable)
- Continuas: Cuando F_X es una función continua. ($X(\Omega)$ es un intervalo o unión de intervalos)
- Mixtas: Cuando F_X es una combinación de al menos una función de distribución acumulativa escalonada y una función de distribución acumulativa continua.

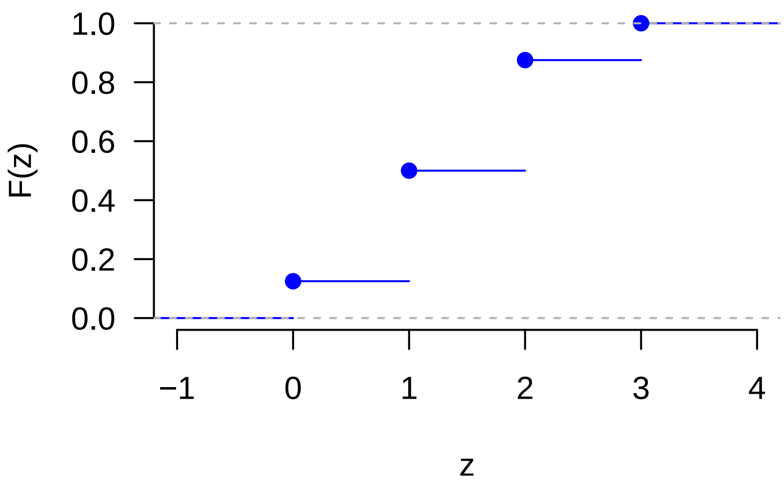
Ejercicio 0.6. Determine y haga el gráfico de la función de distribución acumulativa asociada a la siguiente variable aleatoria:

$Z :=$ “El número de caras que aparecen en el lanzamiento de una moneda normal tres veces”.

Solución

Valores que toma la variable aleatoria Z:
$z = 0, 1, 2, 3.$

$$F_Z(z) = \begin{cases} \frac{0}{8} = 0 & \text{si } z < 0 \\ \frac{1}{8} & \text{si } 0 \le z < 1 \\ \frac{4}{8} & \text{si } 1 \le z < 2 \\ \frac{7}{8} & \text{si } 2 \le z < 3 \\ \frac{8}{8} = 1 & \text{si } 3 \le z \end{cases}$$



Función de masa de probabilidad y función de densidad

Función de masa de probabilidad

La **función de masa de probabilidad** f_X de una variable aleatoria **discreta** X se define como:

$$\begin{aligned} f_X(x) &:= F_X(x) - F_X(x^-) \\ &= P[X = x] \end{aligned}$$

donde $x \in \mathbb{R}$.

Para que una función (f_X) sea una función de masa de probabilidad, de una variable aleatoria discreta X, debe cumplir lo siguiente:

1. $f_X(x) \geq 0$, para todo $x \in \mathbb{R}$.

2. $\sum_{t\in\mathbb{R}} f_X(t) = 1.$

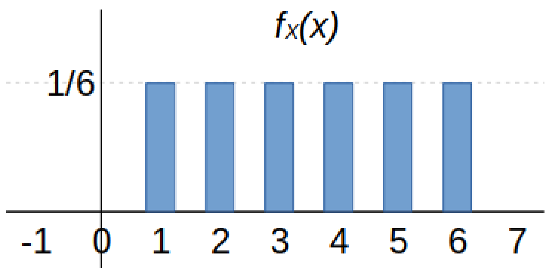
Ejercicio 0.7. Determine y haga el gráfico de la función de masa de probabilidad asociada a la siguiente variable aleatoria:

$X :=$ "Resultado al lanzar, una vez, un dado común y corriente".

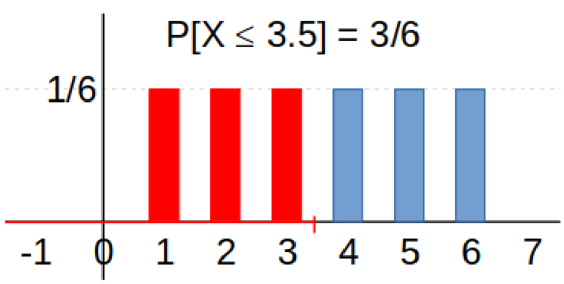
Solución

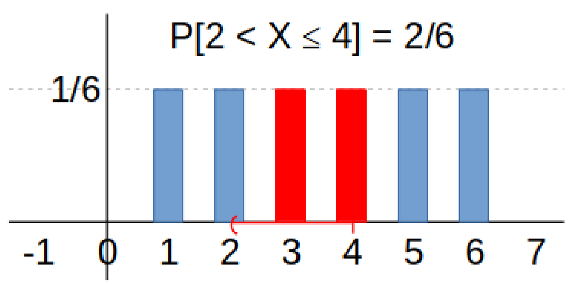
Valores que toma la variable aleatoria X:
$x = 1, 2, 3, 4, 5, 6.$

x	1	2	3	4	5	6
$f_X(x) = P[X = x]$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$



¿Cómo hallar cualquier probabilidad deseada, a partir de la función de masa de probabilidad?





Si se conoce la función de masa de probabilidad de una variable aleatoria, entonces,

$$P[X \in B] = \sum_{t \in B} f_X(t)$$

$$P[a < X \leq b] = \sum_{a < t \leq b} f_X(t)$$

$$P[a \leq X < b] = \sum_{a \leq t < b} f_X(t)$$

$$\vdots$$

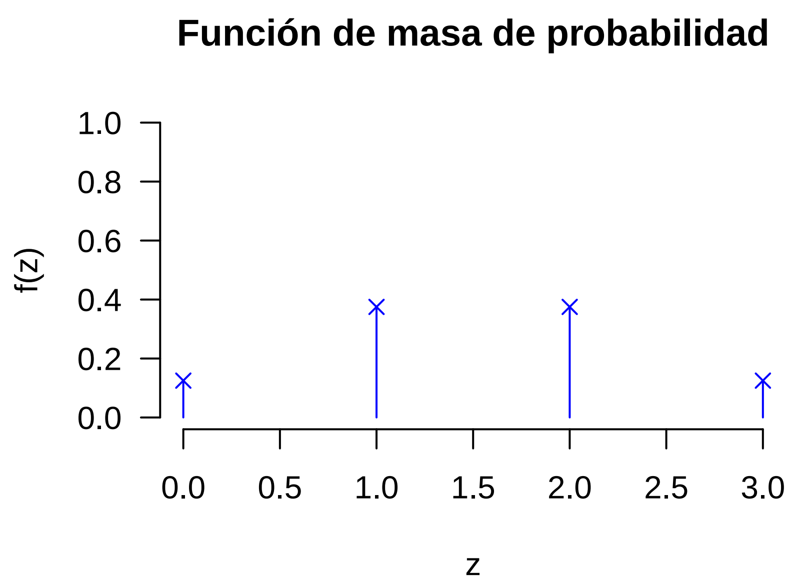
Ejercicio 0.8. Determine y haga el gráfico de la función de masa de probabilidad asociada a la siguiente variable aleatoria:

$Z :=$ "El número de caras que aparecen en el lanzamiento de una moneda normal tres veces".

Solución

Valores que toma la variable aleatoria Z:
$z = 0, 1, 2, 3$.

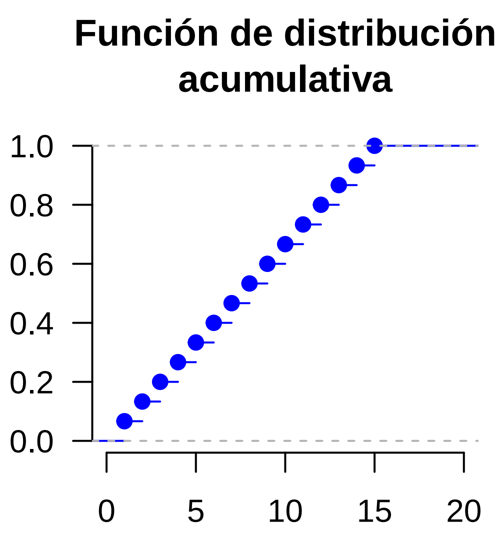
x	0	1	2	3
$f_X(x) = P[X = x]$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

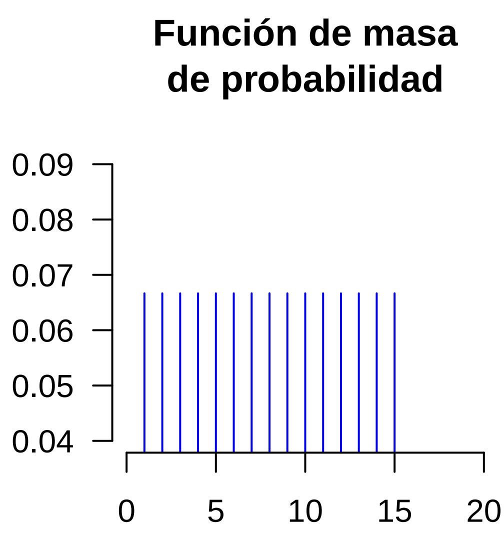


Función de densidad

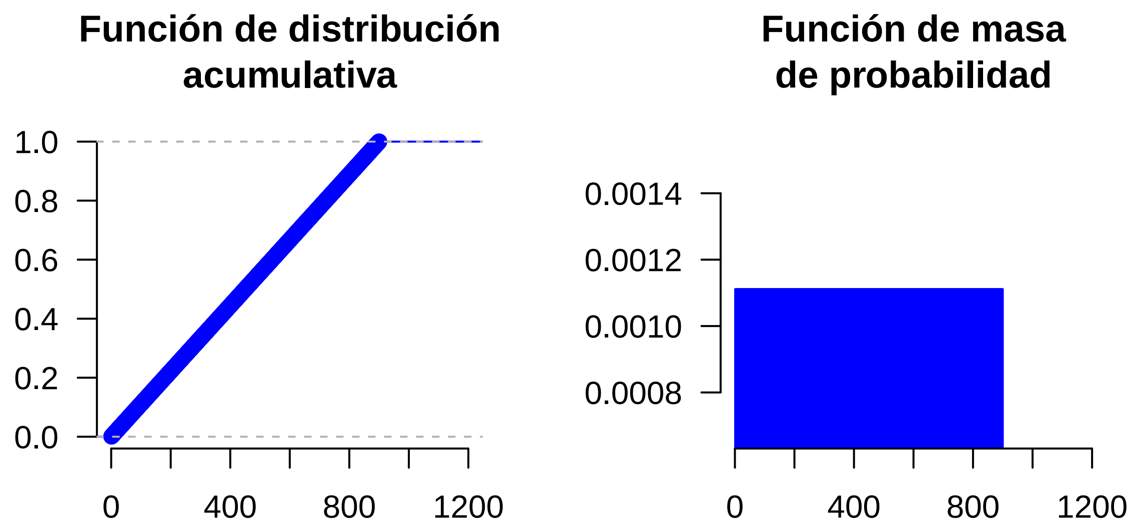
Sea una variable aleatoria X definida como "el tiempo que se debe esperar para…".

Supongamos que una persona **erronea o incorrectamente** dice que dicha variable aleatoria es discreta con valores $x = 1, 2, \dots, 15$ minutos, todos ellos igualmente probables, entonces,

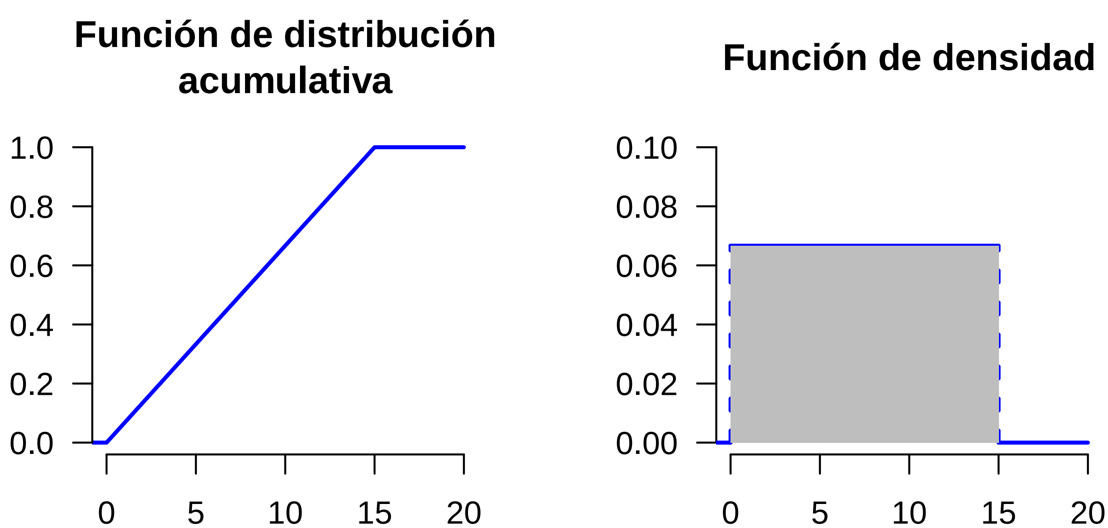




Ahora, supongamos que esa persona empieza a darse cuenta que algo anda mal e intenta corregir su error pasando de minutos a segundos, de donde asume **erronea o incorrectamente** que la variable aleatoria es discreta con valores $x = 1, 2, \ldots, 900$ segundos, todos ellos igualmente probables, entonces,



Finalmente la persona se da cuenta de que el error que está cometiendo consiste en considerar como discreta una variable que **es continua** (desde la parte de estadística descriptiva hemos evidenciado que una variable relacionada con el tiempo es cuantitativa continua con escala de medida de razón). Sin embargo, gracias al error cometido podemos imaginar lo que ocurrirá cuando el tiempo se divida en partes cada vez más pequeñas, hasta llegar a unidades de tiempo infinitesimales. Cuando el tiempo que debo esperar varía infinitesimalmente entre 0 y 15 minutos ($0 < x < 15$) con igual probabilidad, entonces,



⚠ Si para una variable aleatoria discreta se tiene que $f_X(x) := F_X(x) - F_X(x^-)$, entonces ¿cual sería la fórmula equivalente (infinitesimal) para una aleatoria continua?

La **función de densidad** f_X de una variable aleatoria **continua** X se define como:

$$\begin{aligned} f_X(x) &:= \lim_{h \to 0} \frac{F_X(x+h) - F_X(x)}{(x+h) - x} \\ &= \frac{d}{dx} F_X(x) \\ &= F'_X(x) \end{aligned}$$

Una función f_X es función de densidad de una variable aleatoria continua X si:

1. $f_X(x) \geq 0$, para todo $x \in \mathbb{R}$.
2. $\int\limits_{t \in \mathbb{R}} f_X(t)\, dt = \int\limits_{-\infty}^{\infty} f_X(t)\, dt = 1$

Note que,

$$\begin{aligned} F_X(x) &= P[X \leq x] \\ &= \int\limits_{-\infty}^{x} f_X(t)\, dt \end{aligned}$$

Es decir que $P[X \leq x]$ es igual al área acumulada bajo la curva $y = f_X(t)$ desde $-\infty$ hasta el valor dado para x.

Ejercicio 0.9.

La vida útil, en días, para frascos de cierta medicina de prescripción es una variable aleatoria que tiene la siguiente función de densidad:

$$f_X(x) = \begin{cases} \frac{20000}{(x+100)^3} & x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Calcule la probabilidad de que un frasco de esta medicina tenga una vida útil de: (a) al menos 200 días; (b) cualquier lapso entre 80 y 120 días.

Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación. Ejercicio 3.6.

Si se conoce la función de densidad de una variable aleatoria, entonces,

$$P[X \in B] = \int\limits_{t \in B} f_X(t)dt$$

$$\begin{aligned} P[a \leq X \leq b] &= \int\limits_{a \leq t \leq b} f_X(t)dt \\ &= \int\limits_a^b f_X(t)dt \end{aligned}$$

Es decir que $P[a \leq X \leq b]$ es igual al área bajo la curva dada por $f_X(t)$, entre las rectas $x = a$ y $x = b$.

> ⚠ Si X una variable aleatoria continua, entonces $P[X = a] =$ ¿?.

A partir de todo lo visto en las anteriores secciones, concluimos que basta con conocer la función de distribución acumulativa o la función de densidad (la función de masa de probabilidad para el caso discreto) para conocer por completo el mecanismo aleatorio que rige a la variable aleatoria y al experimento aleatorio respectivo (y por ende, se puede calcular cualquier probabilidad que se quiera).

Cuantiles

El **cuantil** x_p, para $p \in [0, 1]$, de una variable aleatoria X es un valor que toma la variable tal que:

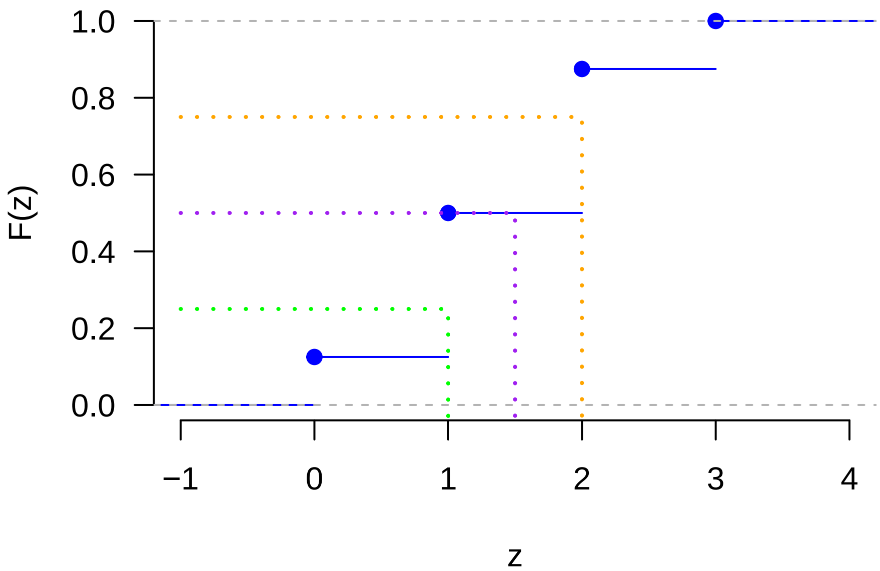
$$P[X \leq x_p] \geq p \qquad \text{y} \qquad P[X \geq x_p] \geq 1 - p$$

Un cuantil x_q no es necesariamente único. Cuando un cuantil no es único, existe un intervalo en el que cada punto del intervalo satisface las dos condiciones para ser cuantil.

Ejercicio 0.10. Encuentre los **cuartiles** 1, 2, y 3 (**cuantiles** 0.25, 0.5 y 0.75) del número de caras que aparecen en el lanzamiento tres veces consecutivas de una moneda común y corriente.

> **Tip**
>
> $Z :=$ “El número de caras que aparecen en el lanzamiento tres veces consecutivas de una moneda común y corriente”.

Valores que toma la variable: $z = 0, 1, 2, 3$.



Note que para una variable aleatoria continua X, se tiene que,

$$
\begin{aligned}
P[X \leq x_p] &= p \\
F_X[x_p] &= p \\
x_p &= F_X^{-1}(p)
\end{aligned}
$$

Una función inversa de la función de distribución acumulativa de la variable aleatoria X se denomina **función cuantil** de X.

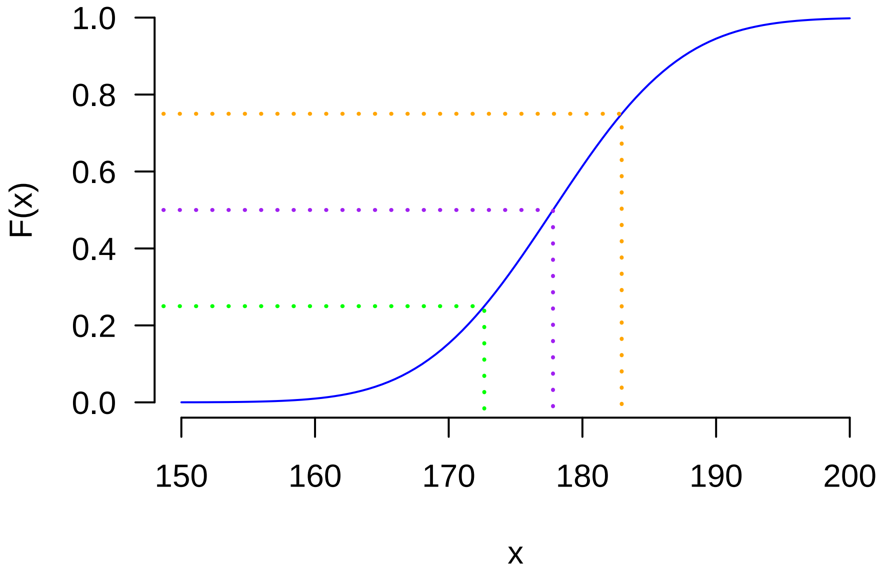
Ejercicios

A continuación se encuentran una serie de ejercicios tomados del siguiente libro:

> Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación.

Ejercicio 0.11 (Walpole 3.1)**.**

> Clasifique las siguientes variables aleatorias como discretas o continuas:
>
> X: el número de accidentes automovilísticos que ocurren al año en Virginia.
>
> Y: el tiempo para jugar 18 hoyos de golf.



M: la cantidad de leche que una vaca específica produce anualmente.

N: el número de huevos que una gallina pone mensualmente.

P: el número de permisos para construcción que los funcionarios de una ciudad emiten cada mes.

Q: el peso del grano producido por acre.

Ejercicio 0.12 (Walpole 3.3)**.**

Sea W la variable aleatoria que da el número de caras menos el número de cruces en tres lanzamientos de una moneda. Liste los elementos del espacio muestral S para los tres lanzamientos de la moneda y asigne un valor w de W a cada punto muestral.

Ejercicio 0.13 (Walpole 3.5)**.**

Determine el valor c de modo que cada una de las siguientes funciones sirva como distribución de probabilidad de la variable aleatoria discreta X:

a. $c\,(x^2+4)$, para $x = 0, 1, 2, 3$;

b. $c\binom{2}{x}\binom{3}{3-x}$, para $x = 0, 1, 2$.

Ejercicio 0.14 (Walpole 3.7)**.**

El número total de horas, medidas en unidades de 100 horas, que una familia utiliza una aspiradora en un periodo de un año es una variable aleatoria continua X que tiene la siguiente función de densidad:

$$f(x) = \begin{cases} x & 0 < x < 1 \\ 2 - x & 1 \leq x < 2 \\ 0 & \text{en otro caso} \end{cases}$$

Calcule la probabilidad de que en un periodo de un año una familia utilice su aspiradora,

a. menos de 120 horas;

b. entre 50 y 100 horas.

Ejercicio 0.15 (Walpole 3.13)**.**

La distribución de probabilidad de X, el número de imperfecciones que se encuentran en cada 10 metros de una tela sintética que viene en rollos continuos de ancho uniforme, está dada por

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

Construya la función de distribución acumulativa de X.

Valor esperado o esperanza

El valor esperado / la esperanza / el promedio para una variable aleatoria discreta X es:

$$\begin{aligned} \mu_X &= E[X] \\ &= \sum_{t \in \mathbb{R}} t\, f_X(t) \\ &= \sum_{t \in \mathbb{R}} t\, P[X = t] \end{aligned}$$

El valor esperado / la esperanza / el promedio para una variable aleatoria continua X es:

$$\begin{aligned} \mu_X &= E[X] \\ &= \int_{t \in \mathbb{R}} t\, f_X(t)\, dt \\ &= \int_{-\infty}^{\infty} t\, f_X(t)\, dt \end{aligned}$$

Propiedades del valor esperado:

- Si $X = a$ constante, entonces $E[X] = E[a] = a$.
- $E[X + b] = E[X] + b$.
- $E[a\, X] = a\, E[X]$.

Ejercicio 0.16. Encuentre el valor esperado / la esperanza / el promedio del número de caras que aparecen al lanzar tres veces consecutivas una moneda común y corriente.

Solución

$Z :=$ "El número de caras que aparecen al lanzar tres veces consecutivas una moneda común y corriente".
Valores que toma la variable: $z = 0, 1, 2, 3$.

$$
\begin{aligned}
E[Z] &= \sum_{t\in\mathbb{R}} t\, f_Z(t) = \sum_{t\in\mathbb{R}} t\, P[Z = t] \\
&= \sum_{z=0}^{3} z\, f_Z(z) = \sum_{z=0}^{3} z\, P[Z = z] \\
&= (0)\, P[Z = 0] + (1)\, P[Z = 1] + (2)\, P[Z = 2] + (3)\, P[Z = 3] \\
&= (0)\, \frac{1}{8} + (1)\, \frac{3}{8} + (2)\, \frac{3}{8} + (3)\, \frac{1}{8} \\
&= \frac{3}{8} + \frac{6}{8} + \frac{3}{8} \\
&= \frac{12}{8} \\
&= 1.5
\end{aligned}
$$

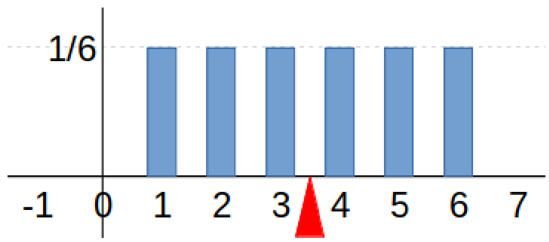
Ejercicio 0.17. Encuentre el valor esperado / la esperanza / el promedio del número de puntos que aparecen en la cara superior de un dado común y corriente de seis caras que se lanza una vez.

Solución

$X :=$ "Resultado al lanzar, una vez, un dado común y corriente".
Valores que toma la variable: $x = 1, 2, 3, 4, 5, 6$.

x	$f_X(x)$	$x\, f_X(x)$
1	$\frac{1}{6}$	$\frac{1}{6}$

2	$\frac{1}{6}$	$\frac{2}{6}$
3	$\frac{1}{6}$	$\frac{3}{6}$
4	$\frac{1}{6}$	$\frac{4}{6}$
5	$\frac{1}{6}$	$\frac{5}{6}$
6	$\frac{1}{6}$	$\frac{6}{6}$
	Suma:	$\sum_{x=0}^{3} x\, f_X(x) = \frac{21}{6} = 3.5$



Ejercicio 0.18. Suponga que en un juego de azar inicialmente se debe apostar \$10000 pesos, luego se lanza un dado común y corriente de seis caras y por cada punto que salga en la cara superior del dado, el apostador recibe \$2500. ¿A la larga, cuál será la ganancia-perdida promedio de el(los) apostador(es) por cada vez que se apueste?

Solución

Si X := "Resultado al lanzar, una vez, un dado común y corriente", entonces $Y = \$2500X - \$10000$ sería la variable aleatoria asociada a la ganancia-perdida de el(los) apostador(es).

x	$f_X(x)$	y	$f_Y(y)$	$y\, f_Y(y)$
1	$\frac{1}{6}$	-7500	$\frac{1}{6}$	$\frac{-7500}{6}$
2	$\frac{1}{6}$	-5000	$\frac{1}{6}$	$\frac{-5000}{6}$
3	$\frac{1}{6}$	-2500	$\frac{1}{6}$	$\frac{-2500}{6}$
4	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{0}{6}$
5	$\frac{1}{6}$	2500	$\frac{1}{6}$	$\frac{2500}{6}$
6	$\frac{1}{6}$	5000	$\frac{1}{6}$	$\frac{5000}{6}$
			Suma:	$\sum_{i=1}^{6} y_i\, f_Y(y_i) = \frac{-7500}{6} = -1250$

Si hubiese tenido en cuenta las propiedades del valor esperado:

$$\begin{aligned} E[Y] &= E[2500X - 10000] \\ &= 2500E[X] - 10000 \\ &= 2500(3.5) - 10000 \\ &= -1250 \end{aligned}$$

Es decir que, a la larga y en promedio, el(los) apostador(es) perderán $1250 pesos por cada vez que se apueste.

Valor esperado de una función

El valor esperado de una función $g(X)$ de una variable aleatoria discreta X es:

$$\begin{aligned} \mu_{g(X)} &= E[g(X)] \\ &= \sum_{t \in \mathbb{R}} g(t)\, f_X(t) \\ &= \sum_{t \in \mathbb{R}} g(t)\, P[X = t] \end{aligned}$$

El valor esperado de una función $g(X)$ de una variable aleatoria continua X es:

$$\begin{aligned} \mu_{g(X)} &= E[g(X)] \\ &= \int_{t \in \mathbb{R}} g(t)\, f_X(t)\, dt \\ &= \int_{-\infty}^{\infty} g(t)\, f_X(t)\, dt \end{aligned}$$

Ejercicio 0.19. Encuentre la ganancia-perdida promedio de un apostador, si de entrada tiene que apostar $500, y si gana $100 por el número al cuadrado de caras que aparecen en el lanzamiento de una moneda normal tres veces.

Solución

Sea Z := "El número de caras que aparecen en el lanzamiento de una moneda normal tres veces". La función de ganancia-perdida sería $g(Z) = 100\, Z^2 - 500$.

$$
\begin{aligned}
E[100\,Z^2-500] &= \sum_{t\in\mathbb{R}} \left(100\,t^2-500\right)\; P[Z=t] \\
&= \sum_{z=0}^{3} \left(100\,z^2-500\right)\; P[Z=z] \\
&= (-500)\,\frac{1}{8} + (-400)\,\frac{3}{8} + (-100)\,\frac{3}{8} + (400)\,\frac{1}{8} \\
&= (-500)\,\frac{1}{8} + (-1200)\,\frac{1}{8} + (-300)\,\frac{1}{8} + (400)\,\frac{1}{8} \\
&= (-1600)\,\frac{1}{8} \\
&= -200
\end{aligned}
$$

Ejercicio 0.20. Si inicialmente se debe apostar \$7500 pesos, y se gana el resultado de un dado común elevado al cubo y multiplicado por \$100. ¿A la larga, cuál será la ganancia-perdida promedio de un apostador por cada vez que apueste?

Solución

Sea X := "Resultado al lanzar, una vez, un dado común y corriente". $Y = g(X) = 100X^3 - 7500$ sería la variable aleatoria asociada a la ganancia-perdida.
Además,

$$
\begin{aligned}
E[Y] &= E\left[100X^3 - 7500\right] \\
&= 100E\left[X^3\right] - 7500.
\end{aligned}
$$

x	$f_X(x)$	$g(x) = x^3$	$g(x)\,f_X(x)$
1	$\frac{1}{6}$	1	$\frac{1}{6}$
2	$\frac{1}{6}$	8	$\frac{8}{6}$
3	$\frac{1}{6}$	27	$\frac{27}{6}$
4	$\frac{1}{6}$	64	$\frac{64}{6}$
5	$\frac{1}{6}$	125	$\frac{125}{6}$
6	$\frac{1}{6}$	216	$\frac{216}{6}$
		Suma:	$\sum_{x=1}^{6} g(x)\,f_X(x) = \frac{441}{6} = 73.5$

$$
\begin{aligned}
E[Y] &= 100E\left[X^3\right] - 7500 \\
&= 100(73.5) - 7500 \\
&= -150
\end{aligned}
$$

Es decir que por cada vez que se apueste, a la larga en promedio se perderán \$150.

Varianza

La varianza de una variable aleatoria X es:

$$
\begin{aligned}
\sigma_X^2 &= Var[X] \\
&= E\left[\left(X - E[X]\right)^2\right] \\
&= E[X^2] - \left(E[X]\right)^2
\end{aligned}
$$

Propiedades de la varianza:

- $Var[X] \geq 0$, para toda variable aleatoria X.
- Si $X = a$ constante, entonces $Var[X] = Var[a] = 0$.
- $Var[X + b] = Var[X]$.
- $Var[a\,X] = a^2\,Var[X]$.

Ejercicio 0.21. Encuentre la varianza del número de caras que aparecen al lanzar tres veces consecutivas una moneda común y corriente.

Solución

Z := "El número de caras que aparecen al lanzar tres veces consecutivas una moneda común y corriente".
Valores que toma la variable: $z = 0, 1, 2, 3$.

$$
\begin{aligned}
Var[Z] &= E\left[Z^2\right] - \left(E[Z]\right)^2 \\
&= \left(\sum_{z=0}^{3} z^2\, P[Z=z]\right) - (1.5)^2 \\
&= (0)\,\frac{1}{8} + (1)\,\frac{3}{8} + (4)\,\frac{3}{8} + (9)\,\frac{1}{8} - (1.5)^2 \\
&= \frac{3}{8} + \frac{12}{8} + \frac{9}{8} - (1.5)^2 \\
&= \frac{24}{8} - (1.5)^2 \\
&= 3 - (1.5)^2 \\
&= 0.75
\end{aligned}
$$

Ejercicio 0.22. Encuentre la varianza y la desviación estándar del número de puntos que aparecen en la cara superior de un dado común y corriente de seis caras que se lanza una vez.

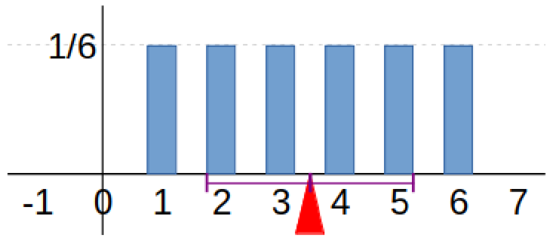
Solución

$X :=$ "Resultado al lanzar, una vez, un dado común y corriente".
Valores que toma la variable: $x = 1, 2, 3, 4, 5, 6$.

x	$f_X(x)$	$x^2\, f_X(x)$
1	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{6}$	$\frac{4}{6}$
3	$\frac{1}{6}$	$\frac{9}{6}$
4	$\frac{1}{6}$	$\frac{16}{6}$
5	$\frac{1}{6}$	$\frac{25}{6}$
6	$\frac{1}{6}$	$\frac{36}{6}$
	Suma:	$\sum_{x=1}^{6} x^2\, f_X(x) = \frac{91}{6} = 15.1666\bar{6}$

$$Var[X] = 15.1666\bar{6} - (3.5)^2 = 2.9166\bar{6}$$

$$SD[X] = \sqrt{2.9166\bar{6}} = 1.707825$$



Desigualdad de Chebyshev

Para toda variable aleatoria X (denotando $\mu_X = E[X]$ y $\sigma_X^2 = Var[X]$), la probabilidad de que ésta tome valores a k desviaciones estándar de la media es por lo menos (es como mínimo) $1 - \frac{1}{k^2}$. Es decir,

$$\begin{aligned} P[\mu_X - (k)\,\sigma_X < X < \mu_X + (k)\,\sigma_X] &\geq 1 - \frac{1}{k^2} \\ P[-(k)\,\sigma_X < X - \mu_X < (k)\,\sigma_X] &\geq 1 - \frac{1}{k^2} \\ P[|X - \mu_X| < (k)\,\sigma_X] &\geq 1 - \frac{1}{k^2} \end{aligned}$$

para k dado.

Por ejemplo, para $k = 1$ se concluye que,

$$\begin{aligned} P[|X - \mu_X| < (1)\,\sigma_X] &\geq 1 - \frac{1}{(1)^2} \\ P[|X - \mu_X| < \sigma_X] &\geq 0 \\ P[\mu_X - \sigma_X < X < \mu_X + \sigma_X] &\geq 0 \end{aligned}$$

para toda variable aleatoria X. Conclusión que es obvia, ya que toda probabilidad es mayor que cero. Sin embargo, para $k > 1$ tendríamos conclusiones más interesantes. Por ejemplo, para $k = 2.5$ se concluye que,

$$\begin{aligned} P[|X - \mu_X| < (2.5)\,\sigma_X] &\geq 1 - \frac{1}{(2.5)^2} \\ P[\mu_X - (2.5)\,\sigma_X < X < \mu_X + (2.5)\,\sigma_X] &\geq 0.84 \end{aligned}$$

para toda variable aleatoria X (sin importar su distribución).

Por otro lado, si en

$$P[\mu_X - (k)\,\sigma_X < X < \mu_X + (k)\,\sigma_X] \geq 1 - \frac{1}{k^2}$$

decimos que $\epsilon = (k)\,\sigma$ entonces la desigualdad quedaría de la siguiente manera,

$$P[\mu_X - \epsilon < X < \mu_X + \epsilon] \geq 1 - \frac{\sigma_X^2}{\epsilon^2}$$

para $\epsilon > 0$ y cualquier variable aleatoria X (denotando $\mu_X = E[X]$ y $\sigma_X^2 = Var[X]$).

> ¿De qué me sirven los anteriores resultados? ¿para qué los puedo utilizar?

Ejemplo 0.1. Supongamos que una fabrica produce 1500 unidades al día, con una desviación estándar de aproximadamente 300 unidades.

Entonces, sin importar la distribución (el mecanismo aleatorio detrás) de la variable aleatoria X : "número de unidades producidas por la fabrica al día", podemos concluir que,

$$P[\mu_X - (k)\,\sigma_X < X < \mu_X + k\,\sigma_X] \geq 1 - \frac{1}{k^2}$$
$$P[1500 - (k)(300) < X < 1500 + (k)(300)] \geq 1 - \frac{1}{k^2}$$

De allí (sin conocer nada más acerca de la variable aleatoria X) puedo asegurar que como mínimo el 84% de los días se producen entre 750 y 2250 unidades al día (llego a esta conclusión tomando $k = 2.5$).

Ejercicio 0.23. ¿A partir de la información el ejemplo anterior, como mínimo, qué **PROPORCIÓN** de los días se producirán entre 1000 y 2000 unidades?

> **Tip**
>
> Tome $\epsilon = 500$.

Ejemplo 0.2. Supongamos que la estatura media de los hombres adultos en cierto país es de aproximadamente 177.8 cm, con una desviación estándar de alrededor de 7.62 cm.

Entonces, sin importar la distribución (el mecanismo aleatorio detrás) de la variable aleatoria X : "estatura media de los hombres adultos", se tiene que,

$$
\begin{aligned}
P[\mu_X - k\,\sigma_X < X < \mu_X + k\,\sigma_X] &\geq 1 - \frac{1}{k^2}\\
P[177.8 - 2(7.62) < X < 177.8 + 2(7.62)] &\geq 1 - \frac{1}{2^2}\\
P[162.56 < X < 193.04] &\geq 0.75
\end{aligned}
$$

Es decir, al menos (como mínimo) el 75% de los hombres adultos de ese cierto país tienen una estatura entre 162.56 y 193.04 cm.

Ejercicio 0.24. ¿A partir de la información el ejemplo anterior, entre qué estaturas estaría al menos (como mínimo) el 90% de las estaturas de los hombres adultos de ese país?

Ejercicios

A continuación se encuentran una serie de ejercicios tomados del siguiente libro:

> Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación.

Ejercicio 0.25 (Walpole 4.1 (3.13))**.**

> En el ejercicio 3.13 de la página 92 se presenta la siguiente distribución de probabilidad de X, el número de imperfecciones que hay en cada 10 metros de una tela sintética, en rollos continuos de ancho uniforme

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

> Calcule el número promedio de imperfecciones que hay en cada 10 metros de esta tela.

Ejercicio 0.26 (Walpole 4.13 (3.7))**.**

> La función de densidad de la variable aleatoria continua X, el número total de horas que una familia utiliza una aspiradora durante un año, en unidades de 100 horas, se da en el ejercicio 3.7 de la página 92 como

$$
f(x) = \begin{cases} x & 0 < x < 1 \\ 2 - x & 1 \leq x < 2 \\ 0 & \text{en otro caso} \end{cases}
$$

Calcule el número promedio de horas por año que las familias utilizan sus aspiradoras.

Ejercicio 0.27 (Walpole 4.21 (4.12)).

¿Cuál es la utilidad promedio por automóvil que obtiene un distribuidor, si la utilidad en cada uno está dada por $g(X) = X^2$, donde X es una variable aleatoria que tiene la función de densidad del ejercicio 4.12?

Ejercicio 4.12: Si la utilidad para un distribuidor de un automóvil nuevo, en unidades de \$5000, se puede ver como una variable aleatoria X que tiene la siguiente función de densidad

$$f(x) = \begin{cases} 2(1-x) & 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Calcule la utilidad promedio por automóvil.

Ejercicio 0.28 (Walpole 4.49 (4.32)).

Considere la situación del ejercicio 4.32 de la página 119. La distribución del número de imperfecciones por cada 10 metros de tela sintética está dada por

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

Calcule la varianza y la desviación estándar del número de imperfecciones.

Ejercicio 0.29 (Walpole 4.58).

El tiempo total que una adolescente utiliza su secadora de pelo durante un año, medido en unidades de 100 horas, es una variable aleatoria continua X que tiene la siguiente función de densidad

$$f(x) = \begin{cases} x & 0 < x < 1 \\ 2-x & 1 \leq x < 2 \\ 0 & \text{en otro caso} \end{cases}$$

Utilice las propiedades de valor esperado para evaluar la media de la variable aleatoria $Y = 60X^2 + 39X$, donde Y es igual al número de kilowatts-hora que gasta al año.

Ejercicio 0.30 (Walpole 4.75).

Una empresa eléctrica fabrica una bombilla de luz de 100 watts que, de acuerdo con las especificaciones escritas en la caja, tiene una vida media de 900 horas con una desviación estándar de 50 horas. A lo sumo, ¿qué porcentaje de las bombillas no duran al menos 700 horas? Suponga que la distribución es simétrica alrededor de la media.

Funciónes generadora de momentos y característica

Definición 0.2 (Función generadora de momentos)**.** Sea X una variable aleatoria tal que $E\left[\mathrm{e}^{tX}\right]$ es finito para todo $t \in (-\alpha, \alpha)$, con α real positivo. Se define la función generadora de momentos de X, denotada por $m_X(\cdot)$ como:

$$m_X(t) = E\left[\mathrm{e}^{tX}\right]$$

para $t \in (-\alpha, \alpha)$.

Teorema 0.1. *Si X es una variable aleatoria cuya función generadora de momentos existe, entonces, $E\left[X^r\right]$ existe para toda $r \in \mathbb{Z}^+$ (el recíproco no es válido).*

Teorema 0.2. *Si X es una variable aleatoria cuya función generadora de momentos $m_X(\cdot)$ existe, entonces, existe $h \in (0, \infty)$ tal que:*

$$m_X(t) = \sum_{k=0}^{\infty} \frac{E\left[X^k\right]}{k!} t^k$$

para todo $t \in (-h, h)$, y por lo tanto,

$$E\left[X^k\right] = \frac{d^r}{dt^r} m_X(t)\bigg|_{t=0}$$

Definición 0.3 (Función característica)**.** Sea X una variable aleatoria. La función característica de X es la función $\varphi_X : \mathbb{R} \longrightarrow \mathbb{C}$ definida por:

$$\varphi_X(t) := E\left[\mathrm{e}^{\mathrm{i}tX}\right]$$

donde $\mathrm{i} = \sqrt{-1}$ y $\mathrm{e}^{\mathrm{i}tX} = \cos(tX) + \mathrm{i}\sin(tX)$.

Teorema 0.3. *Si X es una variable aleatoria discreta o absolutamente continua entonces $\varphi_X(t)$ existe para todo $t \in \mathbb{R}$.*

Teorema 0.4. *Sea X una variable aleatoria. La función característica $\varphi_X(t)$ de X satisface:*

- $\varphi_X(0) = 1$.

- *$\varphi_X(t) \leq 1$ para todo t.*
- *Si $E\left[X^k\right]$ existe, entonces,*

$$\left.\frac{d^r}{dt^r}\varphi_X(t)\right|_{t=0} = \mathrm{i}^k\, E\left[X^k\right].$$

Teorema 0.5. *Si X y Y son variables aleatorias, y,*

$$\varphi_X(t) = \varphi_Y(t)$$

para todo t, entonces, X y Y tienen la misma distribución.

EN CONSTRUCCION.

Ejercicios

Determine una expresión para la función característica de Y en términos de las constantes en los reales, la función característica de X o las funciones características de los X_i según corresponda.

1. Si $Y = a \in \mathbb{R}$, entonces \$ _Y(t) = \$¿?
2. Si $Y = aX + b$, para $a, b \in \mathbb{R}$, entonces \$ _Y(t) = \$¿?
3. X y Y son variables aleatorias independientes si y solo si \$ _{X+Y}(t) = \$¿?
4. Si $X_1, \dots, X_m$ es un conjunto de variables aleatorias independientes, $a_1, \dots, a_m \in \mathbb{R}$ y $Y = a_1X_1 + \cdots + a_mX_m$, entonces \$ _Y(t) = \$¿?
5. Si $X_1, \dots, X_n$ es una muestra aleatoria y $Y = \bar{X}$, entonces \$ _Y(t) = \$¿?

EN CONSTRUCCION.

Distribuciones discretas

EN CONSTRUCCION.

En esta sección se hará una revisión de algunos temas relacionados con modelos/distribuciones de variables aleatorias discretas.

Uniforme discreta

Ejercicio 0.31.

> De un equipo de 10 empleados, y mediante la selección al azar de una etiqueta contenida en una caja que contiene 10 etiquetas numeradas del 1 al 10, se elige a uno para que supervise cierto proyecto. Calcule la fórmula para la distribución de probabilidad de X que represente el número en la etiqueta que se saca. ¿Cuál es la probabilidad de que el número que se extrae sea menor que 4?
>
> Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación. Ejercicio 5.3.

Características:

Todos los valores que puede tomar la variable aleatoria son igualmente probables.

Valores que puede tomar la variable:

$x = 1, 2, \ldots, n$

Parámetro:

$n \in \mathbb{Z}^+$, número o cantidad de valores que puede tomar la variable.

Notación:

$X \sim \mathcal{U}(n)$, esto se lee así: la variable aleatoria X tiene una distribución uniforme discreta de parámetro n.

Función de masa de probabilidad:

$$f_X(x) = \begin{cases} \frac{1}{n} & \text{Si } x = 1, 2, \dots, n \\ 0 & \text{en otro caso} \end{cases}$$

x	1	2	...	n
$f_X(x)$	$\frac{1}{n}$	$\frac{1}{n}$	...	$\frac{1}{n}$

Función de distribución acumulativa:

$$F_X(x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{1}{n} & \text{si } 1 \leq x < 2 \\ \frac{2}{n} & \text{si } 2 \leq x < 3 \\ \vdots & \vdots \\ \frac{n-1}{n} & \text{si } n-1 \leq x < n \\ 1 & \text{si } n \leq x \end{cases}$$

Valor esperado y varianza:

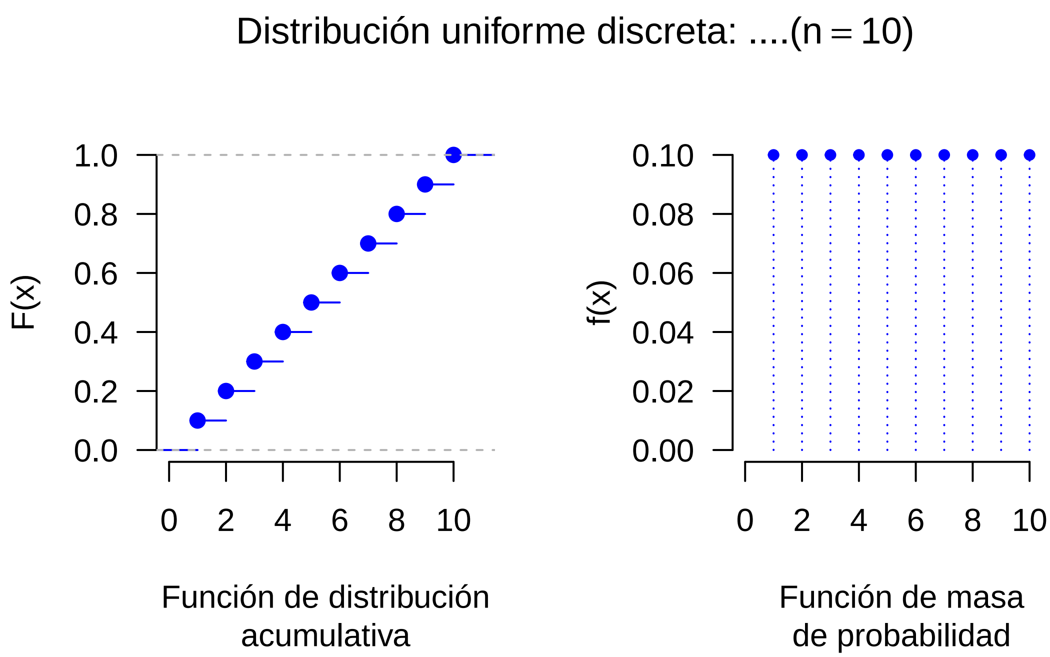
- $\mu_X = E[X] = \frac{n+1}{2}$.
- $\sigma_X^2 = Var[X] = \frac{n^2-1}{12}$.

Función característica:

$$\varphi_X(t) = \frac{e^{\mathrm{i}t} - e^{\mathrm{i}(n+1)t}}{n(1 - e^{\mathrm{i}t})}$$

Para ilustrar:

```
layout(matrix(c(1,2), 1, 2))
n <- 10
x <- 1:n
y <- rep(1/n, n)
plot(ecdf(x), bty="n", col="blue", xlim=c(0,11), las=1, main="",
     sub="Función de distribución\nacumulativa", xlab="", ylab="F(x)")
title(bquote("Distribución uniforme discreta:  ("*n == .(n)*")"),
      line= -1.25, outer=TRUE)
plot(x, y, type= "h", bty="n", col="blue", xlim=c(0,11), ylim=c(0,max(y)), las=1, lty=3,
     main="", sub="Función de masa\nde probabilidad", xlab="", ylab="f(x)")
points(x, y, pch=20, col="blue")
```



- Special Distribution Calculator

Algunas situaciones en donde aplica:

- Las variables en donde todos sus valores son igualmente probables.
- Número resultante en el lanzamiento de un dado de cualquier número de caras que no "esté cargado".

Bernoulli

Ejercicio 0.32.

> La probabilidad de que llueva mañana es 0.40 y la probabilidad de que no llueva es 0.60.
>
> Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación. Ejercicio 2.49.

Características:

La variable aleatoria sólo toma dos posibles valores, uno asociado a "acierto" y otro a "fracaso".

Valores que puede tomar la variable:

$x = 0$ ("fracaso"), 1 ("acierto").

Parámetro:

$p \in [0, 1]$, probabilidad de "acierto".

Notación:

$X \sim \mathcal{B}(p)$, esto se lee así: la variable aleatoria X tiene una distribución Bernoulli de parámetro p.

Función de masa de probabilidad:

$$f_X(x) = \begin{cases} 1-p & \text{Si } x = 0 \\ p & \text{Si } x = 1 \\ 0 & \text{en otro caso} \end{cases}$$

x	0	1
$f_X(x)$	$1-p$	p

Función de distribución acumulativa:

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1-p & \text{si } 0 \le x < 1 \\ 1 & \text{si } 1 \le x \end{cases}$$

Valor esperado y varianza:

- $\mu_X = E[X] = p$.
- $\sigma_X^2 = Var[X] = p(1-p)$.

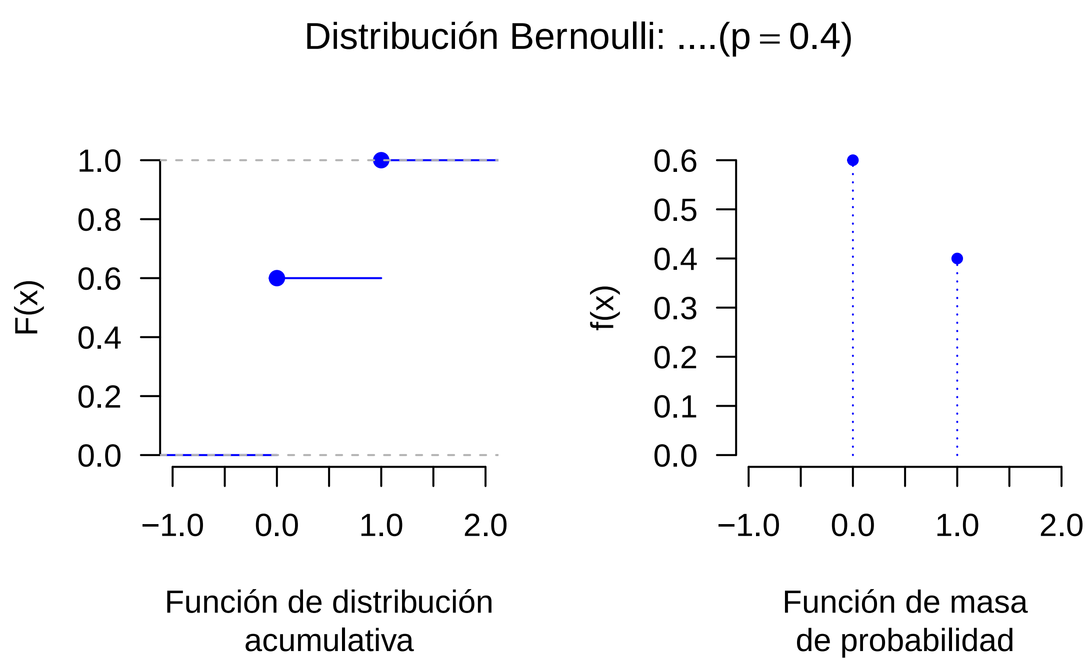
Función característica:

$$\varphi_X(t) = (1-p) + p\,\mathrm{e}^{\mathrm{i}t}$$

Para ilustrar:

```
layout(matrix(c(1,2), 1, 2))
p <- 0.4
x <- c(0, 1)
y <- c(1-p, p)
cdf <- approxfun(x, cumsum(y), method = "constant", yleft = 0, yright = 1, f = 0,
```

```r
                 ties = "ordered")
class(cdf) <- c("ecdf", "stepfun", class(cdf))
attr(cdf, "call") <- sys.call()
plot(cdf, bty="n", col="blue", xlim=c(-1,2), las=1, main="",
     sub="Función de distribución\nacumulativa", xlab="", ylab="F(x)")
title(bquote("Distribución Bernoulli:  ("*p == .(p)*")"),
      line= -1.25, outer=TRUE)
plot(x, y, type= "h", bty="n", col="blue", xlim=c(-1,2), ylim=c(0,max(y)), las=1, lty=3,
     main="", sub="Función de masa\nde probabilidad", xlab="", ylab="f(x)")
points(x, y, pch=20, col="blue")
```



- Binomial Coin Experiment. n = 1

Algunas situaciones en donde aplica:

- Resultado del lanzamiento de una moneda (1: cara y 0: sello).
- Ganar o perder (Resultado al lanzar un dado donde si sale 5 o 6 gano y en otro caso pierdo).
- Seleccionar un individuo o elemento y ver si tiene una característica de interés (Resultado de seleccionar un estudiante al azar de la clase e identificar si ha aprobado todos los exámenes).

⚠ A partir de un experimento Bernoulli se pueden derivar tres distribuciones:

- Binomial
- Hipergeométrica
- Binomial Negativa (incluyendo la Geométrica como un caso particular)

Binomial

Ejercicio 0.33.

> La probabilidad de que un paciente se recupere de una delicada operación de corazón es 0.9. ¿Cuál es la probabilidad de que exactamente 5 de los siguientes 7 pacientes intervenidos sobrevivan?
>
> Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación. Ejercicio 5.11.

Características:

La variable aleatoria corresponde al número de "aciertos" de un experimento Bernoulli que se repite exactamente N veces y donde el resultado de cada repetición no depende de ninguna otra repetición, es decir, los resultados de cada repetición son independientes entre si.

Valores que puede tomar la variable:

$x = 0, 1, \dots, n$

Parámetros:

$n \in \mathbb{Z}^+$, número de repeticiones, y $p \in [0, 1]$, probabilidad de "acierto".

Notación:

$X \sim \mathcal{B}(n, p)$, esto se lee así: la variable aleatoria X tiene una distribución Binomial de parámetros n y p.

Función de masa de probabilidad:

$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{Si } x = 0, 1, \dots, n \\ 0 & \text{en otro caso} \end{cases}$$

Función de distribución acumulativa:

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \binom{n}{0} p^0 \, (1-p)^{n-0} & \text{si } 0 \leq x < 1 \\ \binom{n}{0} p^0 \, (1-p)^{n-0} + \binom{n}{1} p^1 \, (1-p)^{n-1} & \text{si } 1 \leq x < 2 \\ \vdots & \vdots \\ \sum_{i=0}^{n-1} \binom{n}{i} p^i \, (1-p)^{n-i} & \text{si } n-1 \leq x < n \\ 1 & \text{si } n \leq x \end{cases}$$

Valor esperado y varianza:

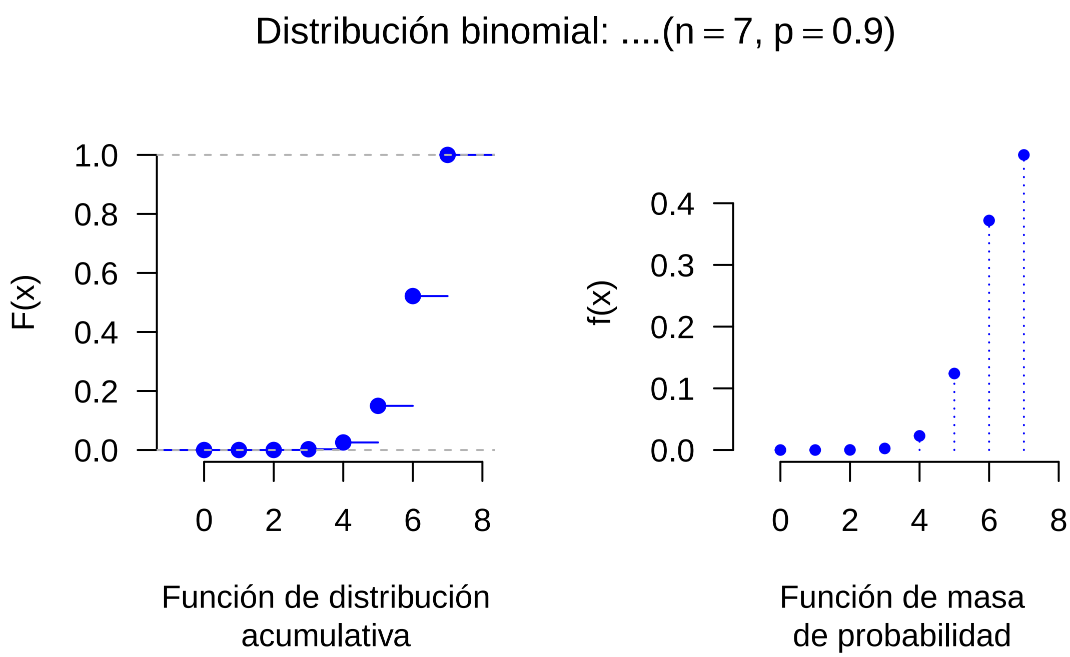
- $\mu_X = E[X] = (n)(p)$.
- $\sigma_X^2 = Var[X] = (n)(p)(1-p)$.

Función característica:

$$\varphi_X(t) = \left((1-p) + p \, \mathrm{e}^{\mathrm{i}t}\right)^n$$

Para ilustrar:

```
layout(matrix(c(1,2), 1, 2))
n <- 7
p <- 0.9
x <- 0:n
y <- dbinom(x, n, p)
cdf <- approxfun(x, cumsum(y), method = "constant", yleft = 0, yright = 1, f = 0,
                 ties = "ordered")
class(cdf) <- c("ecdf", "stepfun", class(cdf))
attr(cdf, "call") <- sys.call()
plot(cdf, bty="n", col="blue", xlim=c(-1,n+1), las=1, main="",
     sub="Función de distribución\nacumulativa", xlab="", ylab="F(x)")
title(bquote("Distribución binomial:  ("*n == .(n)*","~p == .(p)*")"),
      line= -1.25, outer=TRUE)
plot(x, y, type= "h", bty="n", col="blue", xlim=c(-1,n+1), ylim=c(0, max(y)), las=1, lty=3
     main="", sub="Función de masa\nde probabilidad", xlab="", ylab="f(x)")
points(x, y, pch=20, col="blue")
```



- Special Distribution Calculator
- Binomial Coin Experiment
- Binomial Distribution
- Binomial Distribution Applet/Calculator

Algunas situaciones en donde aplica:

- Número de caras obtenidas al lanzar una moneda n veces.
- Número de veces que gano al lanzar un dado n veces (gano si sale 5 o 6).
- Número de elementos e individuos que tienen una característica, de n seleccionados con reemplazamiento ó de tal forma que garantizan el que son independientes entre sí.

Hipergeométrica

Ejercicio 0.34.

¿Cuál es la probabilidad de que una camarera se rehúse a servir bebidas alcohólicas a sólo 2 menores si verifica al azar 5 identificaciones de 9 estudiantes, de los cuales 4 son menores de edad?

Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación. Ejercicio 5.34.

Características:

La variable aleatoria corresponde al número de "aciertos" de n experimentos Bernoulli realizados con respecto a un total de N posibles, donde se sabe que R (de los posibles N) son efectivamente "aciertos", además, los experimentos son realizados simultaneamente, sin reemplazamiento o son dependientes entre sí.

Valores que puede tomar la variable:

$x \in \mathbb{Z}$ con $\max\{0, n-(N-R)\} \leq x \leq \min\{n, R\}$

Parámetros:

$N \in \mathbb{Z}^+$, número de experimentos Bernoulli posibles, $n \in \mathbb{Z}^+$ con $n \leq N$, número de experimentos Bernoulli realizados, y $R \in \mathbb{Z}^+$ con $R \leq N$, número total de todos los posibles "aciertos".

Notación:

$X \sim Hg(n, N, R)$, esto se lee así: la variable aleatoria X tiene una distribución Hipergeométrica de parámetros n, N y R.

Función de masa de probabilidad:

$$f_X(x) = \begin{cases} \frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}} & \text{Si } x \text{ es un valor que puede tomar la variable} \\ 0 & \text{en otro caso} \end{cases}$$

Función de distribución acumulativa:

$$F_X(x) = P[X \leq x] = \sum_{i=0}^{\lfloor x \rfloor} \left[\frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}} \right]$$

Valor esperado y varianza:

- $\mu_X = E[X] = n\,\frac{R}{N}$.
- $\sigma_X^2 = Var[X] = n\,\frac{R}{N}\,\left(1-\frac{R}{N}\right)\left(\frac{N-n}{N-1}\right)$.

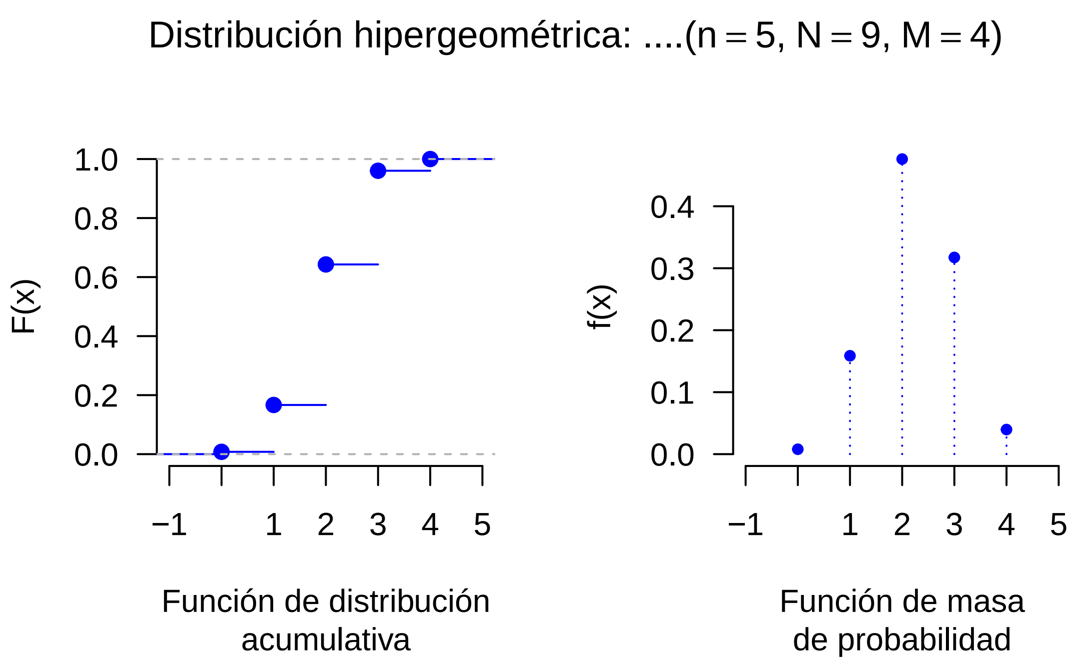
Función característica:

$$\varphi_X(t) = \frac{\binom{N-R}{n}\,{}_2F_1\left(-n, -R; N-R-n+1; \mathrm{e}^{\mathrm{i}t}\right)}{\binom{N}{n}}$$

donde ${}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; z) = \sum_{n=0}^{\infty} \frac{(a_1)_n \cdots (a_p)_n}{(b_1)_n \cdots (b_q)_n} \frac{z^n}{n!}$ (función hipergeométrica generalizada).

Para ilustrar:

```r
layout(matrix(c(1,2), 1, 2))
N <- 9
M <- 4
n <- 5
x <- max(0, n-(N-M)):min(n, M)
y <- dhyper(x, M, N-M, n)
cdf <- approxfun(x, cumsum(y), method = "constant", yleft = 0, yright = 1, f = 0,
                 ties = "ordered")
class(cdf) <- c("ecdf", "stepfun", class(cdf))
attr(cdf, "call") <- sys.call()
plot(cdf, bty="n", col="blue", xlim=c(x[1]-1, x[length(x)]+1), las=1, main="",
     sub="Función de distribución\nacumulativa", xlab="", ylab="F(x)")
title(bquote("Distribución hipergeométrica:  ("*n == .(n)*","~N == .(N)*","~M == .(M)*")")
      line= -1.25, outer=TRUE)
plot(x, y, type= "h", bty="n", col="blue", xlim=c(x[1]-1, x[length(x)]+1), ylim=c(0, max(y
     main="", sub="Función de masa\nde probabilidad", xlab="", ylab="f(x)")
points(x, y, pch=20, col="blue")
```



- Special Distribution Calculator
- The Ball and Urn Experiment
- Hyper-Geometric Distribution Applet/Calculator

Algunas situaciones en donde aplica:

- Número de balotas rojas obtenidas al extraer, simultáneamente o sin reemplazamiento, n balotas de una urna con un total de N balotas, donde se sabe que hay M rojas.
- Número de productos defectuosos obtenidos al extraer, simultáneamente o sin reemplazamiento, n de un lote de N, donde se sabe que hay M defectuosos.

Binomial negativa

Ejercicio 0.35.

> Una pareja decide que continuará procreando hijos hasta tener dos hombres. Suponiendo que $P[hombre] = 0.5$, ¿cuál es la probabilidad de que su segundo niño sea su cuarto hijo? (*A couple decides to continue to have children until they have two males. Assuming that* $P[male] = 0.5$, *what is the probability that their second male is their fourth child?*)
>
> Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación. Ejercicio 5.92.

Forma I

Características:

Número de "fallos" obtenidos al repetir un experimento Bernoulli las veces necesarias para obtener r "aciertos" (el experimento se detiene en el momento en el que sale el r-ésimo "acierto").

Valores que puede tomar la variable:

$x = 0, 1, 2, ...$

Parámetros:

$r \in \mathbb{Z}^+$, número de aciertos deseados, y $p \in [0, 1]$, probabilidad de "acierto".

Notación:

$X \sim \mathcal{BN}_{(I)}(r, p)$.

Función de masa de probabilidad:

$$f_X(x) = \begin{cases} \binom{x+r-1}{r-1} p^r (1-p)^x & \text{Si } x = 0, 1, ... \\ 0 & \text{en otro caso} \end{cases}$$

Valor esperado y varianza

- $\mu_X = E[X] = \frac{r(1-p)}{p}$.
- $\sigma_X^2 = Var[X] = \frac{r(1-p)}{p^2}$.

Función característica:

$$\varphi_X(t) = \left(\frac{p}{1-(1-p)\mathrm{e}^{\mathrm{i}t}} \right)^r$$

Forma II

Características:

Número de repeticiones de un experimento Bernoulli necesarias para obtener r "aciertos" (el experimento se detiene en el momento en el que sale el r-ésimo "acierto").

Valores que puede tomar la variable:

$x = r, r+1, r+2, ...$

Parámetros:

$r \in \mathbb{Z}^+$, número de aciertos deseados, y $p \in [0,1]$, probabilidad de "acierto".

Notación:

$X \sim \mathcal{BN}_{(II)}(r,p)$.

Función de masa de probabilidad:

$$f_X(x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r} & \text{Si } x = r, r+1, ... \\ 0 & \text{en otro caso} \end{cases}$$

Valor esperado y varianza:

- $\mu_X = E[X] = \frac{r}{p}$
- $\sigma_X^2 = Var[X] = \frac{r(1-p)}{p^2}$

⚠ Note que la Forma I y la Forma II están relacionadas entre sí (son en realidad dos formas de ver una misma distribución):

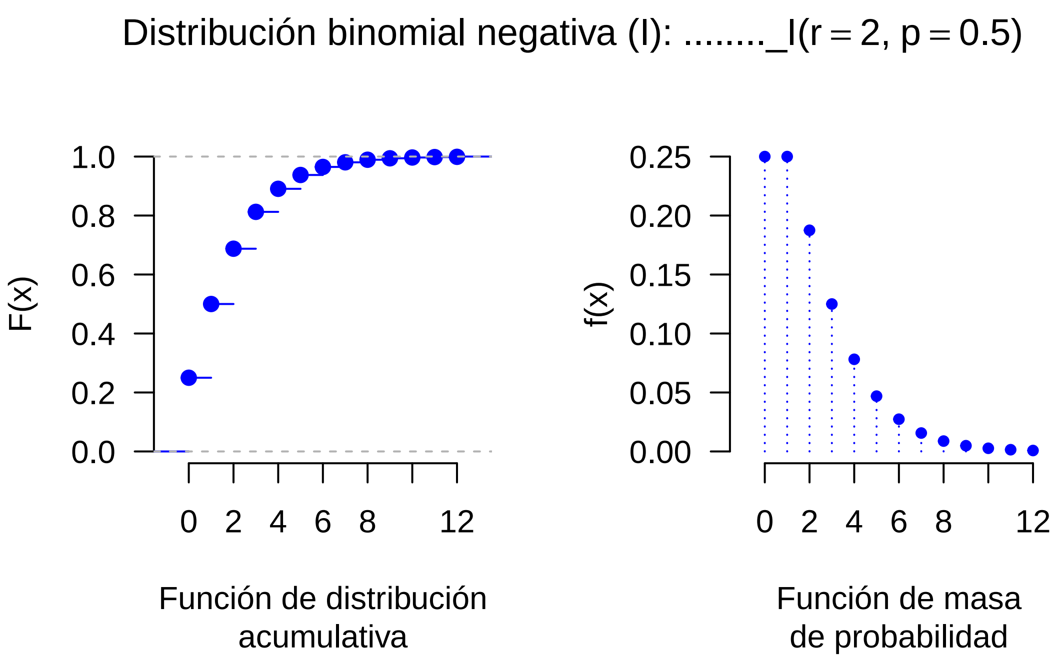
- Si $X \sim \mathcal{BN}_{(I)}$, entonces $Y = X + r \sim \mathcal{BN}_{(II)}$, y
- Si $X \sim \mathcal{BN}_{(II)}$, entonces $Y = X - r \sim \mathcal{BN}_{(I)}$.

Geométrica

Cuando $r = 1$, se dice que la variable aleatoria X tiene una distribución geométrica de parámetro $p \in [0, 1]$, probabilidad de "acierto", y se denota de la siguiente manera: $X \sim \mathcal{G}(p)$.

Para ilustrar:

```
layout(matrix(c(1,2), 1, 2))
r <- 2
p <- 0.5
max_x <- 12
x <- 0:max_x
y <- dnbinom(x, r, p)
cdf <- approxfun(x, cumsum(y), method = "constant", yleft = 0, yright = 1, f = 0,
                 ties = "ordered")
class(cdf) <- c("ecdf", "stepfun", class(cdf))
attr(cdf, "call") <- sys.call()
plot(cdf, bty="n", col="blue", xlim=c(-1, max_x+1), las=1, main="",
     sub="Función de distribución\nacumulativa", xlab="", ylab="F(x)")
title(bquote("Distribución binomial negativa (I):  _I("*r == .(r)*","~p == .(p)*")"),
      line= -1.25, outer=TRUE)
plot(x, y, type= "h", bty="n", col="blue", xlim=c(-1, max_x+1), ylim=c(0, max(y)), las=1,
     main="", sub="Función de masa\nde probabilidad", xlab="", ylab="f(x)")
points(x, y, pch=20, col="blue")
```



- Special Distribution Calculator
- Negative Binomial Experiment
- Negative Binomial Distribution Applet/Calculator (I)
- Negative Binomial Distribution Applet/Calculator (II)

Algunas situaciones en donde aplica:

- Número de lanzamientos de una moneda necesarios para obtener r caras.
- Número de veces que debo lanzar un dado para ganar r veces (gano si sale 5 o 6).
- Número de elementos o individuos independientes que debo observar hasta encontrar r que tienen una característica.

Poisson

Ejercicio 0.36.

> Los baches en ciertas carreteras pueden ser un problema grave y requieren reparación constantemente. Con un tipo específico de terreno y mezcla de concreto la experiencia sugiere que hay, en promedio, 2 baches por milla después de cierta cantidad de uso. Se supone que el proceso de Poisson se aplica a la variable aleatoria "número de baches".
>
> - ¿Cuál es la probabilidad de que no aparezca más de un bache en un tramo de una milla?
> - ¿Cuál es la probabilidad de que no aparezcan más de 4 baches en un tramo determinado de 5 millas?
>
> Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación. Ejercicio 5.72.

Proceso Poisson:

- El número de ocurrencias en un intervalo es independiente del número de ocurrencias en cualquier otro intervalo disyunto.
- La probabilidad de una ocurrencia en un intervalo es proporcional al tamaño del intervalo.

Características:

La variable aleatoria corresponde al número de ocurrencias dentro de un intervalo de cierta amplitud fija, y que cumple las condiciones de un proceso de Poisson, donde λ es el promedio de ocurrencias por intervalo.

Valores que puede tomar la variable:

$x = 0, 1, 2, ...$

Parámetros:

λ, promedio de ocurrencias en un intervalo de la correspondiente amplitud fijada.

Notación:

$X \sim \mathcal{P}(\lambda)$.

Función de masa de probabilidad:

$$f_X(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & \text{Si } x = 0, 1, 2, ... \\ 0 & \text{en otro caso} \end{cases}$$

Valor esperado y varianza:

- $\mu_X = E[X] = \lambda$.
- $\sigma_X^2 = Var[X] = \lambda$.

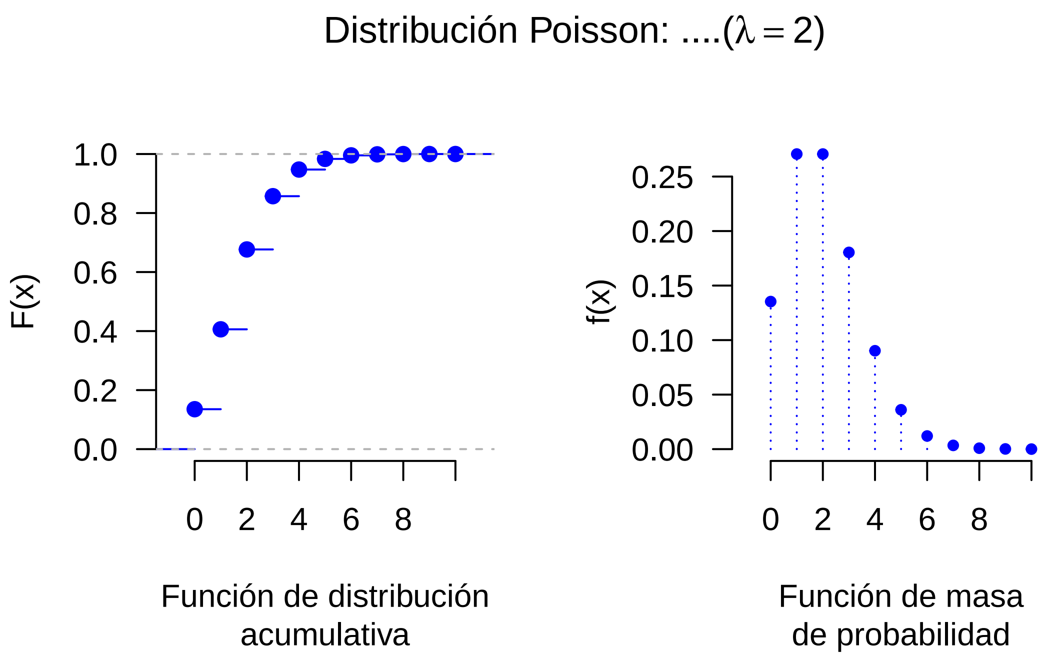
Función característica:

$$\varphi_X(t) = \exp\left[\lambda\left(\mathrm{e}^{\mathrm{i}t} - 1\right)\right]$$

Para ilustrar:

```
layout(matrix(c(1,2), 1, 2))
l <- 2
max_x <- 10
x <- 0:max_x
y <- dpois(x, l)
cdf <- approxfun(x, cumsum(y), method = "constant", yleft = 0, yright = 1, f = 0,
                 ties = "ordered")
class(cdf) <- c("ecdf", "stepfun", class(cdf))
attr(cdf, "call") <- sys.call()
plot(cdf, bty="n", col="blue", xlim=c(-1, max_x+1), las=1, main="",
     sub="Función de distribución\nacumulativa", xlab="", ylab="F(x)")
title(bquote("Distribución Poisson:  ("*lambda == .(l)*")"),
```

```
      line= -1.25, outer=TRUE)
plot(x, y, type= "h", bty="n", col="blue", xlim=c(-1, max_x+1), ylim=c(0, max(y)), las=1,
     main="", sub="Función de masa\nde probabilidad", xlab="", ylab="f(x)")
points(x, y, pch=20, col="blue")
```



- Special Distribution Calculator
- The Poisson Experiment
- Poisson Distribution
- Poisson Distribution Applet/Calculator

Algunas situaciones en donde aplica:

- Número de llamadas que llegan a un centro de contacto (call center) por minuto.
- Número de huecos por kilómetro de vía.

Algunas aproximaciones

Se estima que 1000 de los 10000 votantes registrados en un municipio están a favor de un nuevo impuesto. Si de los 10000 se escogen de manera aleatoria 100 y se les pregunta su opinión, ¿Cuál es la probabilidad de que al menos 90 estén en contra?

$$X \sim \mathcal{H}(n, N, M) \underset{\frac{n}{N} \to 0}{\longrightarrow} Y \sim \mathcal{B}\left(n,\, p = \frac{M}{N}\right)$$

$$X \sim \mathcal{B}(n, p) \underset{n \to \infty \,\wedge\, p \to 0}{\longrightarrow} Y \sim \mathcal{P}(\lambda = (n)(p))$$

En este segundo caso, si $p \to 1$, la situación se puede replantear creando una nueva variable aleatoria, en donde intercambiamos “fracaso” (0) por “acierto” (1), y viceversa.

Ejercicios

A continuación se encuentran una serie de ejercicios tomados del siguiente libro:

> Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación.

Ejercicio 0.37 (Walpole 5.77)**.**

> Durante un proceso de producción, cada día se seleccionan al azar 15 unidades de la línea de ensamble para verificar el porcentaje de artículos defectuosos. A partir de información histórica se sabe que la probabilidad de tener una unidad defectuosa es de 0.05. Cada vez que se encuentran dos o más unidades defectuosas en la muestra de 15, el proceso se detiene. Este procedimiento se utiliza para proporcionar una señal en caso de que aumente la probabilidad de unidades defectuosas.
>
> a. ¿Cuál es la probabilidad de que en un día determinado se detenga el proceso de producción? (Suponga 5% de unidades defectuosas).
>
> b. Suponga que la probabilidad de una unidad defectuosa aumenta a 0.07. ¿Cuál es la probabilidad de que en cualquier día no se detenga el proceso de producción?

Ejercicio 0.38 (Walpole 5.84)**.**

> El propietario de una farmacia local sabe que, en promedio, llegan a su farmacia 100 personas por hora.
>
> a. Calcule la probabilidad de que en un periodo determinado de 3 minutos nadie entre a la farmacia.
>
> b. Calcule la probabilidad de que en un periodo dado de 3 minutos entren más de 5 personas a la farmacia.

Ejercicio 0.39 (Walpole 5.91 (5.90))**.**

Considere la información del ejercicio de repaso 5.90. El perforador cree que "dará en el clavo" si logra el segundo éxito durante o antes del sexto intento. ¿Cuál es la probabilidad de que el perforador "dé en el clavo"?

Ejercicio 5.90: Una empresa que perfora pozos petroleros opera en varios sitios y su éxito o fracaso es independiente de un sitio a otro. Suponga que la probabilidad de éxito en cualquier sitio específico es de 0.25.

a. ¿Cuál es la probabilidad de que un perforador barrene 10 sitios y tenga un éxito?

b. El perforador se declarará en bancarrota si tiene que perforar 10 veces antes de que ocurra el primer éxito. ¿Cuáles son las perspectivas de bancarrota del perforador?

Ejercicio 0.40 (Walpole 5.95)**.**

Un proceso de manufactura produce artículos en lotes de 50. Se dispone de planes de muestreo en los cuales los lotes se apartan periódicamente y se someten a cierto tipo de inspección. Por lo general se supone que la proporción de artículos defectuosos que resultan del proceso es muy pequeña. Para la empresa también es importante que los lotes que contengan artículos defectuosos sean un evento raro. El plan actual de inspección consiste en elegir lotes al azar, obtener muestras periódicas de 10 en 50 artículos de un lote y, si ninguno de los muestreados está defectuoso, no se realizan acciones.

a. Suponga que se elige un lote al azar y 2 de cada 50 artículos tienen defecto. ¿Cuál es la probabilidad de que al menos uno en la muestra de 10 del lote esté defectuoso?

b. A partir de su respuesta en el inciso a), comente sobre la calidad de este plan de muestreo.

c. ¿Cuál es el número promedio de artículos defectuosos encontrados por cada 10 artículos de la muestra?

EN CONSTRUCCION.

Distribuciones continuas

EN CONSTRUCCION.

En esta sección se hará una revisión de algunos temas relacionados con modelos/distribuciones de variables aleatorias continuas.

Uniforme Continua

Ejercicio 0.41.

> Un autobús llega cada 10 minutos a una parada. Se supone que el tiempo de espera para un individuo en particular es una variable aleatoria con **distribución continua uniforme**.
>
> a. ¿Cuál es la probabilidad de que el individuo espere más de 7 minutos?
>
> b. ¿Cuál es la probabilidad de que el individuo espere entre 2 y 7 minutos?
>
> Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación. Ejercicio 6.4.

Características:

Extiende, al caso continuo, la idea de que todos los valores de la variable aleatoria X son igualmente probables.

Dominio (valores que puede tomar la variable aleatoria):

$x \in [a, b]$.

Parámetros:

a, límite inferior del dominio de la variable aleatoria, y b, límite superior del dominio de la variable a aleatoria. $-\infty < a < b < \infty$

Notación:

$X \sim \mathcal{U}(a, b)$, esto se lee así: la variable aleatoria X tiene una distribución uniforme continua de parámetros a y b.

Función de densidad:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{Si } a < x < b \\ 0 & \text{en otro caso} \end{cases}$$

Función de distribución acumulativa:

$$F_X(x) = \begin{cases} 0 & \text{Si } x < a \\ \frac{x-a}{b-a} & \text{Si } a \leq x < b \\ 1 & \text{Si } b \leq x \end{cases}$$

Valor esperado y varianza:

- $\mu_X = E[X] = \frac{a+b}{2}$
- $\sigma_X^2 = Var[X] = \frac{(b-a)^2}{12}$

Función característica:

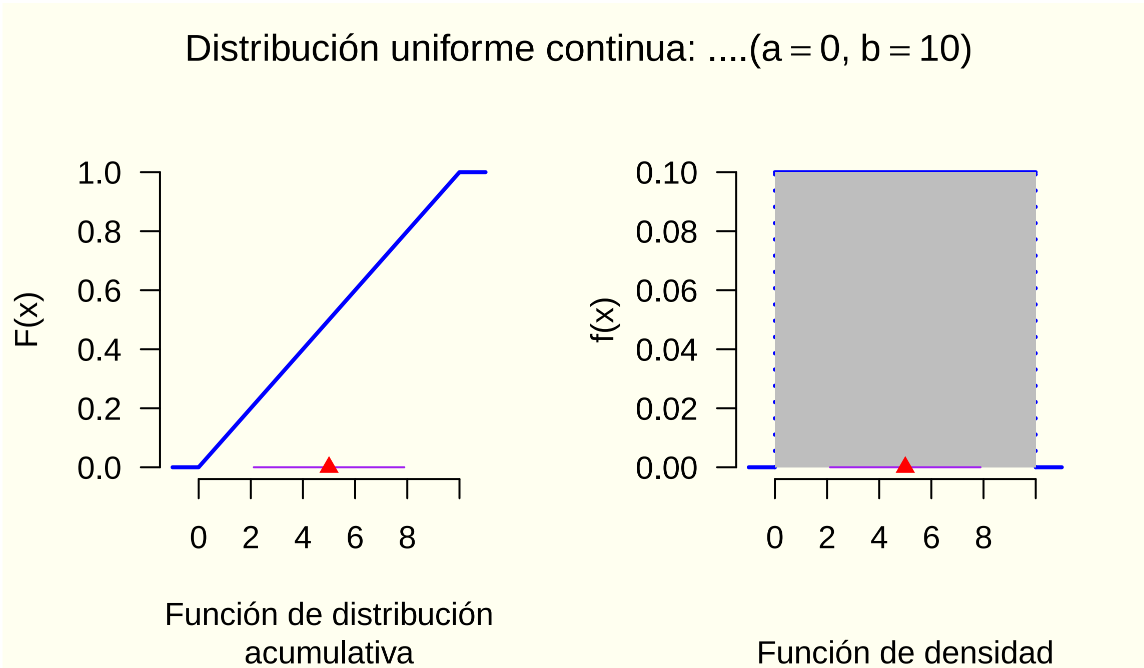
$$\varphi_X(t) = \begin{cases} \frac{\mathrm{e}^{\mathrm{i}tb} - \mathrm{e}^{\mathrm{i}ta}}{\mathrm{i}t(b-a)} & \text{for } t \neq 0 \\ 1 & \text{for } t = 0 \end{cases}$$

Para ilustrar:

- Funciones de distribución acumulativa y de densidad para la variable aleatoria descrita en el Ejercicio 0.41:

```
par(bg = "ivory")
layout(matrix(c(1,2), 1, 2))
a <- 0
b <- 10
mu <- (a + b) / 2
sigma <- sqrt(((b - a)^2) / 12)
y <- 1/(b-a)
plot(c(a-1,a,b,b+1), c(0,0,1,1), type="l", bty="n", col="blue", xlim=c(a-1,b+1), las=1, lw
     sub="Función de distribución\nacumulativa", xlab="", ylab="F(x)")
lines(mu + c(-1,1) * sigma, c(0, 0), col="purple")
points(mu, 0, pch=17, col="red")
plot(c(a,b), c(y,y), type= "l", bty="n", col="blue", xlim=c(a-1,b+1), ylim=c(0,max(y)), la
     main="", sub="Función de densidad", xlab="", ylab="f(x)")
```

```r
segments(c(a,b), c(0,0), c(a,b), c(y,y), lty=3, col="blue", lwd=2)
segments(c(a-1,b), c(0,0), c(a,b+1), c(0,0), lty=1, col="blue", lwd=2)
polygon(c(a,a,b,b), c(0,y,y,0), col="grey", border=NA)
lines(mu + c(-1,1) * sigma, c(0, 0), col="purple")
points(mu, 0, pch=17, col="red")
title(bquote("Distribución uniforme continua:  ("*a == .(a)*","~b == .(b)*")"),
      line= -1.25, outer=TRUE)
```



- Special Distribution Calculator

Ejercicio 0.42. Supongamos que el tiempo de desplazamiento en TM de mi casa a la Universidad puede ser cualquier valor entre 40 y 60 minutos, todos igualmente probables. Si a las 10:05 am estoy tomando el TM desde la estación que queda cerca a mi casa, ¿cuál es la probabilidad de que llegue tarde a clase de 11 am?, ¿En promedio a que horas se espera que llegue?

Distribución Exponencial

Ejercicio 0.43.

Cierto tipo de dispositivo tiene una tasa de fallas anunciada de 0.01 por hora. La tasa de fallas es constante y se aplica la **distribución exponencial**.

a. ¿Cuál es el tiempo promedio que transcurre antes de la falla?

b. ¿Cuál es la probabilidad de que pasen 200 horas antes de que se observe una falla?

Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación. Ejercicio 6.66.

Características:

La longitud de intervalo entre ocurrencias consecutivas de un proceso Poisson es una variable aleatoria X con distribución exponencial.

Dominio (valores que puede tomar la variable aleatoria):

$x \in [0, \infty)$.

Parámetro de escala

Parámetros:

β, promedio de la variable aleatoria (longitud de intervalo promedio entre ocurrencias consecutivas de un proceso Poisson). $\beta > 0$

Notación:

$X \sim \mathcal{E}(\beta)$, esto se lee así: la variable aleatoria X tiene una distribución exponencial de parámetro (de escala) β.

Función de densidad:

$$f_X(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & \text{Si } x \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

Función de distribución:

$$F_X(x) = \begin{cases} 0 & \text{Si } x < 0 \\ 1 - e^{-x/\beta} & \text{Si } 0 \leq x \end{cases}$$

Valor esperado y varianza:

- $\mu_X = E[X] = \beta$

- $\sigma_X^2 = Var[X] = \beta^2$

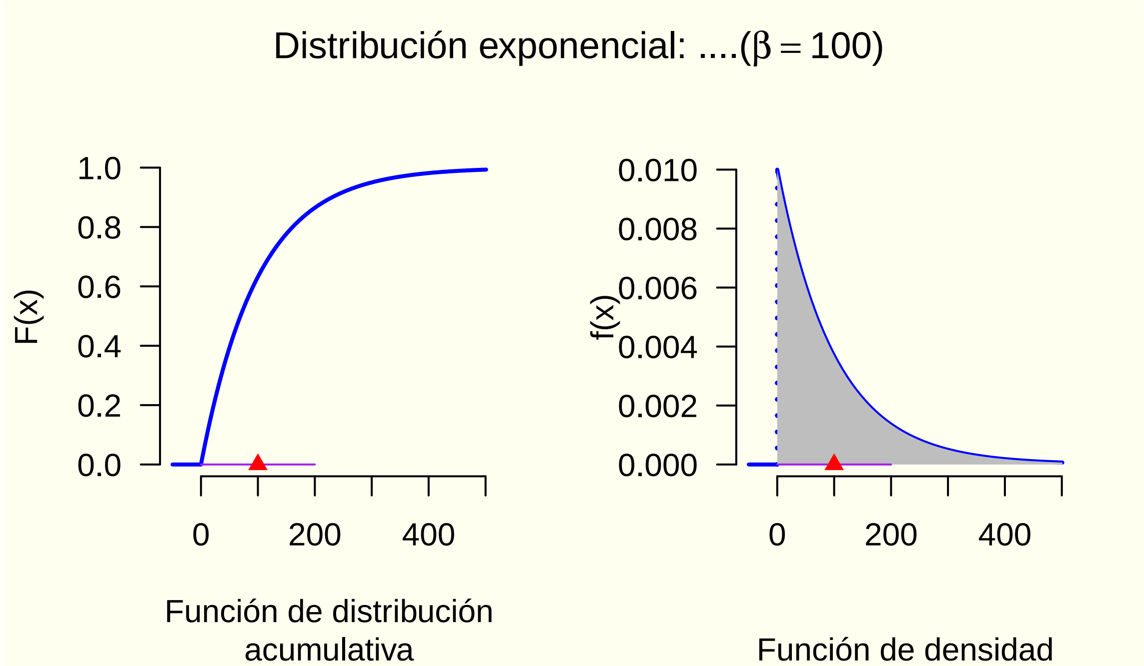
Función característica:

$$\varphi_X(t) = \frac{1}{1 - \mathrm{i}\beta t}$$

Para ilustrar:

- Funciones de distribución acumulativa y de densidad para la variable aleatoria descrita en el Ejercicio 0.43:

```
par(bg = "ivory")
layout(matrix(c(1,2), 1, 2))
lambda <- 0.01
beta <- 1 / lambda
x_max <- 501
x <- seq(0, x_max, length.out=1e3)
aux <- exp(-lambda * x)
fx <- lambda * exp(-lambda * x)
Fx <- 1 - exp(-lambda * x)
plot(x, Fx, type="l", bty="n", col="blue", xlim=c(-50, x_max-1), las=1, lwd=2, main="",
     sub="Función de distribución\nacumulativa", xlab="", ylab="F(x)")
title(bquote("Distribución exponencial:  ("*beta == .(beta)*")"),
      line= -1.25, outer=TRUE)
segments(-50, 0, 0, 0, lty=1, col="blue", lwd=2)
lines(beta + c(-1,1) * beta, c(0, 0), col="purple")
points(beta, 0, pch=17, col="red")
plot(x, fx, type= "l", bty="n", col="blue", xlim=c(-50, x_max-1), ylim=c(0, fx[1]), las=1,
     main="", sub="Función de densidad", xlab="", ylab="f(x)")
segments(-50, 0, 0, 0, lty=1, col="blue", lwd=2)
segments(0, 0, 0, fx, lty=3, col="blue", lwd=2)
polygon(c(0,0,x,x_max), c(0,fx[1],fx,0), col="grey", border=NA)
lines(beta + c(-1,1) * beta, c(0, 0), col="purple")
points(beta, 0, pch=17, col="red")
```



- Special Distribution Calculator
- Exponential Distribution Applet/Calculator

Parámetro de tasa

Relación con la distribución Poisson:

Si Y := "número de ocurrencias" y $Y \sim \mathcal{P}(\lambda)$, entonces X := "longitud de intervalo entre ocurrencias" tiene distribución exponencial de parámetro $\beta = \frac{1}{\lambda}$.

Notación:

$X \sim \mathcal{E}(\lambda)$, esto se lee así: la variable aleatoria X tiene una distribución exponencial de parámetro (de tasa) λ.

Función de densidad:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{Si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Función de distribución acumulativa:

$$F_X(x) = \begin{cases} 0 & \text{Si } x < 0 \\ 1 - e^{-\lambda x} & \text{Si } 0 \leq x \end{cases}$$

Valor esperado y varianza:

- $\mu_X = E[X] = \frac{1}{\lambda}$
- $\sigma_X^2 = Var[X] = \frac{1}{\lambda^2}$

Función característica:

$$\varphi_X(t) = \frac{\lambda}{\lambda - \mathrm{i}t}$$

Para ilustrar:

- Exponential Distribution Applet/Calculator

Ejercicio 0.44. Supongamos que un estudiante siempre llega tarde a clase, que la cantidad de tiempo de clase que se pierde sigue una distribución exponencial y que su promedio de llegada tarde es de 10 minutos, ¿cuál es la probabilidad de que llegue a clase después de las 11:15 am?. Si después de llegar media hora tarde decide no entrar. ¿De 32 clases al semestre, a cuántas se esperaría que no entre?

Distribución Gaussiana

Ejercicio 0.45.

En un proyecto experimental sobre el factor humano se determinó que el tiempo de reacción de un piloto ante un estímulo visual es **distribuido normalmente** con una media de 1/2 segundo y una desviación estándar de 2/5 de segundo.

a) ¿Cuál es la probabilidad de que una reacción del piloto tome más de 0.3 segundos?
b) ¿Qué tiempo de reacción se excede el 95% de las veces?

Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación. Ejercicio 6.80.

Características:

"La distribución normal, distribución de Gauss, distribución gaussiana o distribución de Laplace-Gauss, es una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en estadística y en la teoría de probabilidades.

La gráfica de su función de densidad tiene una forma acampanada y es simétrica respecto de un determinado parámetro estadístico. Esta curva se conoce como campana de Gauss y es el gráfico de una función gaussiana.

La importancia de esta distribución radica en que permite modelar numerosos fenómenos naturales, sociales y psicológicos.

Algunos ejemplos de variables asociadas a fenómenos naturales que siguen el modelo de la normal son:

- caracteres morfológicos de individuos como la estatura;
- caracteres fisiológicos como el efecto de un fármaco;
- caracteres sociológicos como el consumo de cierto producto por un mismo grupo de individuos;
- caracteres psicológicos como el cociente intelectual;
- nivel de ruido en telecomunicaciones;
- errores cometidos al medir ciertas magnitudes;
- etc.

La distribución normal también aparece en muchas áreas de la propia estadística."

- Galton Board
- The Galton Board Experiment

Dominio (valores que puede tomar la variable aleatoria):

$x \in \mathbb{R}$.

Parámetros:

μ, media de la variable aleatoria, y σ, desviación estándar de la variable a aleatoria. $-\infty < \mu < \infty$ y $\sigma > 0$

Notación:

$X \sim \mathcal{N}(\mu, \sigma)$, esto se lee así: la variable X tiene una distribución normal de parámetros μ (localización) y σ (escala).

Función de densidad:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Función de distribución acumulativa:

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(t)\,dt \\ &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}\,dt \end{aligned}$$

Valor esperado y varianza:

- $\mu_X = E[X] = \mu$
- $\sigma_X^2 = Var[X] = \sigma^2$

Función característica:

$$\varphi_X(t) = \exp\left(\mathrm{i}\mu t - \frac{\sigma^2 t^2}{2}\right)$$

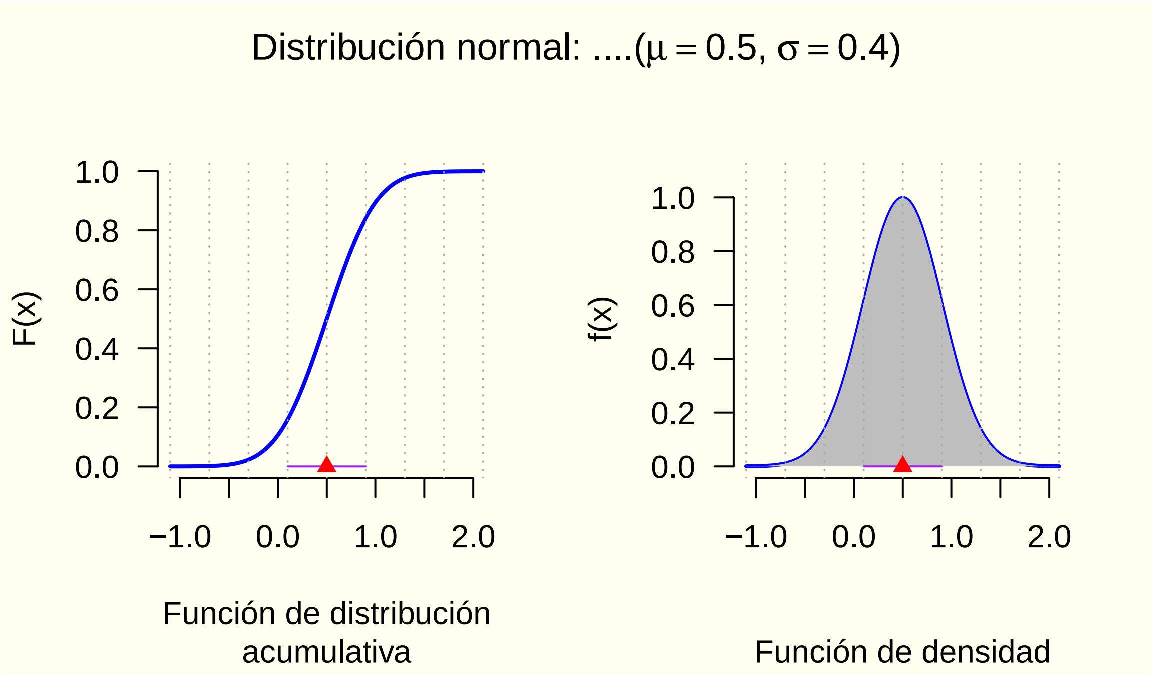
Para ilustrar:

> Efecto de los parámetros
>
> https://cjtorresj.shinyapps.io/introNormalDistribution/

- Funciones de distribución acumulativa y de densidad para la variable aleatoria descrita en el Ejercicio 0.45:

```
par(bg = "ivory")
layout(matrix(c(1,2), 1, 2))
mu <- 0.5
sigma <- 0.4
x_max <- mu + 4 * sigma
x_min <- mu - 4 * sigma
x <- seq(x_min, x_max, length.out=1e3)
fx <- dnorm(x, mu, sigma)
Fx <- pnorm(x, mu, sigma)
plot(x, Fx, type="l", bty="n", col="blue", xlim=c(x_min, x_max), las=1, lwd=2, main="",
     sub="Función de distribución\nacumulativa", xlab="", ylab="F(x)")
abline(v = mu + (-5:5) * sigma, lty=3, col="darkgrey")
lines(mu + c(-1,1) * sigma, c(0, 0), col="purple")
points(mu, 0, pch=17, col="red")
plot(x, fx, type= "l", bty="n", col="blue", xlim=c(x_min, x_max), ylim=c(0, 1.1*dnorm(mu,
     main="", sub="Función de densidad", xlab="", ylab="f(x)")
```

```r
polygon(c(x_min,x,x_max), c(0,fx,0), col="grey", border=NA)
abline(v = mu + (-5:5) * sigma, lty=3, col="darkgrey")
lines(mu + c(-1,1) * sigma, c(0, 0), col="purple")
points(mu, 0, pch=17, col="red")
title(bquote("Distribución normal:  ("*mu == .(mu)*","~sigma == .(sigma)*")"),
      line= -1.25, outer=TRUE)
```



- Special Distribution Calculator
- Normal Distribution Applet/Calculator

⚠ Note que la función de distribución acumulativa de una variable aleatoria gaussiana simplemente se dejó expresada y ya. Resulta que no se conoce una expresión cerrada simple para la antiderivada de la función de densidad de una variable aleatoria gaussiana. Entonces, ¿cómo se supone que voy a obtener las probabilidades que necesito, si la integral correspondiente no tiene solución analítica?, recordemos que para una variable aleatoria continua: $P[x_1 < X < x_2] = \int_{x_1}^{x_2} f_X(t)\, dt$.

Respuesta

La solución a la anterior inquietud sería la siguiente: para una variables aleatoria $X \sim \mathcal{N}(\mu, \sigma)$, **transformamos** el problema $P[x_1 < X < x_2]$ en un problema equivalente que use una variable aleatoria $Z \sim \mathcal{N}(0,1)$, luego, usando aproximaciones numéricas de $F_Z(z) = \int\limits_{-\infty}^{z} f_Z(t)\,dt$, encontramos la respuesta al problema equivalente.

Transformación (Estandarización):
Si $X \sim \mathcal{N}(\mu, \sigma)$ y $Z = \frac{X-\mu}{\sigma}$, entonces $Z \sim \mathcal{N}(0,1)$.

$$\begin{aligned} P[x_1 < X < x_2] &= P\left[x_1 - \mu < X - \mu < x_2 - \mu\right] \\ &= P\left[\frac{x_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{x_2 - \mu}{\sigma}\right] \\ &= P\left[z_1 < Z < z_2\right], \end{aligned}$$

en donde,

$$z_1 = \frac{x_1 - \mu}{\sigma} \quad \text{y} \quad z_2 = \frac{x_2 - \mu}{\sigma},$$

además ya sabíamos que,

$$\begin{aligned} P\left[z_1 < Z < z_2\right] &= P\left[Z < z_2\right] - P\left[Z < z_1\right] \\ &= F_Z\left(z_2\right) - F_Z\left(z_1\right), \end{aligned}$$

y las aproximaciones numéricas para $F_Z(z)$ se encuentran tabuladas (por ejemplo, ver libro de Anderson. Página FM1 o 978) o se pueden obtener mediante herramientas electrónicas, incluyendo algunas calculadoras.

Dentro del conjunto de variables aleatorias que tienen distribución gaussiana, $Z \sim \mathcal{N}(0,1)$ es evidentemente una variable particularmente especial. Este caso especial se denomina distribución normal **estándar** (con media igual a cero y desviación estándar igual a uno) y naturalmente su función de densidad es,

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Ejercicios

Ejercicio 0.46.

Dada una distribución normal con $\mu = 30$ y $\sigma = 6$, calcule

a. el área de la curva normal a la derecha de $x = 17$;

b. el área de la curva normal a la izquierda de $x = 22$;

c. el área de la curva normal entre $x = 32$ y $x = 41$;

d. el valor de x que tiene 80 del área de la curva normal a la izquierda;

e. los dos valores de x que contienen 75 central del área de la curva normal.

Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación. Ejercicio 6.8.

Ejercicio 0.47.

Dada la variable X normalmente distribuida con una media de 18 y una desviación estándar de 2.5, calcule

a) $P(X < 15)$;

b) el valor de k tal que $P(X < k) = 0.2236$;

c) el valor de k tal que $P(X > k) = 0.1814$;

d) $P(17 < X < 21)$.

Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación. Ejercicio 6.9.

Ejercicio 0.48. Supongamos que el tiempo de desplazamiento en TM de mi casa a la Universidad tiene una distribución normal con media 50 minutos y desviación estándar 5 minutos. Si a las 10:05 am estoy tomando el TM desde la estación cerca de mi casa, ¿cuál es la probabilidad de que llegue tarde a clase de 11 am?. ¿De 32 clases, a cuántas se espera que llegue tarde?

Ejercicio 0.49.

El **precio medio** de las acciones de las empresas que forman el S&P 500 es \$30, y la **desviación estándar** es \$8.20 (BusinessWeek, publicación anual especial, primavera de 2003). Suponga que los precios de las acciones **se distribuyen normalmente**.

1. ¿Cuál es la probabilidad de que las acciones de una empresa tengan un precio mínimo de \$40?
2. ¿Cuál es la probabilidad de que el precio de las acciones no supere \$20?
3. ¿Qué tan alto debe ser el precio de las acciones de una firma para situarla en el 10% de las principales empresas?

Anderson, Sweeney, Williams & Camm (2016). Estadística para Negocios y Economía. (12a. ed.) Cengage Learning. Sección 6.2, ejercicio 18.

Relación con la binomial

Se tiene que,

$$X \sim \mathcal{B}(n,p) \underset{n\to\infty}{\longrightarrow} Y \sim \mathcal{N}\left(np, \sqrt{np(1-p)}\right)$$

Es decir, si $X \sim \mathcal{B}(n,p)$, entonces

$$\frac{X-np}{\sqrt{np(1-p)}} \underset{n\to\infty}{\longrightarrow} Z \sim \mathcal{N}(0,1)$$

Generalmente se debe hacer una **corrección por continuidad** (es decir, tengo que tener en cuenta que estoy utilizando una distribución continua para aproximar una distribución discreta):

Si $X \sim \mathcal{B}(n,p)$, $np > 5$ y $n(1-p) > 5$ entonces,

$$\begin{aligned} P[X \leq x] &\approx P\left[Y \leq (x+0.5)\right] \\ &\approx P\left[Z \leq \frac{(x+0.5)-np}{\sqrt{np(1-p)}}\right] \end{aligned}$$

donde $Y \sim \mathcal{N}(np, \sqrt{np(1-p)})$ y $Z \sim \mathcal{N}(0,1)$.

Buscando ilustrar en alguna medida todo lo anterior, se presenta el siguiente gráfico:

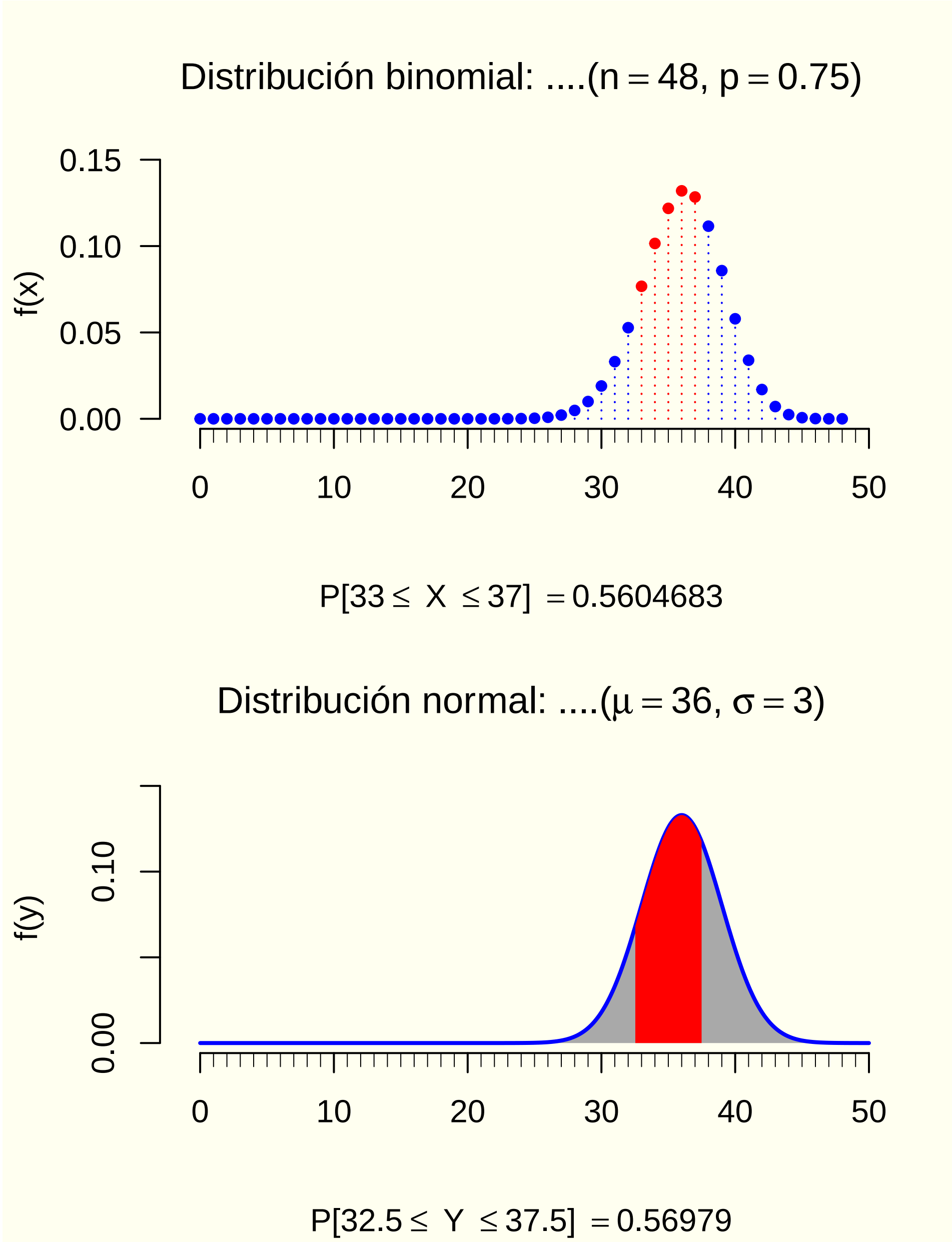
```
par(bg = "ivory")
layout(matrix(c(1,2), 2, 1))

n <- 48
p <- 0.75
x <- 0:n
x1 <- 33
x2 <- 37
idx <- x >= x1 & x <= x2
cols <- ifelse(idx, "red", "blue")
fx <- dbinom(x, n, p)
px <- sum(fx[idx])

plot(x, fx, type= "h", bty="n", col=cols, xlim=c(-1,n+1), ylim=c(0, 1.1*max(fx)),
     las=1, lty=3, xlab="", ylab="f(x)",
     main=bquote("Distribución binomial:  ("*n == .(n)*","~p == .(p)*")"),
     sub=bquote("P["*.(x1) <= ~X~phantom() <= .(x2)*"]"~phantom() == .(px)))
points(x, fx, pch=20, col=cols)
rug(x = x+1, ticksize = -0.05, side = 1)
```

```r
mu <- n * p
sigma <- sqrt(n * p * (1 - p))
y <- seq(0, n+2, length.out=1e3)
fy <- dnorm(y, mu, sigma)
y1 <- x1 - 0.5
y2 <- x2 + 0.5
idy <- (y >= y1 & y <= y2)
py = pnorm(y2, mu, sigma) - pnorm(y1, mu, sigma)

plot(c(-1, n+1), c(0, 1.1*dnorm(mu, mu, sigma)), type= "n", bty="n", xlab="", ylab="f(y)",
     main=bquote("Distribución normal: ("*mu == .(mu)*","~sigma == .(sigma)*")"),
     sub=bquote("P["*.(y1) <= ~Y~phantom() <= .(y2)*"]"~phantom() == .(py)))
polygon(c(0, y, n), c(0, fy, 0), col="darkgrey", border=NA)
lines(y, fy, col="blue", las=1, lwd=2)
polygon(c(min(y[idy]), y[idy], max(y[idy])), c(0, fy[idy], 0), col="red", border=NA)
rug(x = x+1, ticksize = -0.05, side = 1)
```



Ejercicio 0.50. Se cree que 1 de cada 10 votantes registrados en una gran ciudad están a favor de un nuevo impuesto. Si se escogen de manera aleatoria e independiente 1000 votantes y se les pregunta su opinión, ¿cuál es la probabilidad de que al menos 900 estén en contra?

Adicionalmente

Para más información acerca de las relaciones entre distribuciones tanto discretas como continuas, consultar el siguiente enlace: http://www.math.wm.edu/~leemis/chart/UDR/UDR.html

Para más detalles acerca de las características de varias distribuciones discretas y continuas, consultar la siguiente bibliografía clásica básica: Johnson, Kotz y Balakrishnan [2], Johnson, Kotz y Balakrishnan [3], Johnson, Kemp y Kotz [1]

---

[1] N.L. Johnson, A.W. Kemp y S. Kotz. *Univariate Discrete Distributions*. Wiley Series in Probability and Statistics. Wiley, 2005. ISBN: 9780471272465.

[2] N.L. Johnson, S. Kotz y N. Balakrishnan. *Continuous Univariate Distributions, Volume 1*. Wiley Series in Probability and Statistics. Wiley, 1994. ISBN: 9780471584957.

[3] N.L. Johnson, S. Kotz y N. Balakrishnan. *Continuous Univariate Distributions, Volume 2*. Wiley Series in Probability and Statistics. Wiley, 1995. ISBN: 9780471584940.

EN CONSTRUCCION.

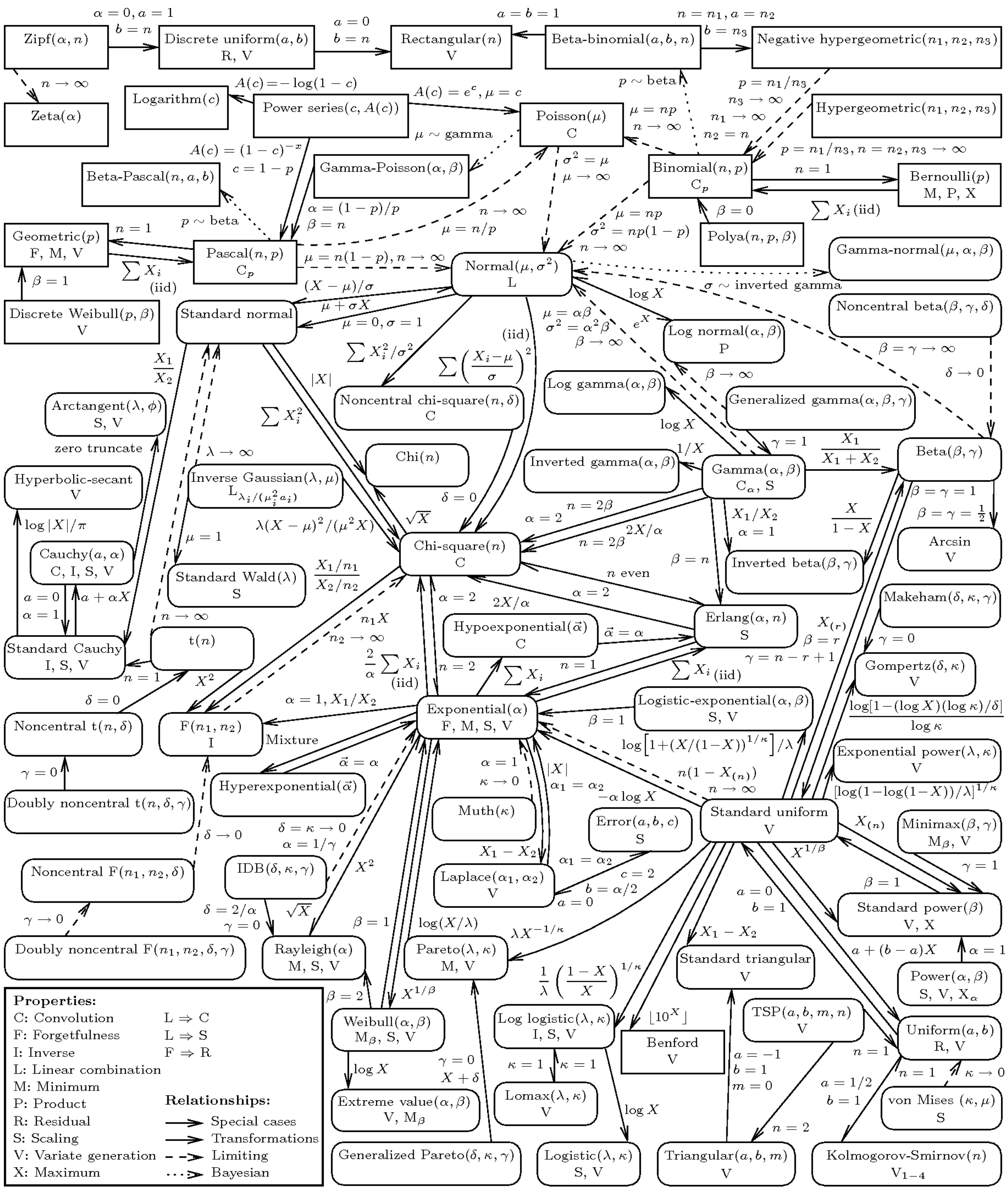


Figura 1: Univariate Distribution Relationship Chart

Transformación inversa de probabilidad

EN CONSTRUCCION.

En esta sección se hará una revisión de la transformación inversa de probabilidad y del método de la transformada inversa para la generación de valores aleatorios.

Método de la transformada inversa

Teorema 0.6 (*Inverse transform sampling*)**.** *Sea U una variable aleatoria uniforme continua en el intervalo $[0,1]$ ($U \sim \mathcal{U}(0,1)$), $F(\cdot)$ una función de distribución acumulativa y $F^{-}(u) = \inf\{x : u \leq F(x)\}$ para $0 < u < 1$, entonces la variable aleatoria $X = F^{-}(U)$ tiene como función de distribución acumulativa $F(\cdot)$.*

Prueba. Como F es continua a derecha, por ser función de distribución acumulativa, entonces $\{x : u \leq F(x)\} = \{x : F^{-}(u) \leq x\}$. Además, como $U \sim \mathcal{U}(0,1)$, entonces, $F_U(u) = u$ para $0 \leq u < 1$. Por lo tanto,

$$\begin{aligned} F_X(x) &= P[X \leq x] \\ &= P\left[F^{-}(U) \leq x\right] \\ &= P\left[U \leq F(x)\right] \\ &= F_U(F(x)) \\ &= F(x) \end{aligned}$$

De donde se concluye que la variable aleatoria $X = F^{-}(U)$ tiene como función de distribución acumulativa $F(\cdot)$.

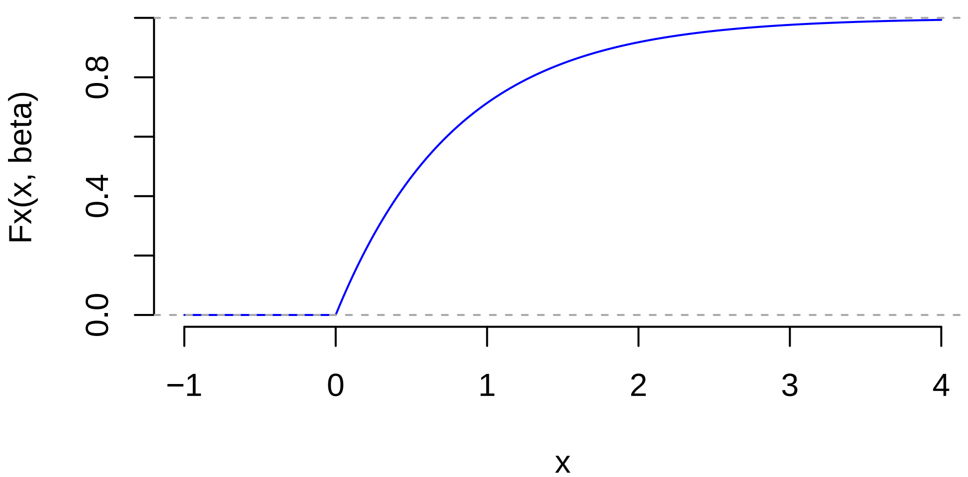
□

Ejercicio 0.51. Sea $F(x;\beta) = 1 - \exp\left(-\frac{x}{\beta}\right)$ para $x > 0$.

$F(\cdot)$ cumple todas las condiciones de una función de distribución acumulativa.

Gráfico de la función $F(x;\beta)$:

```r
# Parámetro beta:
beta <- 0.8
# Función F(x; beta):
Fx <- function(x, beta){
  res <- 1 - exp(-x/beta)
  ifelse(x > 0, res, 0)
}
# Grafica de la función F(x: beta):
curve(Fx(x, beta), from=-1, to=4, n=1e4, bty="n", col="blue")
# Grafica líneas horizontales y = 0 y y = 1:
abline(h=c(0,1), lty=2, col="darkgrey")
```



Para $0 < u < 1$,

$$\begin{aligned} F(x) &= u \\ 1 - \exp\left(-\frac{x}{\beta}\right) &= u \\ \exp\left(-\frac{x}{\beta}\right) &= 1 - u \\ -\frac{x}{\beta} &= \log\left(1 - u\right) \\ x &= -\beta \log\left(1 - u\right) \\ x &= F^{-1}(u) \end{aligned}$$

⚠ ¿Cuál sería la distribución de la variable aleatoria

$$\begin{aligned} X &= F^{-1}(U;\beta) \\ &= -\beta \log\left(1-U\right), \end{aligned}$$

para $U \sim \mathcal{U}(0,1)$?

El resultado del Teorema 0.6 se puede utilizar para generar valores pseudoaleatorios de una distribución para la cual podamos obtener $x = F^{-}(u)$.

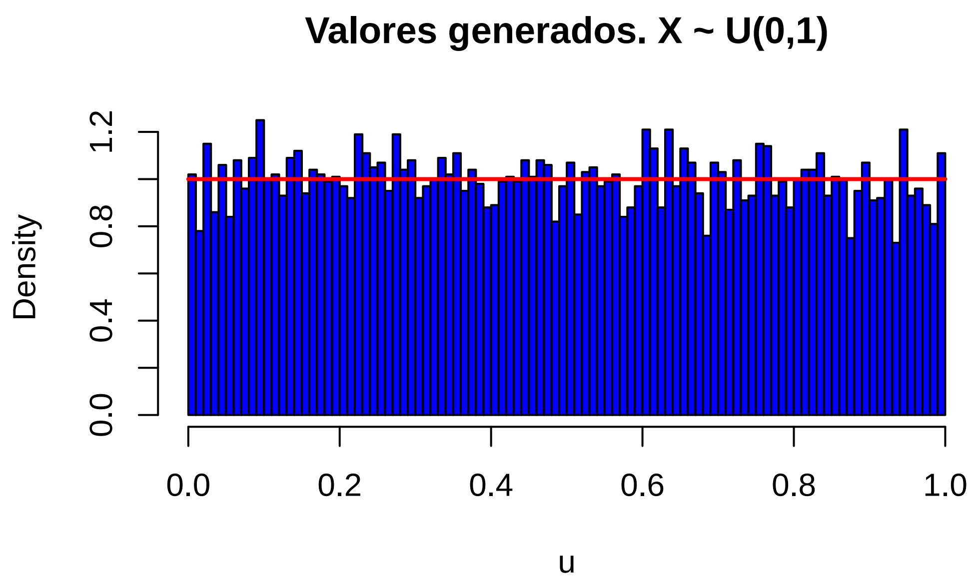
1. Genere un número pseudoaleatorio u a partir de la distribución uniforme continua en $[0,1]$.
2. Calcule $x = F^{-}(u)$.

x será un valor pseudoaleatorio proveniente de una variable aleatoria con función de distribución acumulativa $F(\cdot)$.

Ejemplo 0.3. ¿Qué pasa si generamos una cierta cantidad de valores pseudoaleatorios de una distribución uniforme en $(0,1)$ y luego les aplicamos la función $F^{-1}(U;\beta) = -\beta \log\left(1-u\right)$?

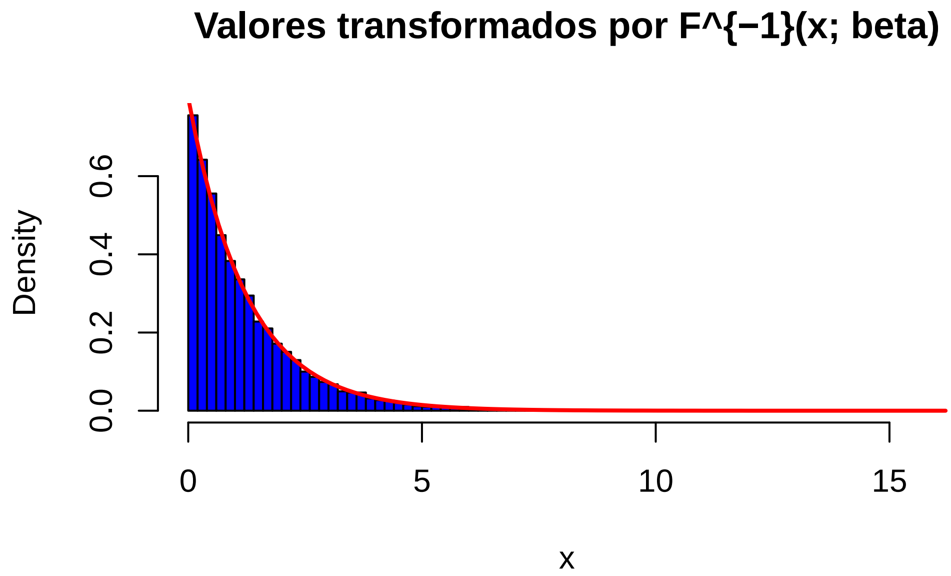
Generación de valores pseudoaleatorios de una distribución uniforme:

```r
# Establecer semilla
set.seed(2023)
# Número de valores a generar:
n <- 1e4
# Generar pseudoaleatorios para una distribución uniforme:
u <- runif(n)
# Histograma de los valores generados:
hist(u, freq=FALSE, col="blue", breaks = sqrt(n),
     main="Valores generados. X ~ U(0,1)")
# Grafica función de densidad uniforme (0,1)
lines(c(0,1), c(1,1), col="red", lwd=2)
```



Aplicamos la función $F^{-1}(U;\beta) = -\beta \log(1-u)$:

```
# Parámetro beta:
beta <- 1.25
# Función F^{-1}(x; beta):
FInvu <- function(u, beta){
  res <- - beta * log(1 - u)
  ifelse(u > 0 | u < 1, res, 0)
}
# Transformar los valores mediante la función F^{-1}(x; beta):
x <- FInvu(u, beta)
# Histograma de los valores transformados por la función:
hist(x, freq=FALSE, col="blue", breaks = sqrt(n),
     main="Valores transformados por F^{-1}(x; beta)")
# Grafica de la función f(x; beta) = exp(-x / beta) / beta
curve(exp(-x / beta) / beta, add=TRUE, col="red", lwd=2)
```



> ⚠ ¿Cuál es la distribución de la variable aleatoria correspondiente a los valores que se obtuvieron al final?

Transformación inversa de probabilidad

Teorema 0.7 (*Probability integral transform*)**.** *Sea X una variable aleatoria con función de distribución acumulativa continua y creciente $F_X(\cdot)$, entonces la variable aleatoria $U = F_X(X)$ tiene distribución uniforme continua sobre el intervalo $[0, 1]$.*

Prueba. Sea $Y = F_X(X)$, entonces, para $0 \leq y < 1$,

$$
\begin{aligned}
F_Y(y) &= P[Y \leq y] \\
&= P[F_X(X) \leq y] \\
&= P[X \leq F_X^{-1}(y)] \\
&= F_X\left(F_X^{-1}(y)\right) \\
&= y
\end{aligned}
$$

Por lo tanto, $Y \sim \mathcal{U}(0, 1)$.

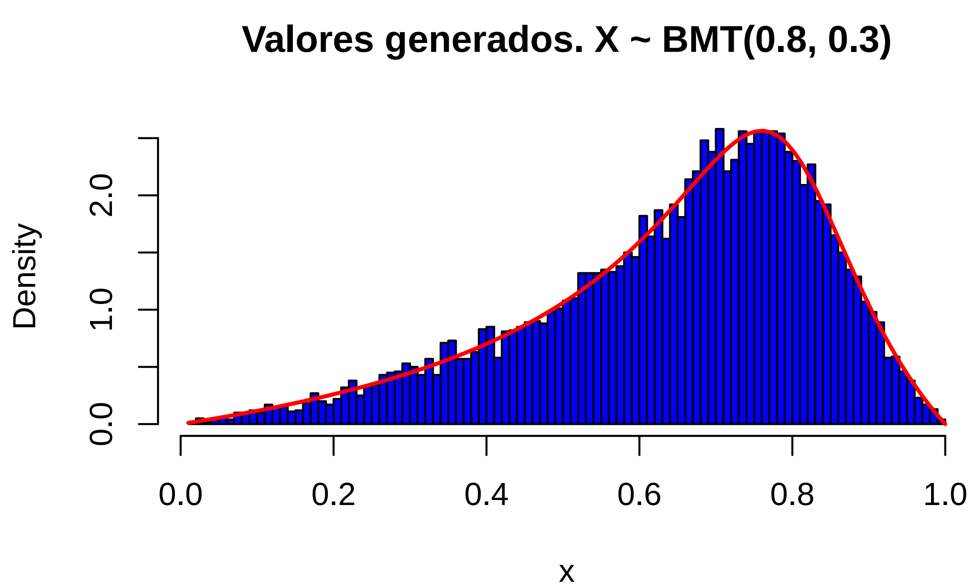
□

> ⚠ Supongamos que tengo unos valores observados provenientes de una variable aleatoria con cierta distribución, ¿qué transformación sería la adecuada para llevarlos a que sean valores observados de una variable aleatorio con cierta distribución de mi preferencia?

Ejemplo 0.4. Supongamos que tengo unos valores observados provenientes de una variable aleatoria con distribución BMT y supongamos que quiero transformarlos adecuadamente para que lleguen a una distribución normal estándar.

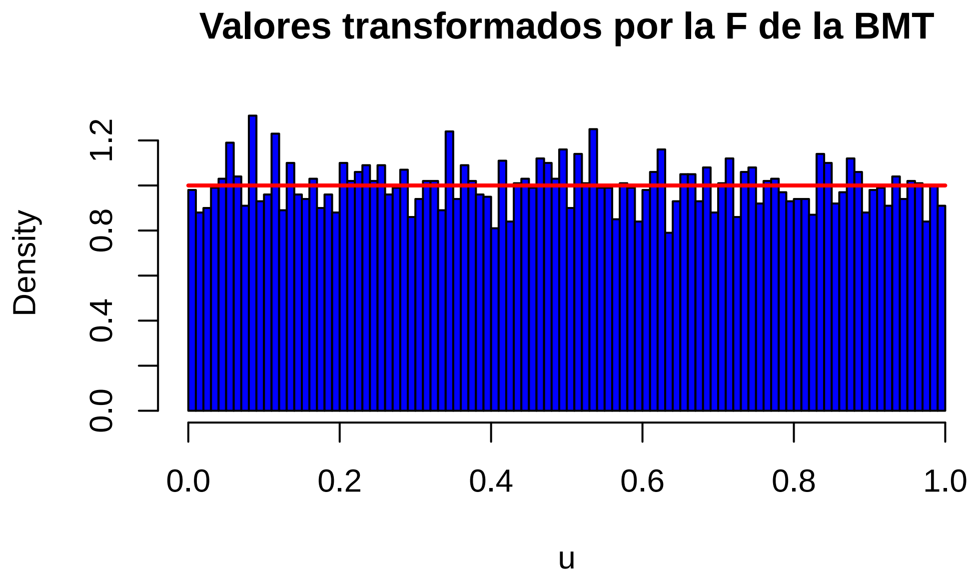
Supongamos que tenemos valores de una distribución BMT:

```r
# Instalar y cargar la libreria de la distribución BMT
if(!require(BMT)) { install.packages("BMT"); require(BMT) }
library(BMT)
# Número de valores a generar:
n <- 1e4
# Valores para los parámetros de la BMT
kappa_l <- 0.8; kappa_r <- 0.3
# Generar pseudoaleatorios para una distribución uniforme:
x <- rBMT(n, kappa_l, kappa_r)
# Histograma de los valores generados:
hist(x, freq=FALSE, col="blue", breaks = sqrt(n),
     main=paste0("Valores generados. X ~ BMT(",
                 kappa_l, ", ", kappa_r, ")"))
# Grafica función de densidad BMT
curve(dBMT(x, kappa_l, kappa_r), add = TRUE, col = "red", lwd=2)
```



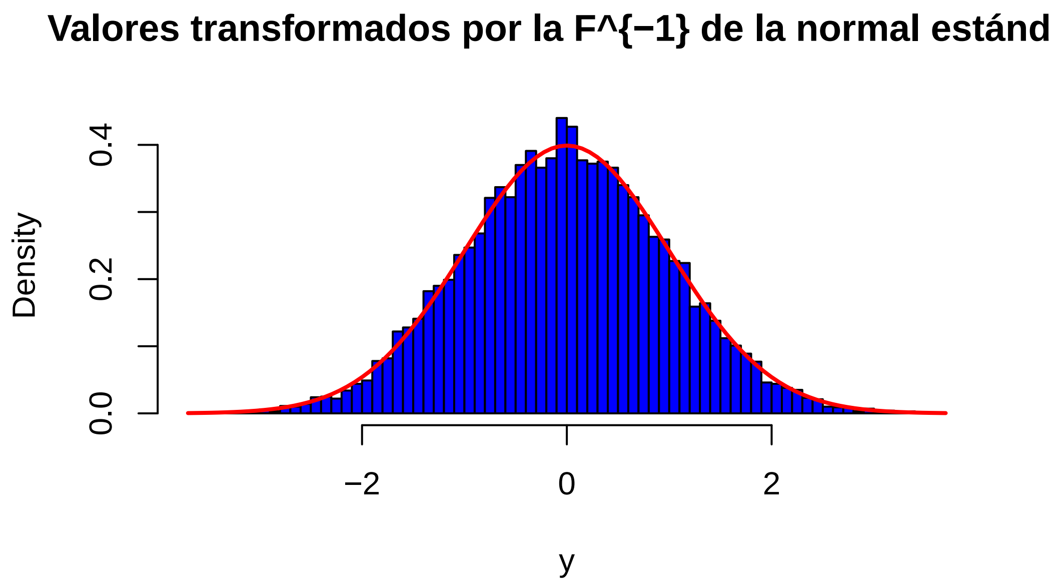
Valores transformados por la función de distribución acumulativa BMT:

```
# Transformar los valores mediante la función F de la BMT:
u <- pBMT(x, kappa_l, kappa_r)
# Histograma de los valores transformados por la F de la BMT:
hist(u, freq=FALSE, col="blue", breaks = sqrt(n),
     main="Valores transformados por la F de la BMT")
# Grafica función de densidad uniforme (0,1)
lines(c(0,1), c(1,1), col="red", lwd=2)
```



Valores transformados por la función cuantil de la normal estándar:

```
# Transformar los valores mediante la función F^{-1} de la normal estándar:
y <- qnorm(u)
# Histograma de los valores transformados por la F de la BMT:
hist(y, freq=FALSE, col="blue", breaks = sqrt(n),
     main="Valores transformados por la F^{-1} de la normal estándar")
# Grafica función de densidad normal estándar
curve(dnorm(x), add = TRUE, col = "red", lwd=2)
```



Partimos de unos datos distribuidos BMT y al transformarlos adecuadamente llegamos a datos con distribución normal estándar.

En el anterior ejemplo (con datos), solamente fue necesario el poder computar la $F(\cdot)$ de la BMT y la $F^{-1}(\cdot)$ de la normal; no fue necesario que dichas funciones tuviesen expresiones analíticas cerradas. Si tenemos expresiones analíticas cerradas para las funciones de distribución acumulativas involucradas y para sus inversas entonces tendremos una expresión analítica cerrada para transformar la respectiva variable aleatoria inicial, a la variable con distribución deseada, tanto de ida como de regreso.

EN CONSTRUCCION.

Convergencia de v. a.

EN CONSTRUCCION.

Consultar sección 1.3 del libro de Inferencia Estadística del profesor Humberto Mayorga: https://repositorio.unal.edu.co/handle/unal/53475

Parte II

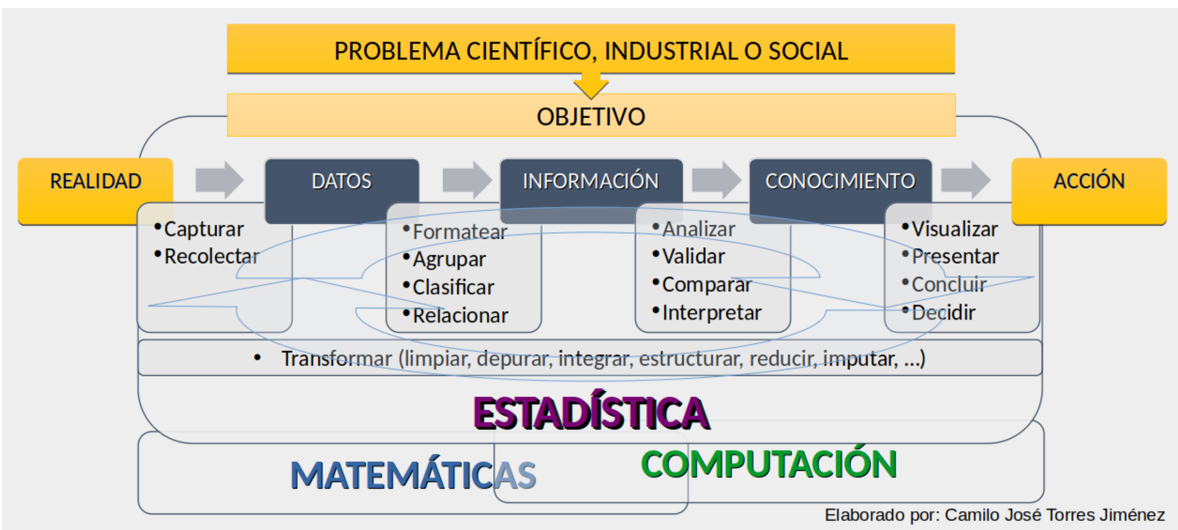
Preliminares

1 Principios de IE

En esta sección se hará una revisión de algunos temas relacionados con los principios de inferencia estadística.

1.1 Acerca de la estadística

1.1.1 El quehacer de la Estadística



Estadística: Recolección, presentación, análisis y uso de datos para la toma de decisiones y resolución de problemas.

> ⚠ La estadística gira alrededor de la incertidumbre (describir, entender, medir, controlar, establecer, explicar). Si no hay incertidumbre (aleatoriedad) entonces no hay estadística.

⚠ La estadística está al servicio de todas las áreas de conocimiento, en su búsqueda de entender la realidad por medio del análisis de datos y en donde la incertidumbre y variabilidad hacen parte de la naturaleza de dicha realidad/datos.

1.1.2 Definiciones

Variable: Una característica, propiedad o atributo observable que medimos.

Por ejemplo: Estatura, Edad, Ingresos, Transacciones.

Constructo o variable latente: Variable no observable. Concepto teórico que no se puede medir directamente.

Por ejemplo: Inteligencia, Felicidad, Inflación, PIB.

Individuo o unidad estadística: Los objetos o entidades a los que les pertenecen las variables. Objetos o entidades a los que se les mide directa o indirectamente las características, propiedades o atributos de interés.

Por ejemplo: Persona, Animal, Empresa, País, Año.

Población: Conjunto de todos los posibles individuos o unidades estadísticas de interés a los cuales se les podría obtener sus características, propiedades o atributos (sus valores para la(s) variable(s)).

Por ejemplo: Estudiantes de una asignatura para el presente semestre académico, Toros de lidia llevados a la Santamaría en las últimas dos décadas, Empresas textiles que tuvieron exportaciones en los últimos cinco años, Años que tuvieron vigente el programa "Bogotá Despierta" el día del amor y la amistad.

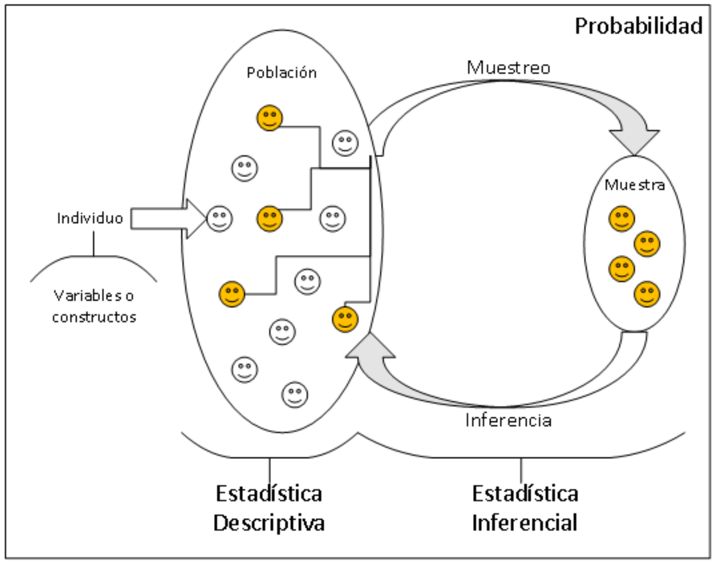
Muestra: Subconjunto de individuos o unidades estadísticas de la población.

1.1.3 Relación entre estadística descriptiva, probabilidad, muestreo e inferencia

A partir del enunciado del siguiente ejercicio, identifique y reflexione acerca de los que serían los individuos, las variables, la población y la muestra, en cada caso.

Ejercicio 1.1.

Defina las poblaciones adecuadas a partir de las cuales se seleccionaron las siguientes muestras:



a. Se llamó por teléfono a personas de 200 casas en la ciudad de Richmond y se les pidió nombrar al candidato por el que votarían en la elección del presidente de la mesa directiva de la escuela.

b. Se lanzó 100 veces una moneda y se registraron 34 cruces

c. Se probaron 200 pares de un nuevo tipo de calzado deportivo en un torneo de tenis profesional para determinar su duración y se encontró que, en promedio, duraron 4 meses.

d. En cinco ocasiones diferentes a una abogada le tomó 21, 26, 24, 22 y 21 minutos conducir desde su casa en los suburbios hasta su oficina en el centro de la ciudad.

Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación. Ejercicio 8.1.

1.2 Acerca del muestreo

- ¿Qué es una muestra?
- ¿Por qué se necesita una muestra?

1.2.1 Muestra aleatoria

Definición 1.1 (Muestra aleatoria)**.** Si $X_1, X_2, \dots, X_n$ son n variables aleatorias *independientes* e *igualmente distribuidas* con distribución $F_X(x)$, entonces $X_1, X_2, \dots, X_n$ es una **muestra aleatoria**, de tamaño n, de la población con distribución $F_X(x)$.

> ⚠ ¿Cómo puedo garantizar que, la manera en que selecciono los individuos, hace que la muestra sea en verdad aleatoria?

1.2.2 Muestreo aleatorio simple

Ejercicio 1.2.

> *Fortune* publica datos sobre ventas, valor del activo, valor de mercado y utilidades por acción de las 500 corporaciones industriales más grandes de Estados Unidos (Fortune 500, 2006). Suponga que usted desea seleccionar una **muestra aleatoria simple** de 10 corporaciones de la lista Fortune 500. Identifique los números de las 10 corporaciones que se tomarán para la muestra.
>
> Anderson, Sweeney, Williams & Camm (2016). Estadística para Negocios y Economía. (12a. ed.) Cengage Learning. Capítulo 7. Ejercicio 3.

Muestreo aleatorio simple (población finita): Una muestra aleatoria simple de tamaño n de una población finita de tamaño N es aquella que es seleccionada de tal manera que cada posible muestra (de tamaño n) tiene la misma probabilidad de ser seleccionada.

Muestreo aleatorio simple (población infinita): Cada individuo se selecciona de manera independiente.

Obtención de valores pseudoaleatorios "igualmente probables":

- *En la calculadora*: `RAN#`
- *Simple Random Sampling Applet*
- *Generate Random Numbers*
- *En una hoja de cálculo*: Funciones `RAND()` y `RANDBETWEEN(a,b)`.
- *En el software estadístico R*: Función `sample()`.

> **!**
>
> Piense (especule) acerca de cuál será la diferencia entre "valores aleatorios" y "valores pseudoaleatorios", luego, investigue cuál es la diferencia y contraste con lo que había pensado (especulado) inicialmente.

1.2.3 Otros tipos de muestreo

Muestreo estratificado: En este caso los individuos de la población se pueden dividir en grupos (homogéneos en su interior y heterogéneos entre ellos) y se desea que en la muestra se tenga una representación adecuada de cada uno de estos grupos. Es así que se toma una muestra aleatoria simple de individuos de cada grupo, de tal manera que en la muestra se conserve la proporción de los tamaños de los grupos.

Muestreo por conglomerados: En este caso los individuos de la población están divididos en segmentos que no necesariamente son homogéneos, cada segmento es una buena representación en menor escala del comportamiento de toda la población y por lo tanto no es necesario tomar individuos de todos los segmentos. Es así que se toma una muestra aleatoria simple de los segmentos y luego para cada segmento seleccionado se toma una muestra aleatoria simple de individuos.

Muestreo sistemático: En este caso, los elementos de la población son seleccionados de manera sistemática a través de la población. La idea es tomar aleatoriamente un individuo por cada cierto número de individuos, es decir, tomo aleatoriamente un individuo de los primeros k, luego uno de los siguientes k y así sucesivamente.

1.3 Acerca de la inferencia estadística

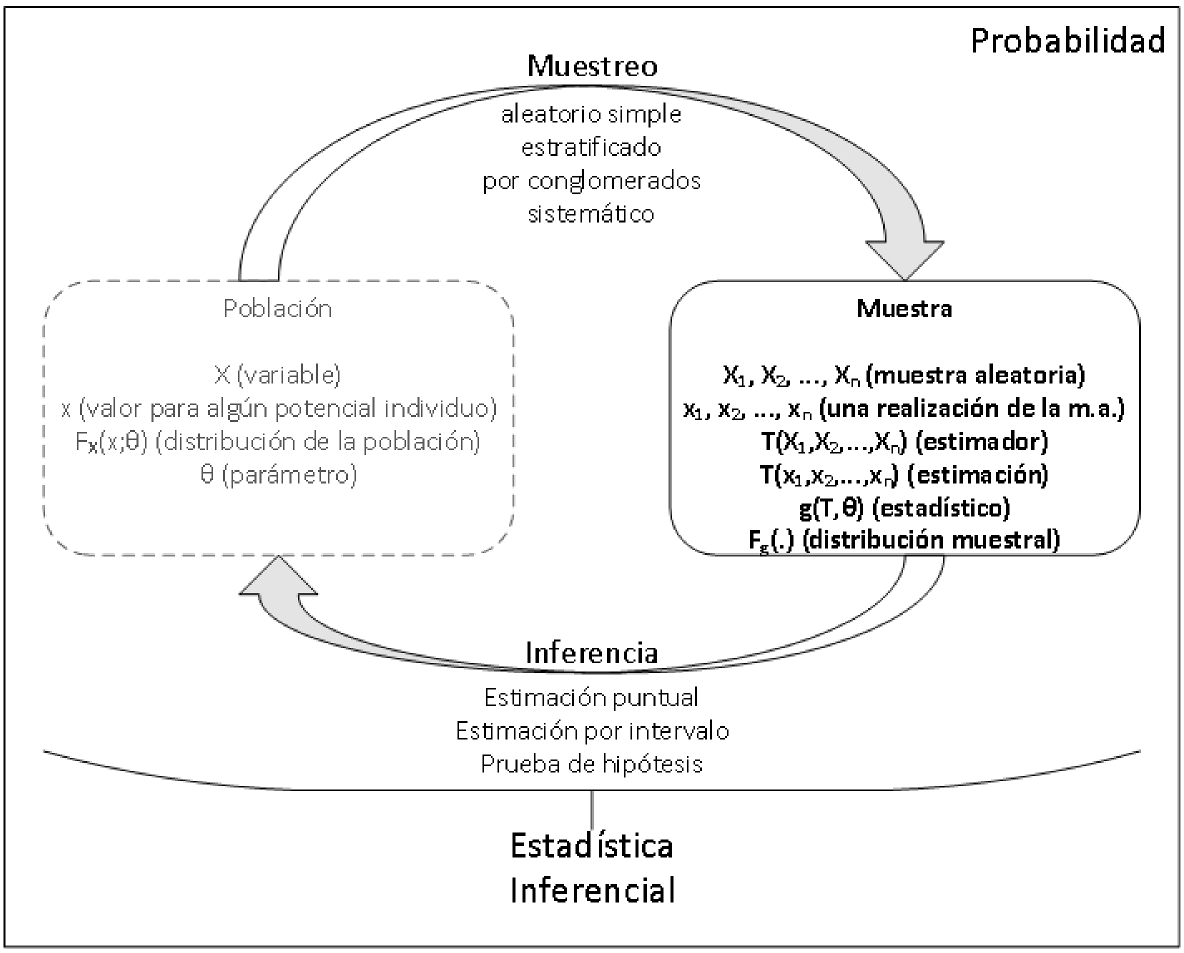
Ejercicio 1.3. Lance una moneda 10 veces consecutivas, registrando los diez resultados obtenidos. Considere que dichos resultados son la realización de una muestra aleatoria de tamaño 10.

- ¿Cuál sería la población en esta situación?
- ¿Cuál sería la variable?
- ¿Cuál sería la distribución de la población?
- ¿Cuál sería el parámetro o los parámetros de dicha distribución?

1.3.1 Ideas iniciales

- **Estadístico**: "Una función de una muestra aleatoria".
- **Estimador**: "Una función de una muestra aleatoria utilizada para estimar un parámetro o una función de un parámetro (no involucra o no tiene parámetros en su expresión)".
- **Estimación**: "Valor que toma un estimador teniendo en cuenta los valores que tomó la muestra aleatoria respectiva (que tomaron los datos observados)".

1.3.2 Relación entre definiciones e ideas iniciales


Probabilidad
Muestreo
aleatorio simple
estratificado
por conglomerados
sistemático
Población
X (variable)
x (valor para algún potencial individuo)
$F_X(x;\theta)$ (distribución de la población)
θ (parámetro)
Muestra
$X_1, X_2, \ldots, X_n$ (muestra aleatoria)
$x_1, x_2, \ldots, x_n$ (una realización de la m.a.)
$T(X_1,X_2,\ldots,X_n)$ (estimador)
$T(x_1,x_2,\ldots,x_n)$ (estimación)
$g(T,\theta)$ (estadístico)
$F_g(.)$ (distribución muestral)
Inferencia
Estimación puntual
Estimación por intervalo
Prueba de hipótesis
Estadística
Inferencial

Parte III

Estimación puntual

2 Obtención de estimadores

En esta sección se hará una revisión de algunos temas relacionados con la estimación puntual de parámetros, específicamente los relacionados con la obtención de estimadores mediante el método de máxima verosimilitud y de otros métodos.

2.1 Preliminares

El objetivo de la **estimación puntual de parámetros** es obtener, a partir de la muestra, un valor que pueda ser una "buena aproximación" de un parámetro (o una función del mismo).

Estimador puntual: Un estimador puntual de un parámetro θ es una función de una muestra aleatoria: $\mathrm{T}\left(X_1, \ldots, X_n\right)$ que es utilizada para estimar el parámetro θ.

Estimación puntual: Una estimación puntual $\hat{\theta}$ de un parámetro θ es el valor que toma un estimador puntual dada una realización de la muestra aleatoria: $\hat{\theta} = \mathrm{T}\left(x_1, \ldots, x_n\right)$.

Definición 2.1. Una **estadística** $T = \mathrm{T}\left(X_1, \ldots, X_n\right)$ cuyos valores son utilizados para estimar una función del parámetro $\mathrm{g}(\theta)$ es un **estimador** de $\mathrm{g}(\theta)$ y las realizaciones del estimador se llaman **estimaciones**.

⚠ ¿Cómo encuentro una función apropiada de la muestra aleatoria para que me sea de utilidad en la estimación de parámetros?

En esta sección se hará una revisión de algunos métodos para obtener estimadores puntuales.

2.2 Método de máxima verosimilitud

La **función de verosimilitud** es la función de densidad conjunta de la muestra aleatoria $X_1, \ldots, X_n$ (de las variables aleatorias $X_1, \ldots, X_n$).

Sea $X_1 \dots, X_n$ una muestra aleatoria de una población con función de densidad (o de masa de probabilidad) $\mathrm{f}_X(x;\theta)$, con $\theta \in \Theta$, la **función de verosimilitud de la muestra aleatoria** se denota y corresponde a,

$$\mathrm{f}_{X_1,\dots,X_n}(x_1,\dots,x_n;\theta) = \mathrm{L}\left(\theta; x_1,\dots,x_n\right),$$

donde $x_1,\dots,x_n$ son realizaciones de las variables aleatorias que constituyen la muestra aleatoria, es decir son los valores observados o los datos con que se cuenta, y por ende se asumen conocidos en la función de verosimilitud $\mathrm{L}\left(\theta; x_1,\dots,x_n\right)$.

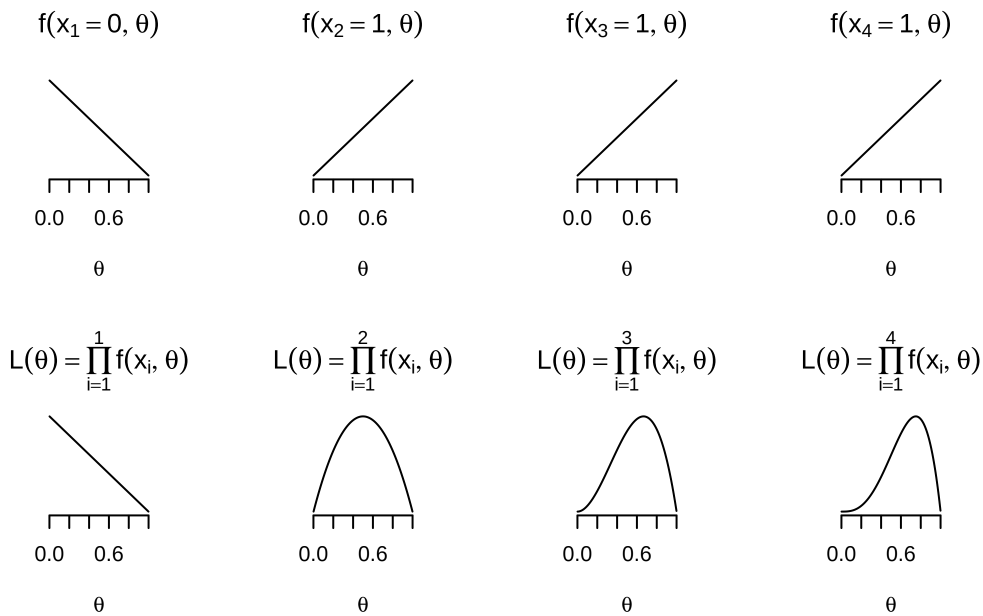
Ejemplo 2.1 (Función de verosimilitud, distribución Bernoulli)**.** Supongamos que nuestra variable aleatoria de interés X toma únicamente dos valores, uno asociado a "acierto" $x = 1$ y otro a "fracaso" $x = 0$. Esta variable tendrá una distribución Bernoulli de parámetro $\theta = p \in (0,1)$, asociado a la probabilidad de "acierto".

En este caso, el objetivo sería encontrar $\hat{\theta}_{ML} = T(x_1,\dots,x_n)$ para una muestra aleatoria $X_1,\dots,X_n$ de una población con $X \sim Bern(p = \theta)$.

En este ejemplo, la función de verosimilitud es $\mathrm{L}\left(\theta; x_1,\dots,x_n\right) = \prod_{i=1}^{n} \mathrm{f}_X(x_i;\theta)$, en donde $\mathrm{f}_X\left(x;\theta\right) = \theta^x(1-\theta)^{1-x}\mathrm{I}_{\{0,1\}}(x)$.

Para poder observar gráficamente la función de verosimilitud, supongamos que tenemos la siguiente realización de la muestra aleatoria: $x_1 = 0, x_2 = 1, x_3 = 1, x_4 = 1$:

```
x <- c(0, 1, 1, 1) # Realización de la muestra aleatoria
n <- length(x) # Tamaño de la muestra
### Gráfica funciones de densidad y funciones de verosimilitud
layout(matrix(c(1:(2*n)), 2, n)) # Malla para 2 x n gráficos
p <- seq(0.001, 0.999, by = 0.001) # valores que toma el parámetro p
l <- rep(1, length(p)) # inicializa la variable para la función de verosimilitud
for(i in 1:n){
    f <- p^(x[i]) * (1-p)^(1-x[i]) # valores que toma la función de densidad
    plot(p, f, type="l", ylim=c(0,1), bty="none", yaxt="n", ylab="",
         xlab=expression(theta), main=bquote(f(x[.(i)]==.(x[i]),theta)))
    l <- l * f # funciones de verosimilitud a medida que incluyo valores x_i
    plot(p, l, type="l", bty="none", yaxt="n", ylab="", xlab=expression(theta),
         main=bquote(L(theta)==prod(f(x[i],theta),i==1,.(i))))
}
```



Se dice que el estimador $\mathrm{T}(X_1 \dots, X_n)$ es el **estimador máximo-verosímil** de θ (***Maximum-Likelihood (ML) Estimator*** *of* θ) si al tomar $\theta = \mathrm{T}(x_1 \dots, x_n) \in \Theta$ se obtiene el supremo de la función de verosimilitud de la muestra aleatoria, es decir,

$$\mathrm{L}\left(\hat{\theta}_{ML}; x_1, \dots, x_n\right) = \sup_{\theta \in \Theta} \mathrm{L}\left(\theta; x_1, \dots, x_n\right),$$

o escrito de otra forma,

$$\hat{\theta}_{ML} = \arg\sup_{\theta \in \Theta} \mathrm{L}\left(\theta; x_1, \dots, x_n\right),$$

donde $\hat{\theta}_{ML} = \mathrm{T}(x_1 \dots, x_n)$ es la **estimación máximo-verosímil** de θ.

Por otra parte se tiene que,

$$\begin{aligned}\hat{\theta}_{ML} &= \arg\sup_{\theta \in \Theta} \mathrm{L}\left(\theta; x_1, \dots, x_n\right) \\ &= \arg\sup_{\theta \in \Theta} \mathrm{g}(\mathrm{L}\left(\theta; x_1, \dots, x_n\right))\end{aligned}$$

para g(.) una función monótona creciente.

En particular,

$$\begin{aligned}\hat{\theta}_{ML} &= \arg\sup_{\theta \in \Theta} \mathrm{L}\left(\theta; x_1, \dots, x_n\right) \\ &= \arg\sup_{\theta \in \Theta} \ \log\left(\mathrm{L}\left(\theta; x_1, \dots, x_n\right)\right)\end{aligned}$$

y teniendo en cuenta que para una muestra aleatoria se tiene que $\mathrm{L}\left(\theta; x_1, \ldots, x_n\right) = \prod_{i=1}^{n} \mathrm{f}_X\left(x_i, \theta\right)$, entonces,

$$\begin{aligned}
\hat{\theta}_{ML} &= \arg\sup_{\theta \in \Theta} \prod_{i=1}^{n} \mathrm{f}_X\left(x_i, \theta\right) \\
&= \arg\sup_{\theta \in \Theta} \sum_{i=1}^{n} \log\left(\mathrm{f}_X\left(x_i, \theta\right)\right).
\end{aligned}$$

¿Qué hago para encontrar el o los valores de θ que maximizan la función de verosimilitud (o su logaritmo)?

2.2.1 Una solución analítica

Bajo ciertas condiciones, los valores de θ que maximizan la función de verosimilitud $\mathrm{L}(\theta; x_1, \ldots, x_n)$ son solución de la ecuación,

$$\frac{\partial}{\partial \theta} \mathrm{L}(\theta; x_1, \ldots, x_n) = 0.$$

Ademas, esos mismos valores de θ maximizan el logaritmo de la función de verosimilitud (función de log-verosimilitud) $\log \mathrm{L}(\theta; x_1, \ldots, x_n)$, y son solución de la ecuación,

$$\frac{\partial}{\partial \theta} \log \mathrm{L}(\theta; x_1, \ldots, x_n) = 0.$$

Manteniendo el mismo objetivo de encontrar $\hat{\theta}_{ML} = \mathrm{T}(x_1 \ldots, x_n)$ cuando $X_1 \ldots, X_n$ es una muestra aleatoria de una población con $X \sim \mathit{Bern}(p = \theta)$ $\left(\mathrm{f}_X\left(x, \theta\right) = \theta^x (1-\theta)^{1-x} \, \mathrm{I}_{\{0,1\}}(x); 0 < \theta < 1\right)$.

La realización de la muestra aleatoria $x_1, \ldots, x_n$ es una secuencia de ceros y unos. La función de verosimilitud es,

$$\begin{aligned}
\mathrm{L}(\theta; x_1, \ldots, x_n) &= \prod_{i=1}^{n} \mathrm{f}_X\left(x_i, \theta\right) \\
&= \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} \, \mathrm{I}_{\{0,1\}}(x_i) \\
&= \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n - \sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \mathrm{I}_{\{0,1\}}(x_i)
\end{aligned}$$

La función de log-verosimilitud es,

$$
\begin{aligned}
\log \mathrm{L}(\theta; x_1, \dots, x_n) &= \sum_{i=1}^{n} \log\left(\mathrm{f}_X\left(x_i, \theta\right)\right) \\
&= \left(\sum_{i=1}^{n} \log\left(\theta^{x_i}(1-\theta)^{1-x_i}\right)\right) \prod_{i=1}^{n} \mathrm{I}_{\{0,1\}}(x_i) \\
&= \left(\sum_{i=1}^{n} \left(x_i \log\left(\theta\right) + (1-x_i)\log(1-\theta)\right)\right) \prod_{i=1}^{n} \mathrm{I}_{\{0,1\}}(x_i) \\
&= \left(\sum_{i=1}^{n} x_i \log\left(\theta\right) + \left(n - \sum_{i=1}^{n} x_i\right) \log(1-\theta)\right) \prod_{i=1}^{n} \mathrm{I}_{\{0,1\}}(x_i)
\end{aligned}
$$

Para $\prod_{i=1}^{n} \mathrm{I}_{\{0,1\}}(x_i) = 1$, con la primera derivada tenemos que,

$$
\begin{aligned}
\frac{\partial}{\partial \theta} \log \mathrm{L}(\theta; x_1, \dots, x_n) &= 0 \\
\frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{n - \sum_{i=1}^{n} x_i}{1-\theta} &= 0 \\
\frac{\sum_{i=1}^{n} x_i}{\theta} &= \frac{n - \sum_{i=1}^{n} x_i}{1-\theta} \\
\frac{1-\theta}{\theta} &= \frac{n - \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i} \\
\frac{1}{\theta} - 1 &= \frac{n}{\sum_{i=1}^{n} x_i} - 1 \\
\theta &= \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}
\end{aligned}
$$

Con la segunda derivada tenemos que,

$$
\frac{\partial^2}{\partial \theta^2} \log \mathrm{L}(\theta; x_1, \dots, x_n) = -\frac{\sum_{i=1}^{n} x_i}{\theta^2} - \frac{n - \sum_{i=1}^{n} x_i}{(1-\theta)^2} < 0
$$

con lo cual se garantiza que $\bar{x}$ es un máximo del logaritmo de la función de log-verosimilitud (y también de la función de verosimilitud).

Concluimos que,

$$
\hat{\theta}_{ML} = \bar{x}.
$$

Teniendo en cuenta que la realización de la muestra aleatoria $x_1, \dots, x_n$ es una secuencia de ceros y unos, entonces $\sum_{i=1}^{n} x_i$ es la cantidad de unos en la muestra y $\bar{x}$ sería la proporción de unos. Es decir, para el caso de la distribución Bernoulli, el estimador máximo-verosímil de θ es la media muestral $\bar{X}$, que coincide con la proporción muestral $\hat{P}$. Para $X \sim Bern(p=\theta)$,

$$\hat{\theta}_{ML} = \bar{x} = \hat{p}.$$

Finalmente, es bueno recordar que $X_1, \dots, X_n$ es una muestra aleatoria únicamente si las observaciones o realizaciones de dichas variables aleatorias se obtuvieron mediante un adecuado muestreo (lo cual garantiza lo independiente identicamente distribuido de la muestra aleatoria).

Ejercicio 2.1 (Distribución binomial)**.** Sabemos que $\sum_{i=1}^{n} X_i \sim Bin(n, p=\theta)$. Por lo tanto, dada una muestra aleatoria Y_1 de tamaño 1 de una población con $Y \sim Bin(n, p=\theta)$, entonces, suponiendo que n ("número de ensayos") es conocido, verifique que el estimador máximo verosímil del parámetro θ ("probabilidad de éxito") es $\frac{y_1}{n}$.

Suponiendo que n es conocido, ¿cuál sería el estimador de máximo verosimilitud del parámetro θ cuando $Y_1, \dots, Y_m$ es una muestra aleatoria de tamaño m de una población con $Y \sim Bin(n, p=\theta)$ $\left(\mathrm{f}_Y(y,\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}\,\mathrm{I}_{\{0,1,\dots,n\}}(y)\right)$?

2.2.2 Otro tipo de solución analítica

Una variable aleatoria X con distribución hipergeométrica $X \sim Hg(n, N, R)$ se puede visualizar o pensar como el número de individuos que tienen cierta característica de interés ("aciertos"), de n que fueron seleccionados simultáneamente o uno a uno sin reemplazamiento. Esta selección se realiza sobre una población finita de tamaño N, en donde R de los N individuos tienen la característica de interés (y por ende, $N-R$ no la tienen). La función de densidad de X estaría dada por

$$\mathrm{f}_X(x; n, N, R) = \frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}}$$

para $x \in \mathbb{Z}$ tal que $\max\{0, n-(N-R)\} \leq x \leq \min\{n, R\}$ (o lo que es lo mismo, para $x \in \mathbb{Z}$ tal que $\mathrm{I}_{\{\max\{0,n-(N-R)\},\dots,\min\{n,R\}\}}(x) = 1$) y donde $n, N, R \in \mathbb{Z}^+$.

El que cada individuo seleccionado tenga o no la característica de interés se asocia a una variable aleatoria con distribución Bernoulli, pero en este caso la selección de los n individuos no cumple el criterio de independencia que se exige en el caso de una distribución Binomial. Si $X_1, \dots, X_n$ son variables aleatorias idénticamente distribuidas con $X_i \sim Bern\left(p = \frac{R}{N}\right)$ y sus realizaciones se tomaron simultáneamente, o uno a uno sin reemplazamiento, de una población finita de tamaño N, entonces $\sum_{i=1}^{m} X_i = Y \sim Hg(n, N, R)$. De lo anterior, estimar alguno de los parámetros n, N o R a partir de las variables $X_1, \dots, X_n$ mencionadas es equivalente a estimar dichos parámetros a partir de una muestra aleatoria Y_1 de tamaño 1 de una población con $Y \sim Hg(n, N, R)$.

2.2.2.1 Estimación del parámetro N

Sea Y_1 (o los $X_1, \ldots, X_n$ equivalentes descritos anteriormente) y supongamos que n y R son conocidos, entonces el único parámetro a estimar sería $\theta = N$ $\left(\mathrm{f}_Y(y, \theta) = \frac{\binom{R}{y}\binom{\theta-R}{n-y}}{\binom{\theta}{n}} \mathrm{I}_{\{\max\{0, n-(\theta-R)\}, \ldots, \min\{n, R\}\}}(y) \right)$. Es decir que en este caso tendríamos conocimiento de que un cierto número de individuos tienen una característica de interés (y_1) dentro una muestra de cierto tamaño (n) (o sabemos que cierto número de individuos tiene la característica (y_1) y cierto número de individuos no la tienen $n - y_1$); ademas, sabemos que en toda la población hay un cierto número de individuos que tienen la característica (R). A partir de esto que conocemos, queremos inferir la cantidad total de individuos de la población ($\theta = N$), la cuál debe ser desconocida.

Ilustremos un poco más lo anterior a partir de la siguiente situación. Supongamos que queremos conocer la cantidad de peces de cierta especie en un lago. Para inferir dicha cantidad, introducimos al lago un cierto número de peces (R) de la misma especie pero marcados de alguna manera especial. Al día siguiente, capturamos simultáneamente (pesca con red) o sin reemplazamiento (uno a uno pero sin devolver los peces capturados) y contamos el número de peces con la marca especial (y_1) del total de los que capturamos (n) (o, el número de capturados con la marca especial (y_1) y el número de capturados sin la marca ($n - y_1$)). Ahora con la información que poseemos se debería poder estimar la cantidad total de peces que había en el lago en el instante previo a proceder con la captura ($\hat{\theta} = N$), y por ende también la cantidad de peces que había en el lago antes de introducir los marcados $\hat{\theta} - R = N - R$. Es así que se requiere establecer una fórmula o mecanismo que permita obtener estimaciones máximo verosímiles para $\theta = N$, a partir de lo observado mediante y_1, y bajo el supuesto de que R y n son conocidos.

Primero, determinemos cuáles podrían ser valores válidos para $\theta = N$:

A partir del valor que podría tomar y_1, es posible establecer los valores válidos que podría tomar $\theta = N$:

$$
\begin{aligned}
\max\{0, n - (\theta - R)\} \leq y_1 &\leq \min\{n, R\} \\
n - (\theta - R) &\leq y_1 \\
-(\theta - R) &\leq y_1 - n \\
\theta - R &\geq n - y_1 \\
\theta &\geq R + (n - y_1)
\end{aligned}
$$

Es decir $\theta = N$ tiene que ser un número entero positivo, mayor o igual que la suma del número de individuos con la característica dentro de toda la población (R) y el número de individuos sin la característica dentro de los n seleccionados ($n - y_1$).

Segundo, determinar cuándo crece, cuando permance constante y cuando decrece la función de verosimilitud:

La función de verosimilitud para el parámetro $\theta = N$ no es una función continua, por tanto no es posible encontrar un máximo usando derivadas. Una alternativa es analizar el comportamiento de la función de verosimilitud, es decir, establecer cuándo crece, cuándo decrece o incluso cuándo permanece constante. Una forma de analizar dicho comportamiento es comparando el cambio que ocurre en los valores que toma la función de verosimilitud para un valor posible del parámetro (θ^*) y su valor inmediatamente anterior (θ^*). Es así que comparamos $L(\theta^*)$ con respecto a $L(\theta^* - 1)$ y establecemos que, **al pasar de** $(\theta^* - 1)$ **a** θ^*,

- $L(.)$ crece si y sólo si $L(\theta^* - 1) < L(\theta^*)$, o equivalentemente si $1 < \frac{L(\theta^*)}{L(\theta^*-1)}$;
- $L(.)$ permanece constante si y sólo si $L(\theta^* - 1) = L(\theta^*)$, o equivalentemente si $1 = \frac{L(\theta^*)}{L(\theta^*-1)}$;
- $L(.)$ decrece si y sólo si $L(\theta^* - 1) > L(\theta^*)$, o equivalentemente si $1 > \frac{L(\theta^*)}{L(\theta^*-1)}$;

donde

$$\begin{aligned} L(\theta) &= L(\theta; x_1, \dots, x_n) \\ &= L(\theta; y_1) \\ &= \frac{\binom{R}{y_1}\binom{\theta-R}{n-y_1}}{\binom{\theta}{n}} \mathrm{I}_{\{\max\{0,n-(\theta-R)\},\dots,\min\{n,R\}\}}(y_1). \end{aligned}$$

Dada y_1, la realización de la variable aleatoria Y_1 (o $x_1, \dots, x_n$ una realización de $X_1, \dots, X_n$ Bernoulli no independientes), y para θ^* tal que $L(\theta^* - 1) \neq 0$, es decir para $\theta^* - 1 \geq R + (n - y_1)$, tenemos que,

$$\begin{aligned} \frac{L(\theta^*)}{L(\theta^* - 1)} &= \frac{\binom{R}{y_1}\binom{\theta^*-R}{n-y_1}}{\binom{\theta^*}{n}} \frac{\binom{\theta^*-1}{n}}{\binom{R}{y_1}\binom{\theta^*-R-1}{n-y_1}} \\ &= \frac{\binom{\theta^*-R}{n-y_1}}{\binom{\theta^*}{n}} \frac{\binom{\theta^*-1}{n}}{\binom{\theta'-R-1}{n-y_1}} \\ &= \frac{\frac{(\theta^*-R)!}{(n-y_1)!(\theta^*-R-(n-y_1))!}}{\frac{\theta^*!}{n!(\theta^*-n)!}} \frac{\frac{(\theta^*-1)!}{n!(\theta^*-n-1)!}}{\frac{(\theta^*-R-1)!}{(n-y_1)!(\theta^*-R-(n-y_1)-1)!}} \\ &= \frac{(\theta^* - R)!(\theta^* - n)!(\theta^* - 1)!(\theta^* - R - (n - y_1) - 1)!}{(\theta^* - R - (n - y_1))!\theta^*!(\theta^* - n - 1)!(\theta^* - R - 1)!} \\ &= \frac{(\theta^* - R)(\theta^* - n)}{(\theta^* - R - (n - y_1))\theta^*} \end{aligned}$$

Lo que quiere decir que $\frac{L(\theta^*)}{L(\theta^*-1)} > 1$ (la función de verosimilitud crece) cuando,

$$\frac{(\theta^* - R)(\theta^* - n)}{(\theta^* - R - (n - y_1))\theta^*} > 1.$$

Además, como $\theta^* \in \mathbb{Z}^+$ y $\theta^* - R - (n - y_1) \geq 1$ entonces $\frac{L(\theta^*)}{L(\theta^*-1)} > 1$ cuando,

$$\begin{aligned}
(\theta^* - R)(\theta^* - n) &> (\theta^* - R - (n - y_1))\theta^* \\
\theta^{*2} - R\,\theta^* - n\,\theta + R\,n &> \theta^{*2} - R\,\theta^* - n\,\theta^* + y_1\,\theta^* \\
R\,n &> y_1\,\theta^* \\
\frac{R\,n}{y_1} &> \theta^*
\end{aligned}$$

Concluimos que, al pasar de $(\theta^* - 1)$ a θ^*,

- $L(.)$ crece $(L(\theta^*) > L(\theta^* - 1))$ si y sólo si $\theta^* < \frac{R\,n}{y_1}$;
- $L(.)$ permanece igual $(L(\theta^*) = L(\theta^* - 1))$ si y sólo si $\theta^* = \frac{R\,n}{y_1}$.
- $L(.)$ decrece $(L(\theta^*) < L(\theta^* - 1))$ si y sólo si $\theta^* > \frac{R\,n}{y_1}$;

Entonces, $\frac{R\,n}{y_1}$ será un valor de referencia para saber si la función de verosimilitud crece, decrece, o permanece igual.

Tercero, sabiendo cuando crece, permanece igual o decrece la función de verosimilitud, identificar el valor del parámetro que hace que dicha función alcance su valor máximo:

Sin embargo, tenemos dos situaciones distintas, dependiendo de si $\frac{R\,n}{y_1}$ es o no un número entero

Si $\frac{R\,n}{y_1}$ NO ES entero, entonces,

$$\left[\frac{R\,n}{y_1}\right] < \frac{R\,n}{y_1} < \left[\frac{R\,n}{y_1}\right] + 1,$$

y como $L(.)$ crece al pasar de $(\theta * -1)$ a $\theta*$ para todo $\theta*$ entero menor que $\frac{R\,n}{y_1}$, entonces,

$$\begin{aligned}
L\left(\left[\frac{R\,n}{y_1}\right]\right) &> L\left(\left[\frac{R\,n}{y_1}\right] - 1\right) \\
L\left(\left[\frac{R\,n}{y_1}\right] - 1\right) &> L\left(\left[\frac{R\,n}{y_1}\right] - 2\right) \\
&\dots
\end{aligned}$$

además, como $L(.)$ decrece al pasar de $(\theta * -1)$ a $\theta*$ para todo $\theta*$ entero mayor que $\frac{R\,n}{y_1}$, entonces,

$$\begin{aligned}
L\left(\left[\frac{R\,n}{y_1}\right] + 1\right) &< L\left(\left(\left[\frac{R\,n}{y_1}\right] + 1\right) - 1\right) = L\left(\left[\frac{R\,n}{y_1}\right]\right) \\
L\left(\left[\frac{R\,n}{y_1}\right] + 2\right) &< L\left(\left(\left[\frac{R\,n}{y_1}\right] + 2\right) - 1\right) = L\left(\left[\frac{R\,n}{y_1}\right] + 1\right) \\
&\dots
\end{aligned}$$

De los dos resultados anteriores, concluimos que $L\left(\left[\frac{R\,n}{y_1}\right]\right)$ es el máximo valor posible de la función $L(.)$ y por ende

$$\hat{\theta}_{ML} = \left[\frac{R\,n}{y_1}\right].$$

Ahora, **si** $\frac{R\,n}{y_1}$ **ES entero**, entonces,

$$\left[\frac{R\,n}{y_1}\right] = \frac{R\,n}{y_1}$$

a partir de lo cual se puede ver que:

- $L(.)$ crece para $\theta < \left[\frac{R\,n}{y_1}\right]$, es decir,

$$L\left(\left[\frac{R\,n}{y_1}\right]-1\right) > L\left(\left[\frac{R\,n}{y_1}\right]-2\right) > L\left(\left[\frac{R\,n}{y_1}\right]-3\right) > \ldots,$$

- $L(.)$ permanece igual para $\theta = \left[\frac{R\,n}{y_1}\right]$, es decir,

$$L\left(\left[\frac{R\,n}{y_1}\right]\right) = L\left(\left[\frac{R\,n}{y_1}\right]-1\right),$$

- $L(.)$ decrece para $\theta > \left[\frac{R\,n}{y_1}\right]$, es decir,

$$L\left(\left[\frac{R\,n}{y_1}\right]\right) > L\left(\left[\frac{R\,n}{y_1}\right]+1\right) > L\left(\left[\frac{R\,n}{y_1}\right]+2\right) > \ldots.$$

De donde se concluye que $\left[\frac{R\,n}{y_1}\right]$ y $\left[\frac{R\,n}{y_1}\right]-1$ son los valores del parámetro que hacen que $L(\theta)$ alcance su valor máximo. En este caso, $\left[\frac{Rn}{y_1}\right]$ y $\left[\frac{Rn}{y_1}\right]-1$ serían las estimaciones máximo verosímiles de θ. $\hat{\theta}_{ML}$ no sería único,

$$\hat{\theta}_{ML} = \left\{\left[\frac{Rn}{y_1}\right]-1, \left[\frac{Rn}{y_1}\right]\right\}$$

Cuarto, usamos el estimador que hallamos en todas las aplicaciones prácticas que apliquen

Mediante el anterior desarrollo analítico obtuvimos las expresiones matemáticas para asociadas a las estimaciones máximo verosímiles. Ahora para ilustrar un poco las cosas, primero **simulemos** la selección simultanea (sin reemplazamiento) de $n = 10$ individuos sobre una población de tamaño $N = 30$, en donde $R = 18$. Eso sí, sin olvidar que en la realidad los datos poblacionales son desconocidos (no se podría saber que $N = 30$). Por eso aquí estamos hablando de datos simulados y no de datos reales (a continuación, `ALL` serían los datos poblacionales desconocidos y tanto `x` como `y_1.op2` serían los datos muestrales simulados).

```r
# Datos poblacionales (que en una situación real no conocería)
ALL <- c(1, 1, 0, 1, 1, 0, 1, 1, 0, 1,
         1, 0, 0, 1, 0, 1, 0, 1, 0, 1,
         0, 1, 1, 0, 1, 0, 1, 1, 0, 1)
N <- length(ALL)
# Datos conocidos
R <- sum(ALL)
n <- 10
# Semilla para la generación de valores pseudoaleatorios
set.seed(123)
```

```r
# Simulación opción 1:
# muestra de tamaño n sin reemplazamiento de una variable Bernoulli
x <- sample(ALL, n, replace = FALSE)
y_1.op1 <- sum(x)
cat("Simulación opción 1.\nx_1, ..., x_n: ",
    paste(x, collapse = ", "),
    "\ny_1: ", y_1.op1)
```

```
Simulación opción 1.
x_1, ..., x_n:  0, 0, 1, 0, 1, 1, 1, 1, 1, 1
y_1:  7
```

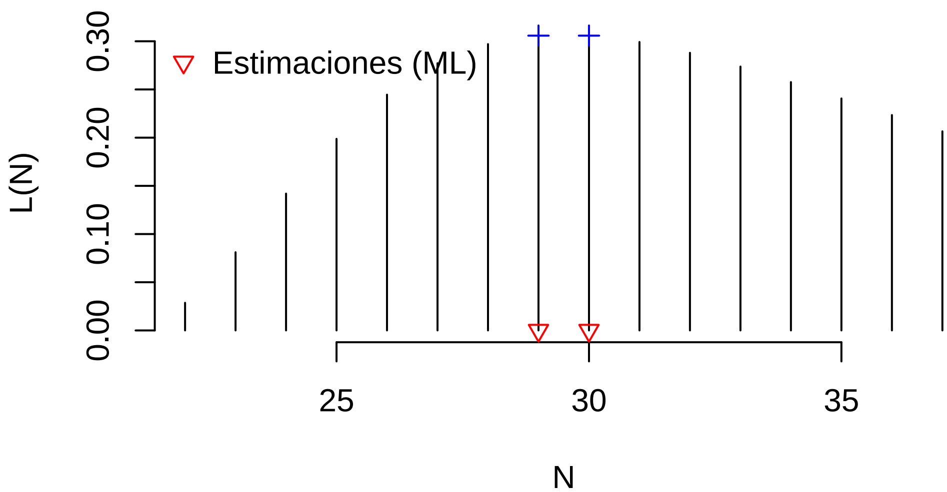
```r
# Simulación opción 2:
# muestra de tamaño 1 de una variable Hipergeométrica
y_1.op2 <- rhyper(1, R, N - R, n)
cat("Simulación opción 2.\ny_1: ", y_1.op2)
```

```
Simulación opción 2.
y_1:  6
```

Luego, veamos la función de verosimilitud asociada a los datos simulados y_1 (es decir, la función de verosimilitud asociada a lo que sería conocido: y_1, n y R), y las estimaciones máximo verosímiles para el parámetro $\theta = N$.

```r
aux.plot <- 0:15
theta <- R + (n - y_1.op2) + aux.plot
L <- dhyper(y_1.op2, R, theta - R, n)
plot(theta, L, type = "h", bty = "n",
```

```
     xlab = "N", ylab = "L(N)", ylim=c(0,max(L)))
val.ref <- R * n / y_1.op2
if(val.ref != trunc(val.ref)){
  theta.hat <- trunc(val.ref)
}else{
  theta.hat <- trunc(val.ref) - c(0,1)
}
points(theta.hat, dhyper(y_1.op2, R, theta.hat - R, n), col = "blue", pch = 3)
points(theta.hat, rep(0, length(theta.hat)), col = "red", pch = 6)
legend("topleft", legend = "Estimaciones (ML)", col = "red", pch = 6, bty = "n")
```



¿Qué pasaría si en vez de Y_1 (o las equivalente $X_1, \ldots, X_n$ variables aleatorias Bernoulli no independientes), se tiene una muestra aleatoria $Y_1, \ldots, Y_m$ de tamaño m, de una población $Y \sim Hg(n, N = \theta, R)$?

Para ilustrar un poco la situación, podemos simular un conjunto de datos asociado a la misma:

```r
m <- 15 # Tamaño de muestra
# R, N y n iguales que antes
y <- rhyper(m, R, N - R, n) # Simular una realización de la muestra aleatoria
cat("y_1, ..., y_m: ", paste(y, collapse = ", "))
```

```
y_1, ..., y_m:  5, 6, 8, 4, 7, 8, 7, 4, 4, 5, 6, 3, 6, 5, 6
```

Calculamos la función de log-verosimilitud (a partir de los datos simulados) para valores que podría tomar el parámetro de interés $\theta = N$.

```r
aux.max.theta <- 35 # No puedo calcular infinitos valores para theta
theta <- seq(R + max(n - y), aux.max.theta) # Valores que podría tomar theta
for(i in 1:length(theta)){
    cat("theta: ", theta[i], "\t", "log-Likelihood: ",
        sum(dhyper(y, R, theta[i] - R, n, log = TRUE)), "\n", sep = "")
}
```

```
theta: 25   log-Likelihood: -43.23136
theta: 26   log-Likelihood: -37.04521
theta: 27   log-Likelihood: -33.08715
theta: 28   log-Likelihood: -30.47942
theta: 29   log-Likelihood: -28.7807
theta: 30   log-Likelihood: -27.72891
theta: 31   log-Likelihood: -27.15431
theta: 32   log-Likelihood: -26.94071
theta: 33   log-Likelihood: -27.00532
theta: 34   log-Likelihood: -27.28739
theta: 35   log-Likelihood: -27.74128
```

Para los datos simulados, la **estimación** máximo verosímil sería,

$$\hat{\theta}_{ML} = \hat{N}_{ML} = 32,$$

Ya que es el valor para el cual la función de log-verosimilitud llega a su máximo (suponiendo que para $\theta = 36, 37, ...$ dicha función seguirá decreciendo).
En este caso, se encontró la estimación para un caso particular, pero **desconocemos la expresión matemática del estimador respectivo**. Aún no tenemos la expresión matemática de la función de la muestra correspondiente al estimador máximo verosímil.

¿Será posible demostrar que

$$\hat{\theta}_{ML} = \max\left\{\theta \in \mathbb{Z}^+ : \theta \geq \left(R + \max_{1\leq i\leq m}\{n - x_i\}\right) \wedge \frac{L(\theta+1)}{L(\theta)} \geq 1\right\}$$

para $Y_1, \dots, Y_m$ una muestra aleatoria de una población $Y \sim Hg(n, N = \theta, R)$?
¿Es posible encontrar de forma analítica una expresión más simple para el estimador máximo verosímil del parámetro $\theta = N$, suponiendo que n y R son conocidos?

2.2.2.2 Estimación del parámetro R

Ejercicio 2.2. Sea Y_1 una variable aleatoria como la descrita anteriormente (o la muestra $X_1, \dots, X_n$ equivalente) y supongamos que n y N son conocidos, entonces el único parámetro a estimar sería $\theta = R$ $\left(\mathrm{f}_Y(y, \theta) = \frac{\binom{\theta}{y}\binom{N-\theta}{n-y}}{\binom{N}{n}} \mathrm{I}_{\{\max\{0, n-(N-\theta)\}, \dots, \min\{n,\theta\}\}}(y)\right)$.

¿Cuál sería la expresión matemática asociada a la estimación máximo verosímil del parámetro $\theta = R$, suponiendo que n y N son conocidos??

¿Qué pasaría si en vez de Y_1 (o las equivalente $X_1, \dots, X_n$ variables aleatorias Bernoulli no independientes), se tiene una muestra aleatoria $Y_1, \dots, Y_m$ de tamaño m, de una población $Y \sim Hg(n, N, R = \theta)$?

Para ilustrar un poco la situación, podemos simular un conjunto de datos asociado a la misma:

```r
m <- 15 # Tamaño de muestra
N <- 20; R <- 8; n <- 10 # Supongamos estos valores para los parámetros
y <- rhyper(m, R, N - R, n) # Simular una realización de la muestra aleatoria
cat("y_1, ..., y_m: ", paste(y, collapse = ", "))
```

```
y_1, ..., y_m:  4, 5, 5, 2, 3, 3, 3, 6, 4, 3, 5, 5, 5, 5, 4
```

Calculamos la función de log-verosimilitud (a partir de los datos simulados) para valores que podría tomar el parámetro de interés $\theta = R$.

```r
theta <- seq(max(y), N - max(n - y)) # Valores que podría tomar theta
for(i in 1:length(theta)){
    cat("theta: ", theta[i], "\t", "log-Likelihood: ",
        sum(dhyper(y, theta[i], N - theta[i], n, log = TRUE)), "\n", sep = "")
}
```

```
theta: 6    log-Likelihood: -31.22295
theta: 7    log-Likelihood: -24.92984
theta: 8    log-Likelihood: -22.59759
theta: 9    log-Likelihood: -23.31624
theta: 10   log-Likelihood: -26.8586
theta: 11   log-Likelihood: -33.4468
theta: 12   log-Likelihood: -43.93297
```

Para los datos simulados, la **estimación** máximo verosímil sería,

$$\hat{\theta}_{ML} = \hat{R}_{ML} = 8,$$

Ya que es el valor para el cual la función de log-verosimilitud llega a su máximo (los posibles valores de θ son finitos, entonces estoy seguro de que se logró llegar a un valor que maximiza dicha función).
En este caso, se encontró la estimación para un caso particular, pero **desconocemos la expresión matemática del estimador respectivo**. Aún no tenemos la expresión matemática de la función de la muestra correspondiente al estimador máximo verosímil.

⚠

¿Para este caso (muestra aleatoria $Y_1, \ldots, Y_m$ de tamaño m, de una población $Y \sim Hg(n, N, R = \theta)$), es posible encontrar de forma analítica una expresión relativamente simple para el estimador máximo verosímil del parámetro $\theta = R$, suponiendo que n y N son conocidos?

2.2.2.3 Estimación del parámetro n

De la misma manera en que lo hicimos con los parámetros R y N, podríamos tomar Y_1 (o la muestra $X_1, \ldots, X_n$ equivalente) y suponer que N y R son conocidos, entonces el único parámetro a estimar sería $n = \theta \left(\mathrm{f}_Y(y, \theta) = \frac{\binom{R}{y}\binom{N-R}{\theta-y}}{\binom{N}{\theta}} \mathrm{I}_{\{\max\{0, \theta-(N-R)\}, \ldots, \min\{\theta, R\}\}}(y) \right)$. Matemáticamente hablando, no veríamos problema alguno en tratar de establecer una fórmula o mecanismo que nos permita llegar a estimaciones máximo verosímiles para $n = \theta$. Sin embargo, antes de ponernos a trabajar en ello deberíamos preguntarnos:

> ⚠
>
> ¿Es posible encontrar o proponer una situación real o hipotética en donde se requiera inferir n (el cual naturalmente tiene que ser desconocido) a partir de lo observado mediante y_1, y bajo el supuesto de que N y R son conocidos?

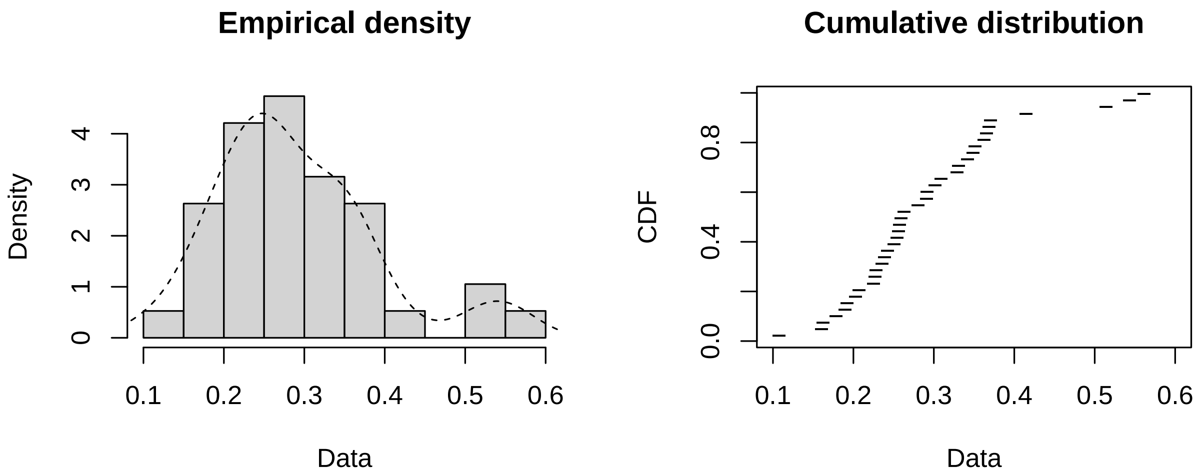
2.2.3 Una solución numérica

Supongamos que queremos estimar los parámetros de la distribución poblacional asociada a la variable ***porcentaje de gasto en alimentación de los hogares***, a partir de los datos a los que se refiere el siguiente texto:

> The data frame contains 38 observations (random sample of 38 households in a large US city) on 3 variables:
>
> - food: household expenditures for food.
> - income: household income.
> - persons: number of persons living in household.
>
> Griffiths, W.E., Hill, R.C., and Judge, G.G. 1993. Learning and Practicing Econometrics New York: John Wiley and Sons (Table 15.4).

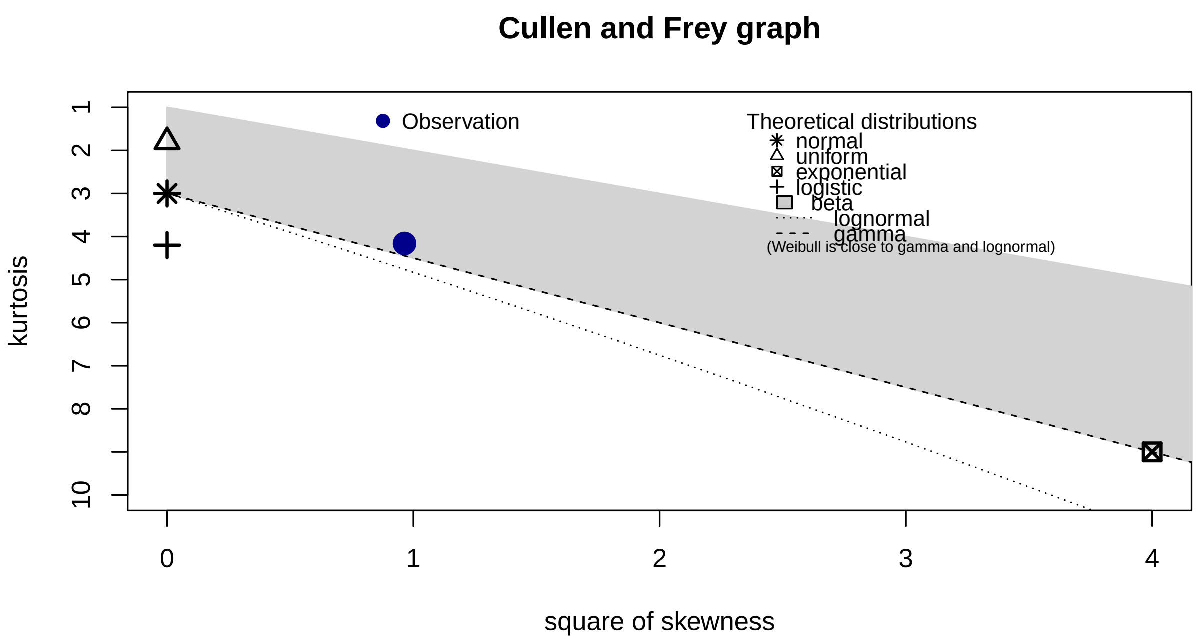
Lo primero que deberíamos hacer es un análisis exploratorio de los datos de la muestra.

```r
# Instalar y cargar la libreria betareg (datos FoodExpenditure)
if(!require(betareg)) { install.packages("betareg"); require(betareg) }
# Cargar los datos de gasto en alimentación (38 hogares)
data("FoodExpenditure", package="betareg")
# Porcentaje de gasto en alimentación
x <- FoodExpenditure$food / FoodExpenditure$income
# Instalar y cargar la libreria fitdistrplus (gráficos y ajuste)
if(!require(fitdistrplus)) { install.packages("fitdistrplus"); require(fitdistrplus) }
# Gráfico
plotdist(x, histo=TRUE, demp=TRUE, pch="-")
```



Al ejecutar la función `descdist()` del paquete `fitdistrplus` en R, se obtienen medidas descriptivas para la muestra y un gráfico que me permite ver el coeficiente de asimetría y el de curtosis muestral con respecto a las regiones en el plano que ocuparían algunas distribuciones de acuerdo a los valores que pueden tomar sus coeficientes de asimetría y de curtosis (es decir los teóricos poblacionales).

```
# Medidas descriptivas y plano asimetría^2 vs kurtosis
descdist(x)
```




```
summary statistics
```

```
------
min:  0.1075258   max:  0.561243
median:  0.2610673
mean:  0.2896702
estimated sd:  0.1013732
estimated skewness:  0.9819276
estimated kurtosis:  4.160473
```

Generalmente una variable que representa un porcentaje tendrá valores entre 0% y 100%, y así mismo, una variable que representa una razón o proporción tendrá valores entre 0 y 1. Para modelar este tipo de variables sería apropiado utilizar una distribución con dominio acotado (con dominio en un intervalo (a, b)). La distribución Beta es una de las distribuciones más usadas para variables que toman valores en $(0, 1)$.

2.2.3.1 Intentamos solución analítica

Para la distribución beta en $(0, 1)$, se tienen las fórmulas que se muestran a continuación.

Función de densidad:

$$\mathrm{f}_X\left(x, \alpha, \beta\right) = \frac{1}{\mathrm{B}(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}\mathrm{I}_{(0,1)}(x)$$

donde $\mathrm{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ y $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$.

Espacio de los parámetros:

$$\theta = (\alpha, \beta)^T \in \Theta = (0, \infty) \times (0, \infty).$$

Función de log-verosimilitud:

$$\begin{aligned}
\log \mathrm{L}(\theta) &= \sum_{i=1}^{n} \log\left(\frac{1}{\mathrm{B}(\alpha, \beta)} x_i^{\alpha-1}(1-x_i)^{\beta-1}\right) \\
&= (\alpha - 1)\sum_{i=1}^{n} \log(x_i) + (\beta - 1)\sum_{i=1}^{n} \log(1 - x_i) - n\log\left(\mathrm{B}(\alpha, \beta)\right) \\
&= (\alpha - 1)\sum_{i=1}^{n} \log(x_i) + (\beta - 1)\sum_{i=1}^{n} \log(1 - x_i) \\
&\quad - n\left[\log\left(\Gamma(\alpha)\right) + \log\left(\Gamma(\beta)\right) - \log\left(\Gamma(\alpha+\beta)\right)\right].
\end{aligned}$$

Gradiente:

$$
\begin{aligned}
\frac{\partial}{\partial\alpha}\log \mathrm{L}(\theta) &= \sum_{i=1}^{n}\log(x_i) - n\frac{\partial}{\partial\alpha}\left[\log\left(\Gamma(\alpha)\right) + \log\left(\Gamma(\beta)\right) - \log\left(\Gamma(\alpha+\beta)\right)\right] \\
&= \sum_{i=1}^{n}\log(x_i) - n\left[\psi(\alpha) - \psi(\alpha+\beta)\right] \\
\frac{\partial}{\partial\beta}\log \mathrm{L}(\theta) &= \sum_{i=1}^{n}\log(1-x_i) - n\frac{\partial}{\partial\beta}\left[\log\left(\Gamma(\alpha)\right) + \log\left(\Gamma(\beta)\right) - \log\left(\Gamma(\alpha+\beta)\right)\right] \\
&= \sum_{i=1}^{n}\log(1-x_i) - n\left[\psi(\beta) - \psi(\alpha+\beta)\right]
\end{aligned}
$$

donde $\frac{\partial}{\partial x}\log\Gamma(x)$ se denotará $\psi(x)$ (función digamma).

Hessiana:

$$
\begin{aligned}
\frac{\partial^2}{\partial\alpha^2}\log \mathrm{L}(\theta) &= -n\frac{\partial}{\partial\alpha}\left[\psi(\alpha) - \psi(\alpha+\beta)\right] \\
&= -n\left[\psi_1(\alpha) - \psi_1(\alpha+\beta)\right] \\
\frac{\partial^2}{\partial\beta^2}\log \mathrm{L}(\theta) &= -n\frac{\partial}{\partial\beta}\left[\psi(\beta) - \psi(\alpha+\beta)\right] \\
&= -n\left[\psi_1(\beta) - \psi_1(\alpha+\beta)\right] \\
\frac{\partial^2}{\partial\alpha\partial\beta}\log \mathrm{L}(\theta) &= -n\frac{\partial}{\partial\alpha}\left[\psi(\beta) - \psi(\alpha+\beta)\right] \\
&= n\psi_1(\alpha+\beta)
\end{aligned}
$$

donde $\frac{\partial}{\partial x}\psi(x) = \frac{\partial^2}{\partial x^2}\log\Gamma(x)$ se denotará $\psi_1(x)$ (función trigamma).

> ⚠
>
> Al parecer debido a la forma de funciones gamma, digamma y trigamma, hasta ahora nadie ha podido obtener una formula o expresión explícita y cerrada para el estimador máximo verosímil de una distribución Beta con parámetros α y β. ¿Qué hacer en una situación como esta?

2.2.3.2 Optimización mediante métodos numéricos

Exceptuando algunos casos especiales, las ecuaciones de verosimilitud,

$$
\frac{\partial}{\partial\theta}\mathrm{L}(\theta; x_1, \dots, x_n) = 0 \quad \text{y} \quad \frac{\partial}{\partial\theta}\log \mathrm{L}(\theta; x_1, \dots, x_n) = 0,
$$

no pueden solucionarse de manera analítica. Por tanto, es necesario usar métodos numéricos o procedimientos iterativos para poder solucionar el problema de optimización asociado a dichas ecuaciones,

$$\arg\sup_{\theta\in\Theta} \mathrm{L}\left(\theta; x_1, \ldots, x_n\right) \quad \text{y} \quad \arg\sup_{\theta\in\Theta} \log \mathrm{L}\left(\theta; x_1, \ldots, x_n\right).$$

Existen variados métodos numéricos para resolver problemas de optimización. Buena parte de ellos toman un valor inicial para θ, notado $\hat{\theta}_0$, y a partir de él buscan obtener una secuencia convergente $\{\hat{\theta}_r\}_{r=0,1,\ldots}$. Los algoritmos más comunmente utilizados, los **Hill climbing algorithms**, se basan en una ecuación de actualización de este tipo,

$$\hat{\theta}_{r+1} = \hat{\theta}_r + \lambda_r \, \mathrm{d}_r\left(\hat{\theta}_r\right)$$

donde el vector $\mathrm{d}_r\left(\hat{\theta}_r\right)$ indica la "dirección del r-ésimo paso" y el escalar λ_r representa la "longitud o tamaño de ese r-ésimo paso".

En el caso de los **Gradient ascent algorithms**, la dirección de cada paso es la del gradiente de la función objetivo, $\mathrm{d}_r\left(\hat{\theta}_r\right) = \Delta\left(\hat{\theta}_r\right)$. El gradiente de la función de log-verosimilitud con respecto al vector de parámetros suele denominarse **score**, $\Delta\left(\theta\right) = \frac{\partial}{\partial\theta}\log \mathrm{L}$.

En el método de **Newton–Raphson**, $\lambda_r = 1$ y $\mathrm{d}_r\left(\hat{\theta}\right) = -\mathrm{H}^{-1}\left(\hat{\theta}_r\right)\Delta\left(\hat{\theta}_r\right)$, donde $\Delta(\theta)$ es el score y $\mathrm{H}^{-1}(\theta)$ es el inverso de la matrix Hessiana de la función de log-verosimilitud $\left(\mathrm{H}(\theta) = \frac{\partial^2}{\partial\theta\partial\theta^T}\log \mathrm{L}\right)$, ambos evaluados en $\hat{\theta}_r$.

El método denominado **Fisher's Scoring** es casi identico al método de Newton-Raphson, pero en vez de utilizar el negativo del inverso de la matrix Hessiana de la función de log-verosimilitud $\left(-\mathrm{H}^{-1}(\theta)\right)$, utiliza el inverso de la matrix de información de Fisher $\left(\mathcal{I}^{-1}\left(\theta\right)\right)$. La matrix de **información de Fisher** es la varianza del score,

$$\mathcal{I}\left(\theta\right) = Var\left[\Delta\left(\theta\right)\right] = E\left[\Delta\left(\theta\right)\Delta\left(\theta\right)^T\right]$$

y bajo ciertas condiciones (condiciones de regularidad), también es igual al inverso aditivo del valor esperado de la matrix Hessiana de la función de log-verosimilitud,

$$\mathcal{I}\left(\theta\right) = Var\left[\Delta\left(\theta\right)\right] = E\left[\Delta\left(\theta\right)\Delta\left(\theta\right)^T\right] = -E\left[\mathrm{H}(\theta)\right].$$

Dado que el cálculo de la matrix Hessiana o el de la matrix de información (más el cálculo de la respectiva inversa) son computacionalmente costosos, otras alternativas han sido propuestas. Una de ellas es, por ejemplo, el algoritmo de **Broyden–Fletcher–Goldfarb–Shanno (BFGS)**.

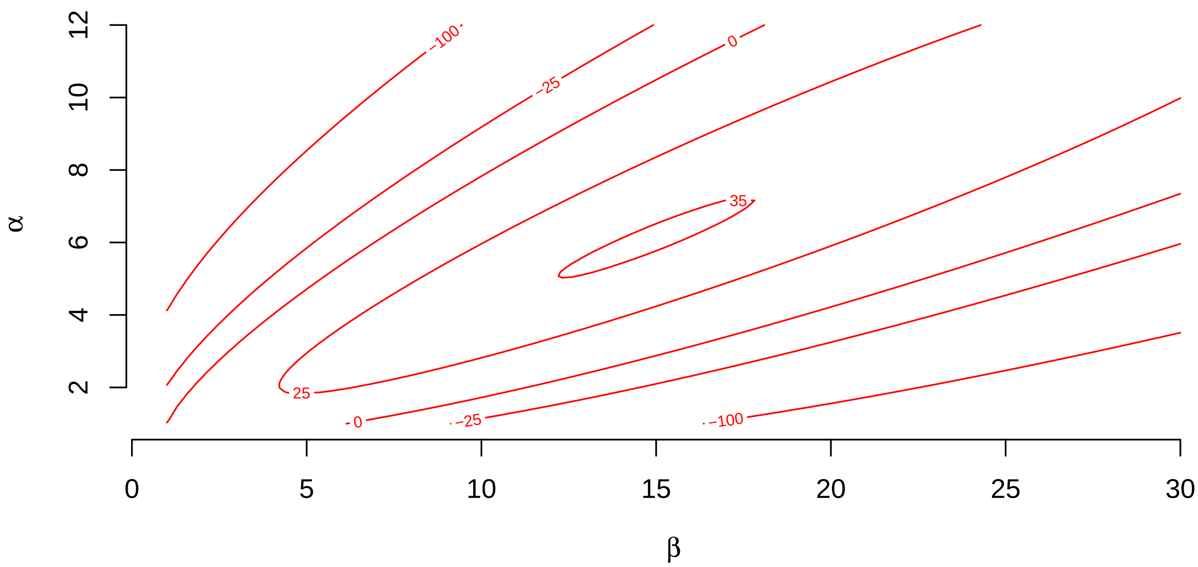
⚠

¿Existen situaciones en las cuales el método de optimización numérica utilizado podría "fallar"? ¿Qué debería hacer para estar seguro de que voy a encontrar o de que encontré el máximo de la función para todos los posibles valores de los parámetros?

2.2.3.3 Aplicando optimización numérica

Retomando los datos mencionados acerca del porcentaje de gasto en alimentación de los hogares, La siguiente gráfica muestra algunas curvas de nivel de la función de log-verosimilitud.

```
### log-verosimilitud con la distribución Beta
### para el porcentaje de gasto en alimentación
loglik.beta <- function(theta, sample)
    { sum( dbeta(sample, theta[1], theta[2], log = TRUE) ) }
### Gráfico de log-verosimilitud (curvas de nivel)
alpha <- seq(1, 12, length.out = 101)
beta <- seq(1, 30, length.out = 101)
z <- apply(expand.grid(alpha, beta), 1, loglik.beta, sample = x)
z <- matrix(z, 101, 101, byrow = TRUE)
contour(beta, alpha, z, drawlabels = TRUE, levels = c(-100,-25,0,25,35),
        col="red", bty = "none", xlab=quote(beta), ylab=quote(alpha))
```



Esperariamos encontrar el máximo $\left(\hat{\alpha}, \hat{\beta}\right)$ de dicha función dentro de la curva de nivel (elipse) etiquetada con el número 35.

Si queremos realizar la optimización numérica usando el método de Newton-Raphson, necesitamos calcular el gradiente y la Hessiana de la función de log-verosimilitud, para un valor de θ requerido.

```
cat("theta:\n")
```

```
theta:
```

```
a <- 6; b <- 15; theta <- c(a, b); print(theta) # un valor para theta
```

```
[1]  6 15
```

```
### Opción 1: Derivación Numérica.
### Instalar y cargar la libreria numDeriv (derivación numérica)
if(!require(numDeriv)) { install.packages("numDeriv"); require(numDeriv) }
### Gradiente numérica de la función loglik.beta
cat("\nGradiente numérica evaluada en theta (función numDeriv::grad):\n")
```

```
Gradiente numérica evaluada en theta (función numDeriv::grad):
```

```
print(numDeriv::grad(loglik.beta, x = theta, sample = x))
```

```
[1]  0.6878986 -0.2714764
```

```
### Hessiana numérica de la función loglik.beta
cat("\nHessiana numérica evaluada en theta (función numDeriv::hessian):\n")
```

```
Hessiana numérica evaluada en theta (función numDeriv::hessian):
```

```
print(numDeriv::hessian(loglik.beta, x = theta, sample = x))
```

```
          [,1]       [,2]
[1,] -5.036981  1.8532913
[2,]  1.853291 -0.7663614
```

```r
### Opción 2: Derivación automática o algorítmica
### Expresión LogL para la función de log-verosimilitud
n <- length(x) # Tamaño de la muestra
slx <- sum(log(x)) #
sl1mx <- sum(log(1-x)) #
logL <- expression((a-1)*slx+(b-1)*sl1mx-
                   n*(log(gamma(a))+log(gamma(b))-log(gamma(a+b))))
### Gradiente y Hessiana automática de la función logL
deriv.logL <- deriv(logL, c("a", "b"), hessian = TRUE) # Expresión
res <- eval(deriv.logL) # Evaluación
cat("\nGradiente automática evaluada en theta (función deriv):\n")
```

```
Gradiente automática evaluada en theta (función deriv):
```

```r
print(attr(res,"gradient"))
```

```
             a          b
[1,] 0.6878986 -0.2714764
```

```r
cat("\nHessiana automática evaluada en theta (función deriv):\n")
```

```
Hessiana automática evaluada en theta (función deriv):
```

```r
print(matrix(attr(res,"hessian"), 2))
```

```
          [,1]       [,2]
[1,] -5.036981  1.8532913
[2,]  1.853291 -0.7663614
```

```r
### Opción 3: Implementar formulas analíticas.
### Gradiente de la función de log-verosimilitud
score <- function(theta, sample){
    n <- length(sample)
```

```r
    c(sum(log(sample)), sum(log(1-sample))) -
        n * (digamma(theta) - digamma(sum(theta)))
}
cat("\nGradiente analítica evaluada en theta (función propia):\n")
```

```
Gradiente analítica evaluada en theta (función propia):
```

```r
print(score(theta, x))
```

```
[1]  0.6878986 -0.2714764
```

```r
### Hessiana de la función de log-verosimilitud
H <- function(theta, sample){
    n <- length(sample); aux1 <- trigamma(sum(theta))
    H <- -n * diag(trigamma(theta) - aux1)
    H[1,2] <- H[2,1] <- n * aux1
    H
}
cat("\nHessiana analítica evaluada en theta (función propia):\n")
```

```
Hessiana analítica evaluada en theta (función propia):
```

```r
print(H(theta, x))
```

```
          [,1]       [,2]
[1,] -5.036981  1.8532913
[2,]  1.853291 -0.7663614
```

Procedemos al método de Newton-Raphson como tal, utilizando una de las tres opciones presentadas para calcular el gradiente y la Hessiana.

```r
### Optimización usando Newton-Raphson, con gradiente y Hessiana analíticos
i <- 0; theta <- c(6, 15) # theta inicial
```

```r
res <- data.frame(iter = i, alpha = theta[1], beta = theta[2])
repeat{
    s <- score(theta, x)
    if(sum(s^2) < 1e-23){ break }
    theta <- theta - as.vector(solve(H(theta, x)) %*% s)
    i <- i + 1
    res <- rbind(res, c(i, theta))
}
res[,1] <- as.factor(res[,1])
knitr::kable(format(res, nsmall=10))
```

iter	alpha	beta
0	6.0000000000	15.0000000000
1	6.0565369274	14.7824825727
2	6.0716009801	14.8219937900
3	6.0716413191	14.8220997664
4	6.0716413194	14.8220997671

```r
cat("\nGradiente analítico evaluado en el theta de la última iteración:\n")
```

```
Gradiente analítico evaluado en el theta de la última iteración:
```

```r
print(s)
```

```
[1] 7.105427e-15 1.243450e-14
```

```r
cat("\nHessiana analítica evaluada en el theta de la última iteración:\n")
```

```
Hessiana analítica evaluada en el theta de la última iteración:
```

```r
print(H(theta, x))
```

```
          [,1]       [,2]
[1,] -4.939202  1.8629437
[2,]  1.862944 -0.7892224
```

Podemos utilizar otros métodos. Sin embargo, es importante recordar que $\theta = (\alpha, \beta) \in \mathbb{R}^+ \times \mathbb{R}^+ = \Theta$, y por ende, debemos garantizar que esto es tenido en cuenta por aquellos métodos que vayamos a utilizar.

```r
### Optimizacion usando múltiples métodos con optimx
### pero sin darle el gradiente o la Hessiana
### (by default optimx performs minimization)
### Instalar y cargar la libreria optimx (optimización)
if(!require(optimx)) { install.packages("optimx"); require(optimx) }
### Espacio de los parámetros: theta \in (0,\infty)^2
m.loglik.beta <- function(theta, x)
    { -loglik.beta(theta, x) } # -1 * log-verosimilitud(theta)
m.opt <- optimx(c(6, 15), m.loglik.beta, x = x, lower = 1e-6,
                control = list(all.methods = TRUE)) # todos
knitr::kable(summary(m.opt, order = value)) # reporte ordenado
```

	p1	p2	value	fevals	gevals	niter	convcode	kkt1	kkt2	xtime
nlminb	6.071641	14.82210	-3.534644e+01	11	28	9	0	TRUE	TRUE	0.001
L-BFGS-B	6.071635	14.82209	-3.534644e+01	9	9	NA	0	TRUE	TRUE	0.003
Rcgmin	6.071653	14.82214	-3.534644e+01	870	406	NA	1	TRUE	TRUE	0.038
spg	NA	NA	8.988466e+307	NA	NA	NA	9999	NA	NA	0.000
Rvmmin	NA	NA	8.988466e+307	NA	NA	NA	9999	NA	NA	0.001
bobyqa	NA	NA	8.988466e+307	NA	NA	NA	9999	NA	NA	0.000
nmkb	NA	NA	8.988466e+307	NA	NA	NA	9999	NA	NA	0.001
hjkb	NA	NA	8.988466e+307	NA	NA	NA	9999	NA	NA	0.001

Algunos métodos solamente funcionan para cuando $\Theta = \mathbb{R}^d$, donde d es el número de parámetros a estimar. Para poder usar esos métodos, podemos reparametrizar nuestro problema de optimización. Esto quiere decir que utilizamos una transformación adecuada para cambiar de parámetros (de espacio de parámetros), obteniendo un problema de optimización equivalente al que teniamos,

$$\underset{\theta \in \Theta}{\arg\sup}\ \log \mathrm{L}\left(\theta; x_1, \dots, x_n\right) = \underset{\theta^* \in \mathbb{R}^d}{\arg\sup}\ \log \mathrm{L}\left(\theta^*; x_1, \dots, x_n\right)$$

En el caso de los parámetros de la distribución Beta, podemos utilizar esta transformación,

$$
\begin{aligned}
\mathbb{R}^+ \times \mathbb{R}^+ &\longrightarrow \mathbb{R}^2 \\
\theta = (\alpha, \beta) &\longrightarrow (e^\alpha, e^\beta) = \theta^*
\end{aligned}
$$

```r
### Optimizacion usando múltiples métodos con optimx
### pero sin darle el gradiente o la Hessiana
### (by default optimx performs minimization)
### Reparametrizando theta = exp(eta). eta \in R x R
t.m.loglik.beta <- function(eta, x)
    { -loglik.beta(exp(eta), x) } # -1 * log-verosimilitud(eta)
t.m.opt <- optimx(c(0.8, 1.2), t.m.loglik.beta, x = x, # sin lower
                control = list(all.methods = TRUE)) # todos
t.m.opt$p1 <- exp(t.m.opt$p1); t.m.opt$p2 <- exp(t.m.opt$p2) # theta
knitr::kable(summary(t.m.opt, order = value)) # reporte ordenado
```

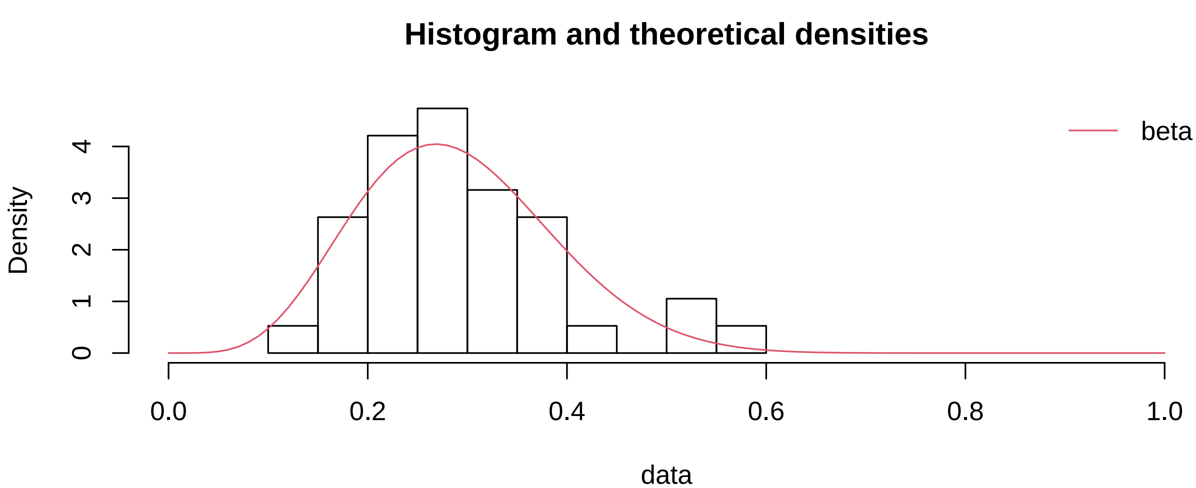
	p1	p2	value	fevals	gevals	niter	convcode	kkt1	kkt2	xtime
nlminb	6.071642	14.82210	-3.534644e+01	16	31	12	0	TRUE	TRUE	0.001
CG	6.071617	14.82204	-3.534644e+01	332	79	NA	0	TRUE	TRUE	0.010
Rvmmin	6.071614	14.82203	-3.534644e+01	261	62	NA	0	TRUE	TRUE	0.012
L-BFGS-B	6.071612	14.82203	-3.534644e+01	13	13	NA	0	TRUE	TRUE	0.001
Rcgmin	6.071606	14.82202	-3.534644e+01	875	254	NA	1	TRUE	TRUE	0.027
nlm	6.071523	14.82180	-3.534644e+01	NA	NA	10	0	TRUE	TRUE	0.001
Nelder-Mead	6.071250	14.82130	-3.534644e+01	75	NA	NA	0	TRUE	TRUE	0.002
BFGS	6.070940	14.82002	-3.534644e+01	25	11	NA	0	TRUE	TRUE	0.006
spg	NA	NA	8.988466e+307	NA	NA	NA	9999	NA	NA	0.000
ucminf	NA	NA	8.988466e+307	NA	NA	NA	9999	NA	NA	0.000
newuoa	NA	NA	8.988466e+307	NA	NA	NA	9999	NA	NA	0.000
bobyqa	NA	NA	8.988466e+307	NA	NA	NA	9999	NA	NA	0.000
nmkb	NA	NA	8.988466e+307	NA	NA	NA	9999	NA	NA	0.001
hjkb	NA	NA	8.988466e+307	NA	NA	NA	9999	NA	NA	0.001

Uno o varios de los métodos de optimización numérica suelen estar implementados e incorporados de alguna manera en paquetes o funciones existentes.

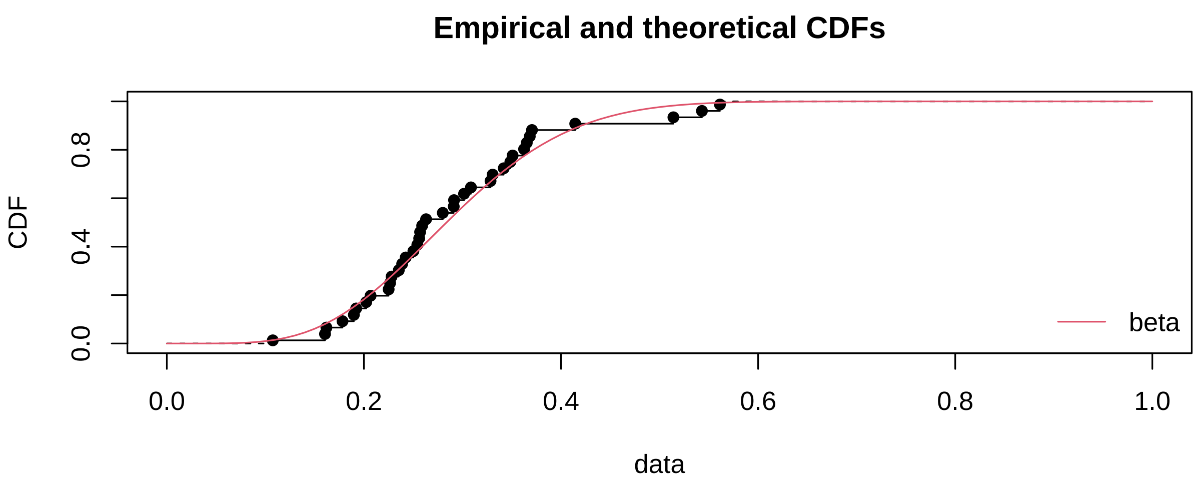
```
### Estimación máximo verosímil usando fitdistrplus::fitdist
### Instalar y cargar la libreria fitdistrplus (gráficos y ajuste)
if(!require(fitdistrplus)) { install.packages("fitdistrplus"); require(fitdistrplus) }
fit.b <- fitdist(x, "beta") # estimación (by default method="mle")
summary(fit.b)
```

```
Fitting of the distribution ' beta ' by maximum likelihood
Parameters :
       estimate Std. Error
shape1  6.07213   1.358727
shape2 14.82243   3.398868
Loglikelihood:  35.34644   AIC:  -66.69289   BIC:  -63.41772
Correlation matrix:
          shape1    shape2
shape1 1.0000000 0.9435681
shape2 0.9435681 1.0000000
```

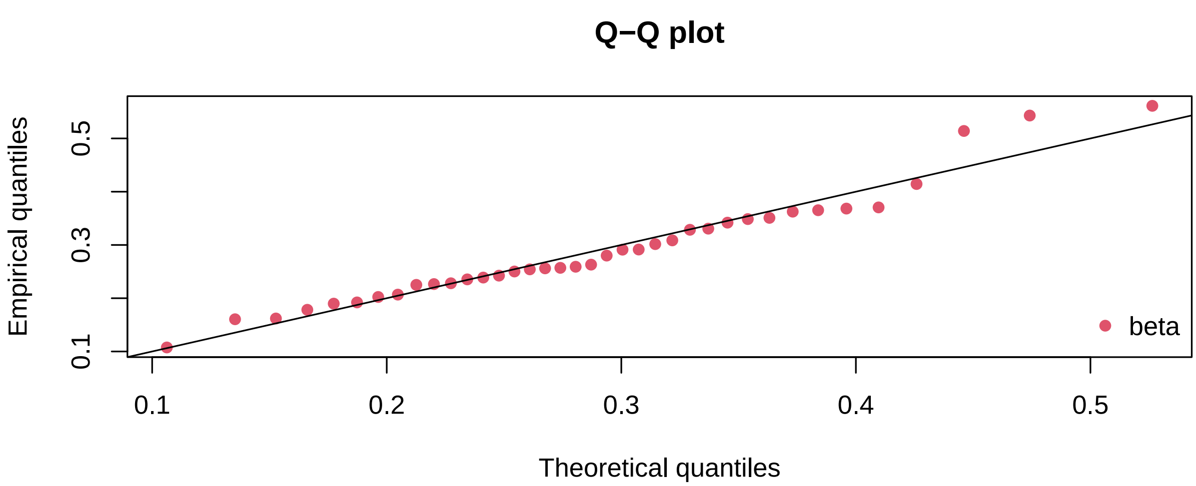
```
denscomp(list(fit.b), xlim = c(0,1)) # Función de densidad ajustada
```



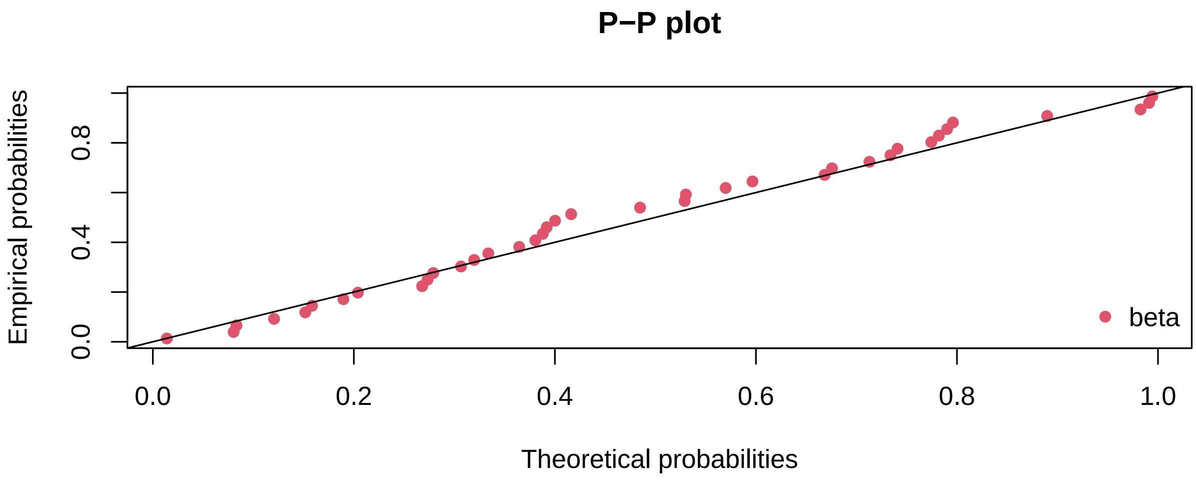

```
cdfcomp(list(fit.b), xlim = c(0,1)) # Función de distribución ajustada
```




```
qqcomp(list(fit.b), fitpch=16) # Gráfico Q-Q (Cuantil-Cuantil)
```




```
ppcomp(list(fit.b), fitpch=16) # Gráfico P-P (Probabilidad-Probabilidad)
```




```
cat("máxima log-verosimilitud alcanzada usando Newton-Raphson:\n",
    format(loglik.beta(theta, x), nsmall=14), "\n",
    "máxima log-verosimilitud alcanzada usando optimx:\n",
    format(-min(m.opt$value), nsmall=14), "\n",
    "máxima log-verosimilitud alcanzada reparametrizando y usando optimx:\n",
    format(-min(t.m.opt$value), nsmall=14), "\n",
    "máxima log-verosimilitud alcanzada usando fitdistrplus::fitdist:\n",
    format(fit.b$loglik, nsmall=14), sep="")
```

```
máxima log-verosimilitud alcanzada usando Newton-Raphson:
35.34644474252015
máxima log-verosimilitud alcanzada usando optimx:
35.34644474252016
máxima log-verosimilitud alcanzada reparametrizando y usando optimx:
35.34644474251994
máxima log-verosimilitud alcanzada usando fitdistrplus::fitdist:
35.34644441043961
```

2.2.4 Principio de invarianza

Dada una muestra aleatoria $X_1, \ldots, X_n$ proveniente de una población con distribución $\mathrm{f}_X(x, \theta)$, donde θ es el vector de parámetros, dada una función del vector de parámetros g(.), y suponiendo que $\mathrm{T}(X_1, \ldots, X_n)$ es el estimador de máxima verosimilitud de θ, entonces el estimador de máxima verosimilitud de $\mathrm{g}(\theta)$ es la estadística $\mathrm{g}(\mathrm{T}(X_1, \ldots, X_n))$.

2.3 Método de momentos

El método de momentos se puede asociar a la resolución de un sistema de ecuaciones,

$$\hat{\theta}_{MM} = \{\theta \in \Theta : (\hat{\eta}_1, \dots, \hat{\eta}_k) = (\eta_1(\theta), \dots, \eta_k(\theta))\}$$

donde $\eta_1(\theta), \dots, \eta_k(\theta)$ pueden ser momentos ordinarios, centrales o estandarizados poblacionales, expresados como funciones de los parámetros, y $\hat{\eta}_1, \dots, \hat{\eta}_k$ serían los respectivos momentos muestrales asociados.

2.3.1 Momentos poblacionales y muestrales

Sean $X_1, \dots, X_n$ una muestra aleatoria de una población con momentos ordinarios y centrales de orden r,

$$\mu_*^{(r)} = E\left[X^r\right] \quad \text{y} \quad \mu^{(r)} = E\left[\left(X - \mu_*^{(1)}\right)^r\right]$$

Los **momentos muestrales ordinarios y centrales de orden** $r = 1, 2, \dots$ se denotan y definen como,

$$M_*^{(r)} = \frac{1}{n}\sum_{i=1}^{n} X_i^r \quad \text{y} \quad M^{(r)} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - M_*^{(1)}\right)^r.$$

2.3.2 Una solución analítica

Encontrar $\hat{\theta}_{MM} = T(x_1, \dots, x_n)$ cuando $X_1, \dots, X_n$ es una muestra aleatoria de una población con $X \sim Gamma(k = \theta_1, \beta = \theta_2)$ $\left(\mathrm{f}_X(x, \theta) = \frac{1}{\theta_2^{\theta_1}\Gamma(\theta_1)} x^{\theta_1 - 1}\, \mathrm{e}^{-\frac{x}{\theta_2}}\, \mathrm{I}_{(0,\infty)}(x)\right)$.

Como tengo dos parámetros por estimar, necesito dos momentos. Dos momentos me permitan tener dos ecuaciones para esas dos incógnitas. El primer momento ordinario de una variable aleatoria X distribuida Gamma con parámetro de forma θ_1 y parámetro de escala θ_2 es,

$$E[X] = \theta_1\theta_2,$$

y el segundo momento central es,

$$E\left[(X - E(X))^2\right] = Var[X] = \theta_1\theta_2^2$$

El sistema de ecuaciones me quedaría así,

$$\theta_1\theta_2 = \mu_*^{(1)} = M_*^{(1)} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

$$\theta_1\theta_2^2 = \mu^{(2)} = M^{(2)} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

Para resolver el sistema de ecuaciones podría tomar la segunda ecuación y dividirla por la primera,

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{\bar{X}}=\frac{\theta_1\theta_2^2}{\theta_1\theta_2}=\theta_2,$$

despejando θ_1 y reemplazando θ_2 en la primera ecuación,

$$\begin{aligned}\theta_1\theta_2&=\bar{X}\\ \theta_1&=\frac{\bar{X}}{\theta_2}\\ \theta_1&=\frac{\bar{X}^2}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\end{aligned}$$

luego $\hat{\theta}_{MM}=\left(\frac{\bar{X}^2}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2},\frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{\bar{X}}\right)$ es un estimador por el método de momentos de $\theta=(\theta_1,\theta_2)$.

2.4 Métodos en general

Sean $X_1,\dots,X_n$ una muestra aleatoria de una población con vector de parámetros θ. La mayoría de los métodos usados para obtener estimadores se pueden dividir en dos grupos; aquellos que tienen que ver con la resolución de un sistema de ecuaciones,

$$\hat{\theta}=\left\{\theta\in\Theta:\hat{h}_n(x)=h(x;\theta)\right\}$$

y aquellos que tienen que ver con la resolución de un problema de optimización,

$$\hat{\theta}=\arg\min_{\theta\in\Theta}\ \mathrm{d}\left[\hat{h}_n(x),h(x;\theta)\right]$$

donde $\mathrm{d}\left[,\right]$ es una distancia generalizada, $\hat{h}_n(x)$ es una función escalar o vectorial asociada a la muestra y $h(x;\theta)$ es una función conformable asociada al modelo teórico seleccionado.

Si tuviese que hacer un listado de métodos:

- Maximum likelihood
- Maximum product of spacing
- Maximum goodness of fit or minimum distance for the Cumulative Distribution Function
  - Kolmogorov-Smirnov
  - Cramér-von Mises
  - Anderson-Darling
- Method of moments or moment matching

- Minimum distance for the Moments
- Minimum distance for the Moment Generating Function
- Minimum distance for the Characteristic Function
- Quantile matching
- Minimum distance for the Quantile Function

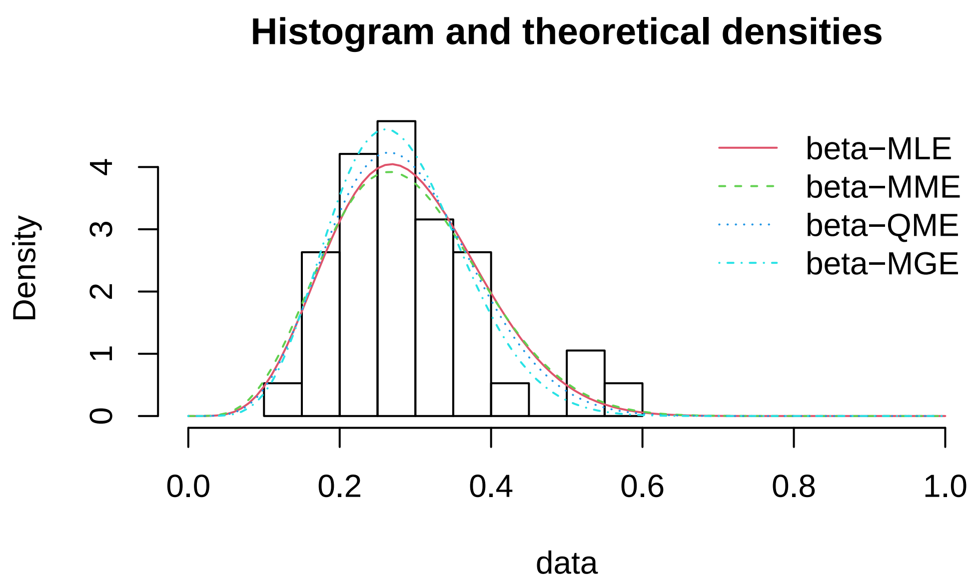
En el caso de los métodos que involucran una distancia, cabe la posibilidad de usar diferentes funciones, es decir diferentes métricas. Generalmente se suele usar la distancia euclidiana pero no es la única opción. Adicionalmente, se pueden introducir pesos diferenciados para darle más o menos importancia a alguna parte de una función. Por ejemplo, darle más peso a la distancia en la cola derecha de una función de distribución acumulativa y darle menor peso a la cola izquierda.

2.4.1 Usando paquete fitdistrplus

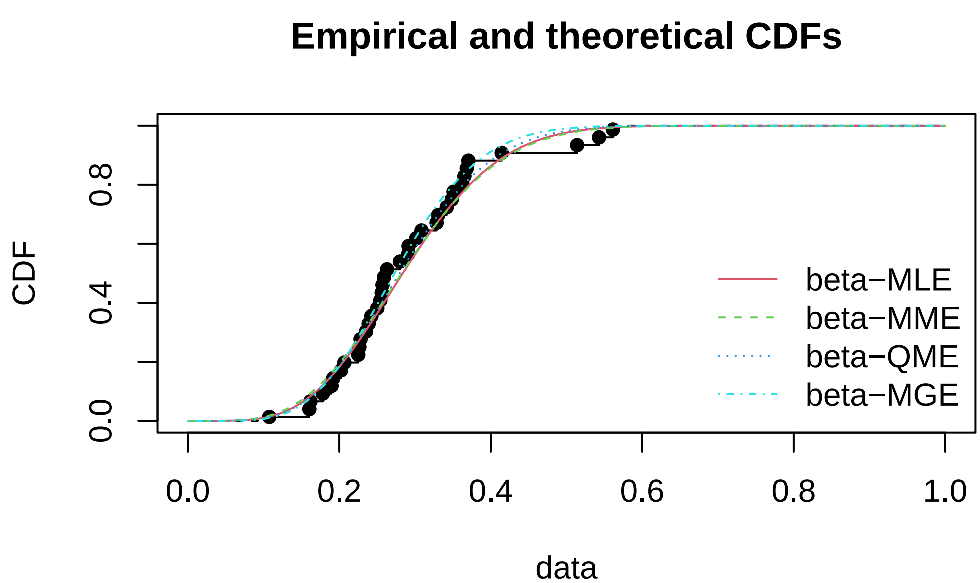
```
### Instalar y cargar la libreria betareg (datos FoodExpenditure)
if(!require(betareg)) { install.packages("betareg"); require(betareg) }
library(betareg)
### Instalar y cargar la libreria fitdistrplus (ajuste)
if(!require(fitdistrplus)) { install.packages("fitdistrplus"); require(fitdistrplus) }
library(fitdistrplus)
### Cargar los datos de gasto en alimentación (38 hogares)
data("FoodExpenditure", package="betareg")
### Porcentaje de gasto en alimentación
x <- FoodExpenditure$food / FoodExpenditure$income
### Estimación fitdistrplus::fitdist
fit.b <- fitdist(x, "beta") # (by default method="mle")
fit.b.mme <- fitdist(x, "beta", method = "mme") # moment matching
fit.b.qme <- fitdist(x, "beta", method = "qme",
                     probs = c(1/3, 2/3)) # quantile matching
fit.b.KSe <- fitdist(x, "beta", method = "mge",
                     gof = "KS") # Kolmogorov-Smirnov
rbind(MLE = fit.b$estimate, MME = fit.b.mme$estimate,
      QME = fit.b.qme$estimate, MGE.KS = fit.b.KSe$estimate)
```

```
         shape1   shape2
MLE    6.072130 14.82243
MME    5.666999 13.89662
QME    6.483029 16.17922
MGE.KS 7.362950 19.04666
```

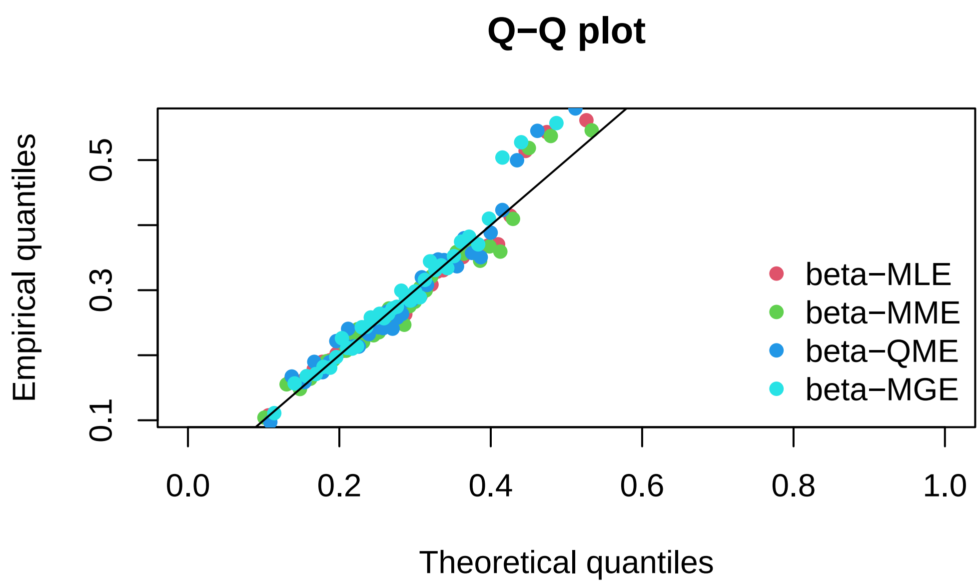
```r
denscomp(list(fit.b, fit.b.mme, fit.b.qme, fit.b.KSe), xlim = c(0,1))
```



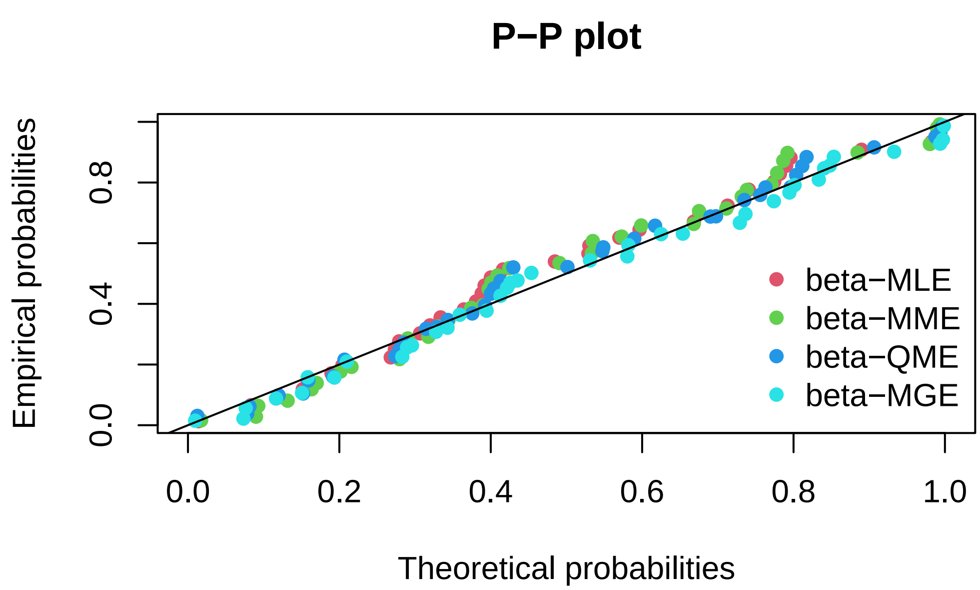

```r
cdfcomp(list(fit.b, fit.b.mme, fit.b.qme, fit.b.KSe), xlim = c(0,1))
```




```r
qqcomp(list(fit.b, fit.b.mme, fit.b.qme, fit.b.KSe), xlim = c(0,1), fitpch=16)
```




```
ppcomp(list(fit.b, fit.b.mme, fit.b.qme, fit.b.KSe), xlim = c(0,1), fitpch=16)
```



2.5 Ejercicios

1. Encontrar $\hat{\theta}_{ML} = T(x_1, \ldots, x_n)$ cuando $X_1 \ldots, X_n$ es una muestra aleatoria de una población con:

- $X \sim \mathcal{U}(a = 0, b = \theta)$
- $X \sim \mathcal{U}(a = -\theta, b = \theta)$
- $X \sim \mathcal{U}(a = \theta - 0.5, b = \theta + 0.5)$
- $X \sim \mathcal{E}(\beta = \theta)$
- $X \sim \mathcal{N}(\mu = \theta, \sigma^2 = \sigma_0^2 \text{ conocido})$
- $X \sim \mathcal{N}(\mu = \mu_0 \text{ conocido}, \sigma^2 = \theta)$
- $X \sim \mathcal{N}(\mu = \theta_1, \sigma^2 = \theta_2)$
- $X \sim \mathcal{Beta}(\alpha = \theta, \beta = 1)$
- $X \sim \mathcal{Gamma}(k = k_0 \text{ conocido}, \beta = \theta)$
- $X \sim \mathcal{P}(\lambda = \theta)$

2. Encontrar $\hat{\theta}_{MM} = T(x_1, \dots, x_n)$ cuando $X_1 \dots, X_n$ es una muestra aleatoria de una población con:

   - $X \sim \mathcal{U}(0, \theta)$
   - $X \sim \mathcal{U}(-\theta, \theta)$
   - $X \sim \mathcal{U}(\theta - 0.5, \theta + 0.5)$
   - $X \sim \mathcal{U}(\theta_1, \theta_2)$
   - $X \sim \mathcal{E}(\theta)$
   - $X \sim \mathcal{N}(\theta_1 = \mu, \theta_2 = \sigma^2)$
   - $X \sim \mathcal{Beta}(\theta_1 = \alpha, \theta_2 = \beta)$
   - $X \sim \mathcal{P}(\theta = \lambda)$
   - $X \sim \mathcal{B}(\theta = p)$
   - $X \sim \mathcal{B}(\theta_1 = n, \theta_2 = p)$
   - $X \sim \mathcal{H}g(\theta_1 = n, \theta_2 = N, \theta_3 = R)$

3 Propiedades algunas estadísticas

En esta sección se hará una revisión rápida de algunas propiedades matemático-estadísticas de interés de algunas estadísticas de particular relevancia.

Si $X_1, \ldots, X_n$ es una muestra aleatoria, entonces, en términos generales $T = \mathrm{T}\left(X_1, \ldots, X_n\right)$ es una estadística. A su vez, T también es una variable aleatoria, y por lo tanto, podemos hablar de su valor esperado como medida de centralidad, de su varianza como medida de dispersión, de su función de distribución acumulativa, de sus convergencias cuando el tamaño de muestra n es suficientemente grande (recordemos que para las variables aleatorias existes diferentes tipos de convergencia: en probabilidad, casi siempre, en distribución, en L^r), entre otros.

En la sección anterior, encontramos que $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ es un estimador para un parámetro θ de ciertas distribuciones. Por tal razón, $\bar{X}$ sería una estadística interesante para la cual querríamos conocer sus propiedades, como variable aleatoria que es. Así mismo, existen otras estadísticas interesantes y de uso frecuente, para las cuales el conocer sus propiedades como variables aleatorias, nos facilitará el **poder evaluarlas en su rol de estimadores puntuales de un parámetro de interés**.

3.1 Momentos muestrales

Recordemos que hemos denotado,

$$\mu_*^{(r)} = E\left[X^r\right],$$

y,

$$\mu^{(r)} = E\left[\left(X - \mu_*^{(1)}\right)^r\right] = E\left[(X - \mu)^r\right]$$

como los momentos (poblacionales) ordinarios y centrales de orden r, respectivamente.

Sean

$$M_*^{(r)} = \frac{1}{n}\sum_{i=1}^{n} X_i^r,$$

y

$$M^{(r)} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - M_*^{(1)}\right)^r,$$

los momentos muestrales ordinarios y centrales de orden r, respectivamente.

3.1.1 Valor esperado

$$
\begin{aligned}
E\left[M_*^{(r)}\right] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^r\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} E\left[X_i^r\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mu_*^{(r)} \\
&= \frac{1}{n} n\mu_*^{(r)} \\
&= \mu_*^{(r)}.
\end{aligned}
$$

3.1.2 Varianza

$$
\begin{aligned}
Var\left[M_*^{(r)}\right] &= \frac{1}{n}\left(E\left[X^{2r}\right] - \left(E\left[X^r\right]\right)^2\right) \\
&= \frac{1}{n}\left[\mu_*^{(2r)} - \left(\mu_*^{(r)}\right)^2\right].
\end{aligned}
$$

3.1.3 Convergencias

$$
M_*^{(r)} \xrightarrow{P} \mu_*^{(r)}
$$

3.2 Promedio muestral o media muestral

La media muestral o media de la muestra es el primer momento ordinario muestral,

$$
\bar{X} = M_*^{(1)} = \frac{1}{n}\sum_{i=1}^{n} X_i.
$$

3.2.1 Valor esperado

$$
E\left[\bar{X}\right] = E\left[M_*^{(1)}\right] = \mu_*^{(1)} = \mu.
$$

3.2.2 Varianza

$$
\begin{aligned}
Var\left[\bar{X}\right] &= Var\left[M_*^{(1)}\right] \\
&= \frac{1}{n}\left(\mu_*^{(2)} - \left(\mu_*^{(1)}\right)^2\right) \\
&= \frac{\sigma^2}{n}.
\end{aligned}
$$

Note que:

$$
\begin{aligned}
Var\left[\bar{X}\right] &= E\left[\bar{X}^2\right] - \Big(E\left[\bar{X}\right]\Big)^2 \\
\frac{\sigma^2}{n} &= E\left[\bar{X}^2\right] - \mu^2 \\
E\left[\bar{X}^2\right] &= \frac{\sigma^2}{n} + \mu^2.
\end{aligned}
$$

3.2.3 Distribución muestral

Si $X \sim Exp(\theta)$, entonces $\bar{X} \sim Gamma\left(n, \frac{\theta}{n}\right)$.

Si $X \sim Gamma(k, \theta)$, entonces $\bar{X} \sim Gamma\left(nk, \frac{\theta}{n}\right)$.

Si $X \sim \mathcal{N}(\mu, \sigma^2)$, entonces $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

3.2.4 Convergencias

$$
\begin{aligned}
&\bar{X} \xrightarrow{P} \mu \\
&\bar{X} \xrightarrow{c.s.} \mu
\end{aligned}
$$

> ⚠ ¿Cuál debería ser el tamaño de la muestra n para tener una probabilidad de al menos el 0.95, de que el promedio muestral no difiera del teórico en más de una cuarta parte de la desviación estándar teórica (suponiendo población infinita y muestra simple)?

Teorema central del límite:

Sea $X_1, X_2, ...$ una sucesión de variables aleatorias independientes e igualmente distribuidas (cualquier distribución), con media denotada por μ y varianza finita denotada por σ^2, entonces,

$$
\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}\left(\bar{X} - \mu\right)}{\sigma} \xrightarrow{d} Z \sim \mathcal{N}(0, 1).
$$

3.3 Varianza muestral

La denominada varianza muestral está dada por la expresión,

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 .$$

> ⚠ ¿Por qué S^2 es la varianza muestral y no es $M^{(2)} = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$?

3.3.1 Valor esperado

$$\begin{aligned}
E\left[M^{(2)}\right] &= E\left[\frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2\right] \\
&= E\left[\frac{1}{n} \sum_{i=1}^{n} \left(X_i^2\right) - \bar{X}^2\right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \Big(E\left[X_i^2\right]\Big) - E\left[\bar{X}^2\right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(\sigma^2 + \mu^2\right) - \left(\frac{\sigma^2}{n} + \mu^2\right) \\
&= \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 \\
&= \sigma^2 - \frac{\sigma^2}{n} = \left(1 - \frac{1}{n}\right) \sigma^2 \\
&= \frac{n-1}{n} \sigma^2 .
\end{aligned}$$

$$\begin{aligned}
E\left[S^2\right] &= E\left[\frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2\right] \\
&= E\left[\frac{n}{n-1} \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2\right] \\
&= \frac{n}{n-1} E\left[\frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2\right] \\
&= \frac{n}{n-1} \left(\frac{n-1}{n} \sigma^2\right) \\
&= \sigma^2 .
\end{aligned}$$

3.3.2 Varianza

$$
\begin{aligned}
Var\left[S^2\right] &= \frac{1}{n}\left(\mu^{(4)} - \frac{n-3}{n-1}\sigma^4\right) \\
&= \frac{\sigma^4}{n}\left((\kappa-1) + \frac{2}{n-1}\right),
\end{aligned}
$$

para $n > 1$ y donde κ es la curtosis poblacional.

3.3.3 Distribución muestral

Si $X \sim \mathcal{N}(\mu, \sigma^2)$, entonces $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.

3.3.4 Convergencias

$$
S^2 \xrightarrow{c.s.} \sigma^2,
$$

y por ende,

$$
S^2 \xrightarrow{P} \sigma^2.
$$

3.4 Estadísticas de orden

La k**-ésima estadística de orden**, $k = 1, 2, \dots, n$, denotada por $X_{(k)}$, se puede definir recursivamente de la siguiente manera,

$$
\begin{aligned}
X_{(1)} &= \min\{X_1, \dots, X_n\} \\
X_{(k)} &= \min\left\{\{X_1, \dots, X_n\} \;\; \left\{X_{(1)}, \dots, X_{(k-1)}\right\}\right\}
\end{aligned}
$$

Al conjunto de estadísticas de orden $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ se le denomina **muestra aleatoria ordenada**.

Mínimo muestral:

$$
X_{(1)}
$$

Máximo muestral:

$$
X_{(n)} = \min\left\{\{X_1, \dots, X_n\} \;\; \left\{X_{(1)}, \dots, X_{(n-1)}\right\}\right\} = \max\{X_1, \dots, X_n\}
$$

Mediana muestral:

$$
M_e = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{si } n \text{ es impar} \\ \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} & \text{si } n \text{ es par} \end{cases}
$$

3.4.1 Distribución muestral

$$F_{X_{(k)}}(x) = \sum_{i=k}^{n} \binom{n}{i} \left[F_X(x)\right]^i \left[1 - F_X(x)\right]^{n-i}$$

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} \left[F_X(x)\right]^{k-1} \left[1 - F_X(x)\right]^{n-k} f_X(x)$$

Distribución del mínimo muestral:

$$F_{X_{(1)}}(x) = 1 - \left[1 - F_X(x)\right]^n$$

Distribución del máximo muestral:

$$F_{X_{(n)}}(x) = \left[F_X(x)\right]^n$$

Distribución de la mediana muestral: Si $X \sim \mathcal{N}(\mu, \sigma^2)$, entonces $M_e \sim \mathcal{N}\left(\mu, \frac{\pi\sigma^2}{2n}\right)$.

3.4.2 Convergencias

Sea $F_X(x)$ estrictamente monotona y $F_X(x_p) = p$ entonces,

$$X_{([np]+1)} \xrightarrow{d} X \sim \mathcal{N}\left(x_p, \frac{p(1-p)}{n\left[f_X(x_p)\right]^2}\right)$$

3.5 Función de distribución acumulativa muestral

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, x]}(X_i) = \begin{cases} 0 & \text{si } x < X_{(1)} \\ k/n & \text{si } X_{(k)} \leq x < X_{(k+1)} \\ 1 & \text{si } X_{(n)} \leq x \end{cases}$$

3.5.1 Valor Esperado

$$E\left[F_n(x)\right] = F(x)$$

3.5.2 Varianza

$$Var[F_n(x)] = \frac{F(x)[1 - F(x)]}{n}$$

3.5.3 Convergencias

$$\mathrm{F}_n(x) \xrightarrow{P} \mathrm{F}(x)$$

para un valor x dado.

Es más, $\mathrm{F}_n(x)$ converge uniformemente a $\mathrm{F}(x)$.

$$\frac{\sqrt{n}\left(\mathrm{F}_n(x) - \mathrm{F}(x)\right)}{\sqrt{\mathrm{F}(x)\left[1 - \mathrm{F}(x)\right]}} \xrightarrow{d} Z \sim \mathcal{N}(0, 1)$$

Ejercicio 3.1. Determinar el valor esperado y la varianza de las estadísticas $\bar{X}$, $M^{(2)}$ y S^2 cuando la muestra aleatoria proviene de un población con,

- $X \sim \mathcal{N}(\mu = 5, \sigma^2 = 1.5^2)$
- $X \sim \mathcal{U}(a = 0, b = 10)$
- $X \sim Exp(\theta = 4/3)$
- $\frac{X}{10} \sim Beta(\alpha = 15, \beta = 1)$

Al tratar de determinar los valores esperados y las varianzas solicitadas, requerirá las formulas específicas de cada distribución para los primeros cuatro momentos centrales poblacionales. Comparando los valores que obtendría con los valores esperados y las varianzas determinadas por usted (usando $n = 100$), contra los siguientes valores aproximados obtenidos mediante simulación, puede identificar inicialmente si alguno de los valores esperados o varianzas determinadas está mal.

```r
n <- 100 # Tamaños de muestra
rep <- 10000 # Repeticiones
d <- c("rnorm(n * rep, 5, 1.5)",
       "runif(n * rep, 0, 10)",
       "rexp(n * rep, 0.75)",
       "10 * rbeta(n * rep, 15, 1)") # Distribuciones
df <- NULL
for(k in 1:4){ # Para cada distribución
  ## Generación de muestras (cada fila es una muestra)
  x <- matrix(eval(parse(text = d[k])), rep, n)
  x.bar <- apply(x, 1, mean) # promedio muestral
  s.2 <- apply(x, 1, var) # varianza muestral
  m.2 <- s.2 * (n - 1) / n # 2do momento central muestral
  df <- rbind(df, c(mean(x.bar), var(x.bar),
                    mean(m.2), var(m.2),
                    mean(s.2), var(s.2)))
```

```r
}
colnames(df) <- c("mean(x.bar)", "var(x.bar)",
                  "mean(m.2)", "var(m.2)",
                  "mean(s.2)", "var(s.2)")
rownames(df) <- c("Normal(0, 1.5^2)","Uniforme(0, 10)",
                  "Exp(4/3)","10*Beta(15, 1)")
cat("Tamaño de muestra utilizado: ", n, ". ",
    "Número de repeticiones utilizadas ", rep, ".\n", sep = "")
```

```
Tamaño de muestra utilizado: 100. Número de repeticiones utilizadas 10000.
```

```r
knitr::kable(df)
```

	mean(x.bar)	var(x.bar)	mean(m.2)	var(m.2)	mean(s.2)	var(s.2)
Normal(0, 1.5^2)	4.999322	0.0228168	2.2321929	0.1030152	2.2547403	0.1051068
Uniforme(0, 10)	4.998394	0.0825600	8.2443497	0.5465572	8.3276260	0.5576545
Exp(4/3)	1.334307	0.0179806	1.7628536	0.2509338	1.7806602	0.2560288
10*Beta(15, 1)	9.375287	0.0034343	0.3412862	0.0066379	0.3447335	0.0067727

4 Evaluación de estimadores

En esta sección se hará una revisión de algunos temas relacionados con diferentes aspectos a tener en cuenta para evaluar y comparar estimadores.

A menos que se especifique lo contrario, en adelante, sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con función de densidad $\mathrm{f}_X(x;\theta)$, $\mathrm{g}(\theta)$ una función del parámetro θ y $T = \mathrm{T}(X_1, \ldots, X_n)$ un estimador de $\mathrm{g}(\theta)$.

4.1 Insesgamiento

Definición 4.1. El estimador T se denomina **estimador insesgado** para $\mathrm{g}(\theta)$, si y solo si,

$$E\left[T\right] = \mathrm{g}(\theta)$$

Definición 4.2. La diferencia

$$\mathrm{Bias}[T] = E\left[T\right] - \mathrm{g}(\theta)$$

se denomina el **sesgo del estimador** T con respecto a $\mathrm{g}(\theta)$.

Definición 4.3. T es un **estimador asintóticamente insesgado** si

$$\lim_{n\to\infty} \left(E[T_n] - \mathrm{g}(\theta)\right) = 0$$

para todo $\theta \in \Theta$.

Ejercicio 4.1. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución $X \sim \mathcal{U}(a = 0, b = \theta)$. Determine el sesgo para,

- el estimador máximo verosimil de θ.
- el estimador por el método de momentos de θ, usando el primer momento ordinario.

Ejercicio 4.2. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución $X \sim \mathcal{N}(\mu = \mu_0 \text{ conocido}, \sigma^2 = \theta)$. Determine el sesgo para,

- el estimador de θ: $T = M^{(2)} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$.
- el estimador de θ: $T = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$.

4.2 Eficiencia - Precisión

Definición 4.4. Sean T y T^* dos estimadores insesgados para $\text{g}(\theta)$, se dice que T es un estimador **uniformemente mejor** que T^* si,

$$Var[T] \leq Var[T^*],$$

para todo $\theta \in \Theta$.

Ejercicio 4.3. Sea $X_1, \dots, X_n$ una muestra aleatoria de una población con distribución $X \sim \mathcal{U}(a = 0, b = \theta)$. Determine la varianza de,

- el estimador máximo verosimil de θ.
- el estimador por el método de momentos de θ, usando el primer momento ordinario.

¿Cuál estimador es uniformemente mejor?

Definición 4.5. Un estimador T es **insesgado y uniformemente de varianza mínima (uniformly minimum variance unbiased estimator UMVUE)** para $\text{g}(\theta)$ si y sólo si T es un estimador insesgado para $\text{g}(\theta)$ y $Var[T] \leq Var[T^*]$ para T^*, cualquier otro estimador insesgado para $\text{g}(\theta)$.

Bajo ciertas condiciones, la desigualdad de Cramer - Rao permitirá determinar cuál es la varianza mínima posible de los estimadores insesgados para $\text{g}(\theta)$. Antes de poder enunciar dicha desigualdad, debemos hablar de la **información de Fisher** y de las **condiciones de regularidad**.

Definición 4.6 (Información de Fisher)**.** Sea X una variable aleatoria cuya función de densidad $f_X(x;\theta)$ es tal que $\frac{\partial}{\partial\theta} \log f_X(x;\theta)$ existe para todo x, con $f_X(x;\theta) > 0$, y para todo $\theta \in \Theta$, la **información de Fisher** en la variable aleatoria X acerca de θ, se define como,

$$\mathcal{I}(\theta) = E\left[\left(\frac{\partial}{\partial\theta} \log f_X(X,\theta)\right)\left(\frac{\partial}{\partial\theta} \log f_X(X,\theta)\right)'\right]$$

o escrito de otra manera,

$$\mathcal{I}_{jk}(\theta) = E\left[\left(\frac{\partial}{\partial\theta_j} \log f_X(X,\theta)\right)\left(\frac{\partial}{\partial\theta_k} \log f_X(X,\theta)\right)\right].$$

De lo anterior, la **información de Fisher** para una muestra aleatoria acerca del parámetro θ estaría dada por,

$$E\left[\left(\frac{\partial}{\partial\theta_j}\log\prod_{i=1}^{n}f_X(X_i,\theta)\right)\left(\frac{\partial}{\partial\theta_k}\log\prod_{i=1}^{n}f_X(X_i,\theta)\right)\right]$$
$$=E\left[\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta_j}\log f_X(X_i,\theta)\right)\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta_k}\log f_X(X_i,\theta)\right)\right]$$

⚠ Sean X_1 y X_2 variables aleatorias independientes con información de Fisher $\mathcal{I}_{X_1}(\theta)$ y $\mathcal{I}_{X_2}(\theta)$, respectivamente. ¿la información de Fisher acerca del parámetro θ contenida en el vector aleatorio $X=(X_1,X_2)$ es $\mathcal{I}_X(\theta)=\mathcal{I}_{X_1}(\theta)+\mathcal{I}_{X_2}(\theta)$?. En general, si $X_1,\dots,X_n$ son variables aleatorias independientes idénticamente distribuidas y si la información de Fisher acerca de θ contenida en cada X_i es $\mathcal{I}(\theta)$, ¿la información de Fisher acerca de θ contenida en $X_1,\dots,X_n$ es $n\mathcal{I}(\theta)$?.

⚠ ¿El anterior resultado se cumplirá para una muestra aleatoria que proviene de una población con:

- $X\sim\mathcal{U}(a=0,b=\theta)$?
- $X\sim Exp(\beta=\theta)$?

Definición 4.7 (Caso regular de estimación o condiciones de regularidad)**.** Se habla de un **caso regular de estimación** o de **cumplimiento de condiciones de regularidad** cuando el modelo escogido para representar el comportamiento de la población y la estadística en consideración cumplen las siguientes condiciones:

1. $\frac{\partial}{\partial\theta}\log f_X(X,\theta)$ existe para todo x, con $f_X(x,\theta)>0$, y para todo $\theta\in\Theta\subseteq\mathbb{R}$.
2. La información de Fisher acerca del parámetro θ en la población es finita para todo $\theta\in\Theta$
3. Si la variable X que representa a la población es continua,

$$\frac{\partial}{\partial\theta}\int\cdots\int\prod_{i=1}^{n}f_X(x_i,\theta)\,dx_1\dots dx_n$$
$$=\int\cdots\int\frac{\partial}{\partial\theta}\prod_{i=1}^{n}f_X(x_i,\theta)\,dx_1\dots dx_n,$$

análogamente cuando X es discreta.

4. Si la variable X que representa a la población es continua,

$$\begin{aligned}
&\frac{\partial}{\partial\theta}\int\cdots\int \mathrm{T}(x_1,\dots,x_n)\prod_{i=1}^{n} f_X(x_i,\theta)\,dx_1\dots dx_n\\
&=\int\cdots\int \mathrm{T}(x_1,\dots,x_n)\frac{\partial}{\partial\theta}\prod_{i=1}^{n} f_X(x_i,\theta)\,dx_1\dots dx_n,
\end{aligned}$$

análogamente cuando X es discreta.

Bajo condiciones de regularidad y si $\frac{\partial^2}{\partial\theta^2}\log f_X(x,\theta)$ existe para todo x, con $f_X(x,\theta)>0$, y para todo $\theta\in\Theta\subseteq\mathbb{R}$, entonces,

$$E\left[\left(\frac{\partial}{\partial\theta}\log f_X(X,\theta)\right)^2\right]=-E\left[\frac{\partial^2}{\partial\theta^2}\log f_X(X,\theta)\right],$$

A partir del resultado anterior, si $X_1,\dots,X_n$ son variables aleatorias independientes idénticamente distribuidas (muestra aleatoria) y si la información de Fisher acerca de θ contenida en cada X_i es $\mathcal{I}(\theta)$, entonces,

$$\begin{aligned}
\mathcal{I}_{X_1,\dots,X_n}(\theta)&=E\left[\left(\frac{\partial}{\partial\theta}\sum_{i=1}^{n}\log f_X(X_i,\theta)\right)\left(\frac{\partial}{\partial\theta}\sum_{i=1}^{n}\log f_X(X_i,\theta)\right)'\right]\\
&=-E\left[\frac{\partial^2}{\partial\theta^2}\sum_{i=1}^{n}\log f_X(X_i,\theta)\right]\\
&=\sum_{i=1}^{n}\left(-E\left[\frac{\partial^2}{\partial\theta^2}\log f_X(X_i,\theta)\right]\right)\\
&=\sum_{i=1}^{n}\mathcal{I}(\theta)=n\,\mathcal{I}(\theta)
\end{aligned}$$

Ejercicio 4.4. Obtenga

$$\mathcal{I}(\theta)=E\left[\left(\frac{\partial}{\partial\theta}\log f_X(X,\theta)\right)^2\right]$$

y

$$-E\left[\frac{\partial^2}{\partial\theta^2}\log f_X(X,\theta)\right]$$

para,

- $X\sim\mathcal{U}(a=0,b=\theta)$
- $X\sim Exp(\beta=\theta)$

- $X \sim \mathcal{N}(\mu = \theta, \sigma^2 = \sigma_0^2 \text{ conocido})$

Teorema 4.1 (Desigualdad de Cramer - Rao)**.** *Dentro de un caso regular de estimación (bajo condiciones de regularidad), si T es un estimador insesgado para* $\mathrm{g}(\theta)$*, entonces,*

$$\frac{\left(\frac{\partial}{\partial\theta}\mathrm{g}(\theta)\right)^2}{n\,\mathcal{I}(\theta)} \leq Var\left[T\right]$$

expresión que corresponde a la versión más difundida del teorema, y en donde, la parte izquierda de la desigualdad es conocida como ***cota de Cramer-Rao****.*

De manera más general, sin la condición de que T sea un estimador insesgado, se tiene que,

$$\frac{\left(\frac{\partial}{\partial\theta}\mathrm{g}(\theta) + \frac{\partial}{\partial\theta}\mathrm{Bias}[T]\right)^2}{n\mathcal{I}(\theta)} \leq E\left[\left(T - \mathrm{g}(\theta)\right)^2\right]$$

Ejercicio 4.5. Sea $X_1, \dots, X_n$ una muestra aleatoria de una población con distribución $X \sim \mathcal{N}(\mu = \theta, \sigma^2 = \sigma_0^2 \text{ conocido})$.

- ¿Cuál es la información de Fisher de θ?
- ¿Cuál es la mínima varianza posible para un estimador insesgado de θ?
- ¿Cuál es el sesgo y la varianza del estimador máximo verosímil de θ?

La igualdad en la desigualdad de Cramer-Rao se tiene cuando,

$$\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f_X(X_i, \theta) = \mathrm{K}(\theta, n)\left(\mathrm{T}(X_1, \dots, X_n) - \mathrm{g}(\theta)\right)$$

En este caso, T es **UMVUE** para $\mathrm{g}(\theta)$.

- Cuando un estimador insesgado tiene varianza igual a la cota de Cramer-Rao, entonces este es un UMVUE. Sin embargo, un UMVUE puede que no alcance la cota de Cramer-Rao.
- Una forma para verificar que un estimador es UMVUE es ver que es insesgado y que alcanza la cota de Cramer-Rao.
- Un estimador sesgado puede tener una varianza inferior a la de todos los estimadores insesgados. Así mismo, un estimador sesgado puede que tenga una varianza igual o menor a la cota de Cramer-Rao y en cualquier caso nunca es UMVUE.

Teorema 4.2. *Si T el estimador máximo verosímil de θ está dado por la solución de la ecuación*

$$\frac{\partial}{\partial\theta}\log L(\theta, x_1, \dots, x_n) = \frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f_X(X_i, \theta) = 0$$

y si T^* es un estimador insesgado para $\mathrm{g}(\theta)$ cuya varianza coincide con la cota de Cramér - Rao, entonces $T^* = \mathrm{g}(T)$.

Definición 4.8. La **eficiencia relativa** de $T^{(2)}$ respecto a $T^{(1)}$, estimadores insesgados para $\mathrm{g}(\theta)$, corresponde a,

$$\frac{Var\left[T^{(1)}\right]}{Var\left[T^{(2)}\right]}$$

Ejercicio 4.6. Sea $X_1, \dots, X_n$ una muestra aleatoria de una población con distribución $X \sim \mathcal{U}(a = 0, b = \theta)$. Determine la eficiencia relativa del estimador máximo verosímil de θ respecto al estimador por el método de momentos de θ (usando el primer momento ordinario). ¿Qué cambiaría en la eficiencia relativa si cada estimador utiliza una muestra de diferente tamaño, por ejemplo, n y m respectivamente?

Definición 4.9. En un caso regular de estimación, la **eficiencia** de un estimador insesgado T para $\mathrm{g}(\theta)$ es,

$$\mathrm{e}(T) = \frac{\frac{\left(\frac{\partial}{\partial\theta}\mathrm{g}(\theta)\right)^2}{n\mathcal{I}(\theta)}}{Var[T]}$$

Definición 4.10. En un caso regular de estimación, la **eficiencia asintótica** de un estimador insesgado T para $\mathrm{g}(\theta)$ es,

$$\lim_{n\to\infty}\mathrm{e}(T_n)$$

4.3 Concentración

Definición 4.11. Sean $T^{(1)}$ y $T^{(2)}$ dos estimadores para $\mathrm{g}(\theta)$, se dice que $T^{(1)}$ es **más concentrado** que $T^{(2)}$, si y solo si,

$$P\left[-\epsilon < T^{(1)} - \mathrm{g}(\theta) < \epsilon\right] \geq P\left[-\epsilon < T^{(2)} - \mathrm{g}(\theta) < \epsilon\right]$$
$$P\left[\mathrm{g}(\theta) - \epsilon < T^{(1)} < \mathrm{g}(\theta) + \epsilon\right] \geq P\left[\mathrm{g}(\theta) - \epsilon < T^{(2)} < \mathrm{g}(\theta) + \epsilon\right]$$

para cada $\epsilon > 0$ y cada $\theta \in \Theta$.

Definición 4.12. Una medida de concentración del estimador T es

$$\mathrm{MSE}_T(\theta) = E\left[\left(T - \mathrm{g}(\theta)\right)^2\right]$$

la cual se denomina **Error cuadrático medio (Mean-Squared Error)**.

Ejercicio 4.7. Muestre que el error cuadrático medio de un estimador se puede descomponer en dos, en la varianza y en el cuadrado del sesgo de dicho estimador.

Solución

$$
\begin{aligned}
\mathrm{MSE}_T(\theta) &= E\left[(T-\mathrm{g}(\theta))^2\right] \\
&= E\left[\left((T-E[T])+(E[T]-\mathrm{g}(\theta))\right)^2\right] \\
&= E\left[(T-E[T])^2+2\,(T-E[T])\,(E[T]-\mathrm{g}(\theta))+(E[T]-\mathrm{g}(\theta))^2\right] \\
&= E\left[(T-E[T])^2\right]+2E\left[(T-E[T])\,(E[T]-\mathrm{g}(\theta))\right]+E\left[(E[T]-\mathrm{g}(\theta))^2\right] \\
&= E\left[(T-E[T])^2\right]+2\,(E[T]-\mathrm{g}(\theta))\,E\left[(T-E[T])\right]+(E[T]-\mathrm{g}(\theta))^2 \\
&= E\left[(T-E[T])^2\right]+2\,(E[T]-\mathrm{g}(\theta))\,(E\,[T]-E[T])+(E[T]-\mathrm{g}(\theta))^2 \\
&= E\left[(T-E[T])^2\right]+(E[T]-\mathrm{g}(\theta))^2 \\
&= Var[T]+(\mathrm{Bias}[T])^2
\end{aligned}
$$

Ejercicio 4.8. Sea $X_1,\dots,X_n$ una muestra aleatoria de una población con distribución $X\sim\mathcal{U}(a=0,b=\theta)$. Determine el error cuadrático medio para,

- el estimador máximo verosímil de θ.
- el estimador por el método de momentos de θ, usando el primer momento ordinario.

Para finalizar, recuerde que, bajo condiciones de regularidad,

$$
\frac{\left(\frac{\partial}{\partial\theta}\mathrm{g}(\theta)+\frac{\partial}{\partial\theta}\mathrm{Bias}[T]\right)^2}{n\mathcal{I}(\theta)}\le E\left[(T-\mathrm{g}(\theta))^2\right]
$$

4.4 Consistencia

Informalmente y en pocas palabras, consistencia trata acerca de si el estimador converge al parámetro a medida que el tamaño de muestra aumenta.

Definición 4.13. El estimador T_n se denomina **estimador consistente en error cuadrático medio** para $\mathrm{g}(\theta)$, si

$$
\lim_{n\to\infty}E\left[(T_n-\mathrm{g}(\theta))^2\right]=0
$$

o lo que es lo mismo,

$$T_n \xrightarrow{L^2} \mathrm{g}(\theta)$$

para todo $\theta \in \Theta$.

Definición 4.14. El estimador T_n se denomina **estimador consistente fuerte** para $\mathrm{g}(\theta)$, si

$$P\left[\lim_{n\to\infty} T_n = \mathrm{g}(\theta)\right] = 1$$

o lo que es lo mismo,

$$T_n \xrightarrow{c.s.} \mathrm{g}(\theta)$$

Definición 4.15. El estimador T_n se denomina **estimador consistente simple o débil** para $\mathrm{g}(\theta)$, si

$$\lim_{n\to\infty} P\left[-\epsilon < T_n - \mathrm{g}(\theta) < \epsilon\right] = 1$$

o lo que es lo mismo,

$$T_n \xrightarrow{P} \mathrm{g}(\theta)$$

para todo $\theta \in \Theta$ y $\epsilon > 0$.

> ⚠ ¿Un estimador obtenido por el método de momentos es un estimador consistente? ¿por qué si o por qué no?

Naturalmente, si T_n es un estimador asintóticamente insesgado y su varianza tiende a cero cuando el tamaño de la muestra tiende a infinito entonces T_n es consistente en error cuadrático medio.

Ejercicio 4.9. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución $X \sim \mathcal{N}(\mu = \theta, \sigma^2 = \sigma_0^2$ conocido). Determine si el estimador máximo verosímil de θ es consistente en error cuadrático medio.

Definición 4.16. T_n se denomina **best asymptotically normal estimator** si y solo si,

1. La sucesión de variables aleatorias $\{\sqrt{n}\,(T_n - \mathrm{g}(\theta))\}$ converge en distribución a una variable aleatoria con distribución Normal de valor esperado cero y varianza $\sigma^2(\theta)$.
2. El estimador es consistente débil para $\mathrm{g}(\theta)$.
3. Siendo T_n^* cualquier otro estimador consistente débil para $\mathrm{g}(\theta)$ para el cual $\{\sqrt{n}\,(T_n^* - \mathrm{g}(\theta))\}$ converge en distribución a una variable aleatoria con distribución Normal de valor esperado cero y varianza ${\sigma^*}^2(\theta)$, se tiene que $\sigma^2(\theta) \leq {\sigma^*}^2(\theta)$.

En un caso regular de estimación, si T_n es el estimador **máximo verosímil** para $\mathrm{g}(\theta)$, entonces T_n es un **estimador consistente asintóticamente normal**, es decir,

$$\sqrt{n}(T_n - \mathrm{g}(\theta)) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\mathcal{I}(\theta)}\right).$$

Por ejemplo: Sea $X_1, \dots, X_n$ una muestra aleatoria de una población con función de densidad $\mathrm{f}_X(x;\theta)$, una función continua en el percentil x_p, para un valor p fijado de antemano, entonces la estadística de orden $X_{([np]+1)}$ es un estimador consistente asintóticamente normal para el percentil x_p,

$$\frac{X_{([np]+1)} - x_p}{\sqrt{\frac{p(1-p)}{n\,\mathrm{f}_X^2(x_p;\theta)}}} \xrightarrow{d} \mathcal{N}(0,1).$$

4.5 Suficiencia

Informalmente y en pocas palabras, suficiencia trata acerca de si un estadístico conserva toda la información contenida en la muestra aleatoria ("no hay pérdida de información") con respecto al parámetro de interés.

Definición 4.17. $T_1, \dots, T_m$ es una colección de estadísticas **conjuntamente suficientes** para θ, si y sólo si la distribución condicional de las variables aleatorias de la muestra aleatoria $X_1, X_2, \dots, X_n$ dado $T_1, \dots, T_m$ no depende de θ.

Ejercicio 4.10. Sea X_1, X_2 una muestra aleatoria de tamaño 2 de una población con distribución $X \sim \mathcal{B}ernoulli(p = \theta)$. Determine si,

- $T = X_1 + X_2$ es una estadística suficiente para θ.
- $T^* = X_1 X_2$ es una estadística suficiente para θ.

Ejercicio 4.11. Sea $X_1, \dots, X_n$ una muestra aleatoria de una población con distribución $X \sim \mathcal{P}oisson(\lambda = \theta)$. Determine si el estimador máximo verosímil de θ es suficiente para θ.

Teorema 4.3 (Criterio de factorización de Fisher-Neyman)**.** *El conjunto $T_1, \dots, T_m$ es una colección de estadísticas **conjuntamente suficientes** para θ, si y solo si, la función de verosimilitud se puede factorizar de la forma,*

$$\mathrm{L}(\theta; x_1, \dots, x_n) = \mathrm{g}(T_1, \dots, T_m; \theta)\, \mathrm{h}(x_1, \dots, x_n)$$
$$\left(\mathrm{f}_{X_1,\dots,X_n}(x_1, \dots, x_n; \theta) = \mathrm{g}(T_1, \dots, T_m; \theta)\, \mathrm{h}(x_1, \dots, x_n)\right)$$

donde h *es una función no negativa que depende exclusivamente de $x_1, \dots, x_n$ y* g *es una función no negativa que depende de θ y de $x_1, \dots, x_n$ a través de $T_1, \dots, T_m$.*

Ejercicio 4.12. Utilice el criterio de factorización de Fisher-Neyman para obtener una colección de estadísticas conjuntamente suficientes para una muestra aleatoria proveniente de una:

- distribución Bernoulli de parámetro $p = \theta$.
- distribución normal de parámetros $\mu = \theta_1$ y $\sigma^2 = \theta_2$.
- distribución beta de parámetros $\alpha = \theta_1$ y $\beta = \theta_2$

Teorema 4.4. *Si T es un estimador suficiente para θ, entonces la información contenida en la v.a. T acerca de θ es la misma información contenida en la muestra aleatoria $X_1, \dots, X_n$ acerca de θ*

Prueba.

$$\begin{aligned}
\mathcal{I}_{X_1,\dots,X_n}(\theta) &= E\left[\left(\frac{\partial}{\partial\theta}\log \mathrm{f}_{X_1,\dots,X_n}(X_1,\dots,X_n)\right)^2\right] \\
&= E\left[\left(\frac{\partial}{\partial\theta}\log\Big(\mathrm{g}(T;\theta)\,\mathrm{h}(X_1,\dots,X_n)\Big)\right)^2\right] \\
&= E\left[\left(\frac{\partial}{\partial\theta}\Big(\log \mathrm{g}(T;\theta) + \log \mathrm{h}(X_1,\dots,X_n)\Big)\right)^2\right] \\
&= E\left[\left(\frac{\partial}{\partial\theta}\log \mathrm{g}(T;\theta)\right)^2\right]
\end{aligned}$$

Por otra parte, para obtener $\mathcal{I}_T(\theta)$, se necesita saber quién es $\mathrm{f}_T(t)$,

$$\begin{aligned}
\mathrm{f}_{X_1,\dots,X_n|T}(x_1,\dots,x_n|t) &= \frac{\mathrm{f}_{X_1,\dots,X_n,T}(x_1,\dots,x_n,t)}{\mathrm{f}_T(t)} \\
\mathrm{f}_T(t) &= \frac{\mathrm{f}_{X_1,\dots,X_n,T}(x_1,\dots,x_n,t)}{\mathrm{f}_{X_1,\dots,X_n|T}(x_1,\dots,x_n|t)} \\
\mathrm{f}_T(t) &= \begin{cases} \frac{\mathrm{f}_{X_1,\dots,X_n}(x_1,\dots,x_n)}{\mathrm{f}_{X_1,\dots,X_n|T}(x_1,\dots,x_n|t)} & \text{Si } T(x_1,\dots,x_n) = t \\ 0 & \text{en otro caso} \end{cases}
\end{aligned}$$

Lo que quiere decir que,

$$
\begin{aligned}
\mathcal{I}_T(\theta) &= E\left[\left(\frac{\partial}{\partial\theta}\log \mathrm{f}_T\,(T)\right)^2\right]\\
&= E\left[\left(\frac{\partial}{\partial\theta}\log \frac{\mathrm{f}_{X_1,\dots,X_n}(X_1,\dots,X_n)}{\mathrm{f}_{X_1,\dots,X_n|T}(X_1,\dots,X_n|T)}\right)^2\right]\\
&= E\left[\left(\frac{\partial}{\partial\theta}\log \frac{\mathrm{g}(T;\theta)\,\mathrm{h}(X_1,\dots,X_n)}{\mathrm{f}_{X_1,\dots,X_n|T}(X_1,\dots,X_n|T)}\right)^2\right]\\
&= E\Bigg[\bigg(\frac{\partial}{\partial\theta}\Big(\log\left(\mathrm{g}(T;\theta)\right)+\log\left(\mathrm{h}(X_1,\dots,X_n)\right)-\\
&\qquad\qquad \log\left(\mathrm{f}_{X_1,\dots,X_n|T}(X_1,\dots,X_n|T)\right)\Big)\bigg)^2\Bigg]\\
&= E\left[\left(\frac{\partial}{\partial\theta}\log \mathrm{g}(T;\theta)\right)^2\right]
\end{aligned}
$$

Concluimos que,

$$
\mathcal{I}_{X_1,\dots,X_n}(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log \mathrm{g}(T;\theta)\right)^2\right] = \mathcal{I}_T(\theta)
$$

□

Teorema 4.5. *Si $T_1,\dots,T_m$ es una colección de estadísticas* ***conjuntamente suficientes****, entonces cualquier transformación uno a uno de $T_1,\dots,T_m$ también es una colección de estadísticas* ***conjuntamente suficientes****.*

Ejercicio 4.13. Determine una colección de estimadores insesgados y conjuntamente suficientes para una muestra aleatoria de una:

- distribución Bernoulli de parámetro $p=\theta$.
- distribución normal de parámetros $\mu=\theta_1$ y $\sigma^2=\theta_2$.

Definición 4.18. Una colección de estadísticas conjuntamente suficientes se denomina **minimal**, si y sólo si, ellas son función de cualquier otro conjunto de estadísticas suficientes.

! Para el estudio de la suficiencia y la completez será particularmente ventajoso cuando la población tenga una distribución perteneciente a la "familia exponencial". En el Apéndice B hay una pequeña y rápida revisión de algunos aspectos básicos relacionados con la familia exponencial de densidades / distribuciones (**Familia exponencial**).

Si $\mathrm{f}_X(x;\theta)$ pertenece a la familia exponencial k-paramétrica de densidades, entonces las estadísticas,

$$\sum_{i=1}^{n} T_1\left(X_i\right), \dots, \sum_{i=1}^{n} T_k\left(X_i\right)$$

son conjuntamente suficientes para θ. Además se puede demostrar que constituyen una colección minimal.

Teorema 4.6. *Si T es una* ***estadística suficiente*** *para θ y si T_{ML} es el único* ***estimador máximo verosímil*** *para θ, entonces T_{ML} es función de T.*

Teorema 4.7.

- *Si T_{ML} es el único* ***estimador máximo verosímil*** *para θ, entonces T_{ML} es función de una* ***colección minimal de estadísticas conjuntamente suficientes****.*
- *Si T_{ML} NO es el único* ***estimador máximo verosímil*** *para θ, entonces existe un estimador máximo verosímil T^*_{ML} que es una función de una* ***colección minimal de estadísticas conjuntamente suficientes****.*

Definición 4.19. Se dice que las estadísticas T^* y T son **equivalentes** si existe una función $\mathrm{g}()$ uno a uno de tal manera que $T^* = \mathrm{g}(T)$.

Teorema 4.8. *Dadas las estadísticas equivalentes T^* y T, si T es una estadística suficiente para θ, entonces T^* también lo es.*

La idea contraria a la suficiencia se formaliza en los siguientes definición y teorema.

Definición 4.20. La estadística T se denomina estadística **auxiliar** para el parámetro θ, si $\mathrm{f}_T(t)$ es una función que no depende de θ. Si específicamente $E[T]$ es un valor que no depende de θ, entonces T se denomina estadística **auxiliar de primer orden**.

Ejercicio 4.14. Determine una estadística auxiliar de primer orden para el parámetro θ de una población con distribución $X \sim \mathcal{U}(a = -\theta, b = \theta)$.

Teorema 4.9 (Teorema de Basu)**.** *Si T^* es una estadística auxiliar para el parámetro θ y T es una estadística suficiente para θ, entonces T^* y T son variables aleatorias independientes.*

Teorema 4.10 (Teorema de Rao-Blackwell)**.** *Si T^* es un estimador insesgado para* $\mathrm{g}(\theta)$ *y $T_1, \dots, T_m$ son conjuntamente suficientes, entonces $T = E[T^*|T_1, \dots, T_m]$ es un estimador insesgado para $g(\theta)$, que es función solamente de estadísticas suficientes y que tiene una varianza menor que la de T^*.*

Ejercicio 4.15. Sea $X_1, \dots, X_n$ una muestra de una población con distribución $X \sim$ *Bernoulli*$(p = \theta)$.

1. Determine si $T^* = X_1$ es un estimador insesgado para θ $(g(\theta) = \theta)$.
2. Determine si $T_1 = \sum_{i=1}^{n} X_i$ es suficiente para θ.
3. Obtenga $T = E[T^*|T_1]$.
4. Verifique que T efectivamente cumple todas las condiciones que debería cumplir dado el anterior teorema ("estimador insesgado para $g(\theta) = \theta$, que es función solamente de estadísticas suficientes y que tiene una varianza menor que la de T^*").

4.6 Completez

Aspecto referente a la distribución muestral de un estadístico que nos será de utilidad para encontrar estimadores UMVUE.

Definición 4.21. La familia de funciones de distribución $\left(\{\mathrm{F}_X(x;\theta) : \theta \in \Theta\}\right)$ es **completa**, si y solo si,

$$E\big[\mathrm{g}(X)\big] = 0 \text{ implica } P\big[\mathrm{g}(X) = 0\big] = 1$$

para toda función de valor real g(), para todo $\theta \in \Theta$ y para todo x tal que $\mathrm{f}_X(x;\theta) > 0$.

Definición 4.22. T es una **estadística completa** si su distribución $\left(\mathrm{F}_T(x;\theta)\right)$ pertenece a una familia de distribuciones completa $\left(\{\mathrm{F}(x;\theta) : \theta \in \Theta\}\right)$.

Ejercicio 4.16. Determine si $T = \sum_{i=1}^{n} X_i$ es una estadística completa para $X_1, \dots, X_n$ una muestra aleatoria de una población con distribución $X \sim \mathcal{Bernoulli}(p = \theta)$.

Ejercicio 4.17. Determine si $T = X_{(n)}$ es una estadística completa para $X_1, \dots, X_n$ una muestra aleatoria de una población con distribución $X \sim \mathcal{U}(a = 0, b = \theta)$.

Teorema 4.11. *Sea $X_1, \dots, X_n$ una muestra aleatoria de una población con distribución perteneciente a la familia exponencial, la estadística natural $\left(T = \sum_{i=1}^{n} \mathrm{T}(X_i)\right)$ es una estadística suficiente y completa para θ.*

Teorema 4.12. *Sea T un estimador insesgado para* $\mathrm{g}(\theta)$*. Si T es una estadística completa para θ, entonces T es el único estimador insesgado para* $\mathrm{g}(\theta)$*.*

Teorema 4.13 (Teorema de Lehmann-Scheffe)**.** *Si $T_1, \ldots, T_m$ es una colección de estadísticas conjuntamente* ***suficientes*** *y* ***completas*** *para θ y si $T = \mathrm{h}(T_1, \ldots, T_m)$ es un estimador* ***insesgado*** *para $\mathrm{g}(\theta)$, entonces T es* ***UMVUE*** *para $\mathrm{g}(\theta)$.*

4.7 Ejercicios

1. Demuestre que:

    - Si $X \sim \mathcal{B}(p = \theta)$ entonces
    $$\mathcal{I}(\theta) = \frac{1}{\theta(1-\theta)}$$

    - Si $X \sim \mathcal{B}(n = n_0 \text{ conocido}, p = \theta)$ entonces
    $$\mathcal{I}(\theta) = \frac{n}{\theta(1-\theta)}$$

    - Si $X \sim \mathcal{P}(\lambda = \theta)$ entonces
    $$\mathcal{I}(\theta) = \frac{1}{\theta}$$

    - Si $X \sim \mathcal{B}eta(\alpha = \theta_1, \beta = \theta_2)$ entonces
    $$\mathcal{I}(\theta) = \begin{pmatrix} \mathcal{I}_{11}(\theta) & \mathcal{I}_{12}(\theta) \\ \mathcal{I}_{21}(\theta) & \mathcal{I}_{22}(\theta) \end{pmatrix} = \begin{pmatrix} \psi_1(\alpha) - \psi_1(\alpha+\beta) & -\psi_1(\alpha+\beta) \\ -\psi_1(\alpha+\beta) & \psi_1(\beta) - \psi_1(\alpha+\beta) \end{pmatrix}$$

    - Si $X \sim \mathcal{N}\left(\mu = \theta_1, \sigma^2 = \theta_2\right)$ entonces
    $$\mathcal{I}_{11}(\theta) = \frac{1}{\theta_2},$$
    $$\mathcal{I}_{22}(\theta) = \frac{1}{2\theta_2^2}$$
    y
    $$\mathcal{I}_{12}(\theta) = 0$$

    - Si $X \sim \mathcal{L}og\mathcal{N}ormal(\theta_1, \theta_2)$ entonces
    $$\mathcal{I}_{11}(\theta) = \frac{1}{\theta_2},$$
    $$\mathcal{I}_{22}(\theta) = \frac{1}{2\theta_2^2}$$
    y
    $$\mathcal{I}_{12}(\theta) = 0$$

- Si $X \sim \mathcal{P}areto(\theta_1, \theta_2)$ entonces

$$\mathcal{I}_{11}(\theta) = \frac{\theta_2^2}{\theta_1^2},$$

$$\mathcal{I}_{22}(\theta) = \frac{1}{\theta_2^2}$$

  y

$$\mathcal{I}_{12}(\theta) = 0$$

2. Resolver ejercicios, acerca de lo visto en esta sección, del capítulo 2 del libro del profesor **Mayorga**

3. Resolver ejercicios, acerca de lo visto en esta sección, de los capítulos 2, 3 y 4 del libro del profesor **Cepeda**.

Parte IV

Estimación por intervalo

5 Aspectos generales

En esta sección se hará una revisión de algunos aspectos generales relacionados con la estimación por intervalo, específicamente en cuanto a conceptos y definiciones iniciales. Adicionalmente, se presentará la idea general detrás del método de la variable pivote.

5.1 Conceptos y definiciones iniciales

Sean $X_1, \dots, X_n$ una muestra aleatoria de una población con función de densidad $\mathrm{f}_X(x;\theta)$, $\theta \in \Theta$, $\mathrm{g}(\theta)$ una función cuyo recorrido es un conjunto de números reales, y las estadísticas $T^{(1)} = \mathrm{T}^{(1)}\left(X_1, \dots, X_n\right)$ y $T^{(2)} = \mathrm{T}^{(2)}\left(X_1, \dots, X_n\right)$.

Un **intervalo aleatorio** es un intervalo en el cual al menos uno de sus extremos es una variable aleatoria.

Sean $T^{(1)}$ y $T^{(2)}$ tales que $P\left[T^{(1)} < T^{(2)}\right] = 1$; el intervalo aleatorio $\left(T^{(1)}, T^{(2)}\right)$ se denomina **intervalo de confianza del** $100(1-\alpha)\%$ para $\mathrm{g}(\theta)$, si

$$\mathrm{P}\left[T^{(1)} < \mathrm{g}(\theta) < T^{(2)}\right] = 1-\alpha$$

probabilidad que no depende de θ. El valor $1-\alpha$ se denomina **nivel de confianza**, $T^{(1)}$ se denomina **límite de confianza inferior** para $\mathrm{g}(\theta)$ y $T^{(2)}$ se denomina **límite de confianza superior** para $\mathrm{g}(\theta)$. Una realización particular del intervalo de confianza se denomina **estimación por intervalo del** $100(1-\alpha)\%$ **de confianza** para $\mathrm{g}(\theta)$.

El intervalo aleatorio $\left(T^{(1)}, \infty\right)$ es un **intervalo de confianza unilateral del** $100(1-\alpha)\%$ para $\mathrm{g}(\theta)$, si

$$P\left[T^{(1)} < \mathrm{g}(\theta)\right] = 1-\alpha$$

probabilidad que no depende de θ. $T^{(1)}$ recibe el nombre de **límite de confianza inferior unilateral** para $\mathrm{g}(\theta)$.

El intervalo aleatorio $\left(-\infty, T^{(2)}\right)$ es otro **intervalo de confianza unilateral del** $100(1-\alpha)\%$ para $\mathrm{g}(\theta)$, si se tiene que,

$$P\left[\mathrm{g}(\theta) < T^{(2)}\right] = 1-\alpha$$

probabilidad que no depende de θ. $T^{(2)}$ recibe el nombre de **límite de confianza superior unilateral** para $\mathrm{g}(\theta)$.

Sea $\left(T^{(1)}, T^{(2)}\right)$ un intervalo de confianza para θ; si $\mathrm{g}(\theta)$ es una función estrictamente monótona con dominio Θ y recorrido un subconjunto de los números reales, entonces, cuando $\mathrm{g}(\theta)$ es estrictamente creciente, $\left(\mathrm{g}\left(T^{(1)}\right), \mathrm{g}\left(T^{(2)}\right)\right)$ es un intervalo de confianza para $\mathrm{g}(\theta)$, y, cuando es estrictamente decreciente, $\left(\mathrm{g}\left(T^{(2)}\right), \mathrm{g}\left(T^{(1)}\right)\right)$ es un intervalo de confianza para $\mathrm{g}(\theta)$.

Sea $\mathrm{A}\left(x_1, \dots, x_n\right)$ un subconjunto del espacio de observaciones. $\mathrm{A}\left(X_1, \dots, X_n\right)$ se denomina **región de confianza del** $100(1-\alpha)\%$ para el parámetro θ, si

$$P[\theta \in \mathrm{A}\left(X_1, \dots, X_n\right)] = 1 - \alpha$$

probabilidad que no depende de θ.

5.2 Método de la variable pivote

Sea $Q_X = \mathrm{Q}(\theta, X_1, \dots, X_n)$ una función del parámetro θ y de las variables que conforman la muestra aleatoria $X_1, \dots, X_n$. Q se denomina **variable aleatoria pivote** para el parámetro θ si la distribución de Q no depende de θ.

Ejercicio 5.1. Para una muestra aleatoria de una población Normal de media $\mu = \theta$ y varianza σ^2 conocida, determine si,

$$Q = \frac{\bar{X} - \theta}{\sigma / \sqrt{n}} = \frac{\sqrt{n}(\bar{X} - \theta)}{\sigma}$$

es una variable pivote para θ (es decir, determine la distribución de la variable aleatoria Q, en donde será evidente si esta depende o no de θ).

El método de la variable pivote para la construcción de intervalos de confianza consiste en:

1. Determinar una variable pivote.
2. Obtener los cuantiles a y b a partir del nivel de confianza dado $(1-\alpha)$, de tal forma que,

$$P\left[a < Q < b\right] = 1 - \alpha,$$

3. Realizar las transformaciones que sean necesarias (por medio de sucesos/eventos aleatorios que sean equivalentes), con el objetivo de llegar a la forma,

$$P\left[T^{(1)} < \mathrm{g}(\theta) < T^{(2)}\right] = 1 - \alpha$$

concluyendo así que el intervalo $\left(T^{(1)}, T^{(2)}\right)$ es un intervalo de confianza del $100(1-\alpha)\%$ para $\mathrm{g}(\theta)$.

Ejercicio 5.2. Para una muestra aleatoria de una población Normal de media $\mu = \theta$ y varianza σ^2 conocida, a partir de una variable pivote, obtenga un intervalo de confianza bilateral del $100(1-\alpha)\%$ para θ.

Solución

De la solución del Ejercicio 5.1, sabríamos que $Q = \frac{\sqrt{n}(\bar{X}-\theta)}{\sigma}$ es una variable pivote y conoceríamos su respectiva distribución (la cual no depende de θ). Es así que para un $1-\alpha$ dado, gracias a esa distribución, podríamos encontrar a y b tales que,

$$P\left[a < Q < b\right] = 1-\alpha$$

Para los a y b que encontremos y transformando la variable pivote Q, tenemos que,

$$\begin{aligned}
P\left[a < Q < b\right] &= 1-\alpha \\
P\left[a < \sqrt{n}\frac{\bar{X}-\theta}{\sigma} < b\right] &= 1-\alpha \\
P\left[\frac{a\sigma}{\sqrt{n}} < \bar{X}-\theta < \frac{b\sigma}{\sqrt{n}}\right] &= 1-\alpha \\
P\left[-\bar{X}+\frac{a\sigma}{\sqrt{n}} < -\theta < -\bar{X}+\frac{b\sigma}{\sqrt{n}}\right] &= 1-\alpha \\
P\left[\bar{X}-\frac{b\sigma}{\sqrt{n}} < \theta < \bar{X}-\frac{a\sigma}{\sqrt{n}}\right] &= 1-\alpha.
\end{aligned}$$

Dados los datos de la muestra $(x_1, x_2, \ldots, x_n)$, una estimación por intervalo (un intervalo de confianza) del $100(1-\alpha)\%$ de nivel de confianza para θ sería,

$$\left(\bar{x}-(b)\frac{\sigma}{\sqrt{n}} < \theta < \bar{x}+(-a)\frac{\sigma}{\sqrt{n}}\right)$$

donde a y b son tales que,

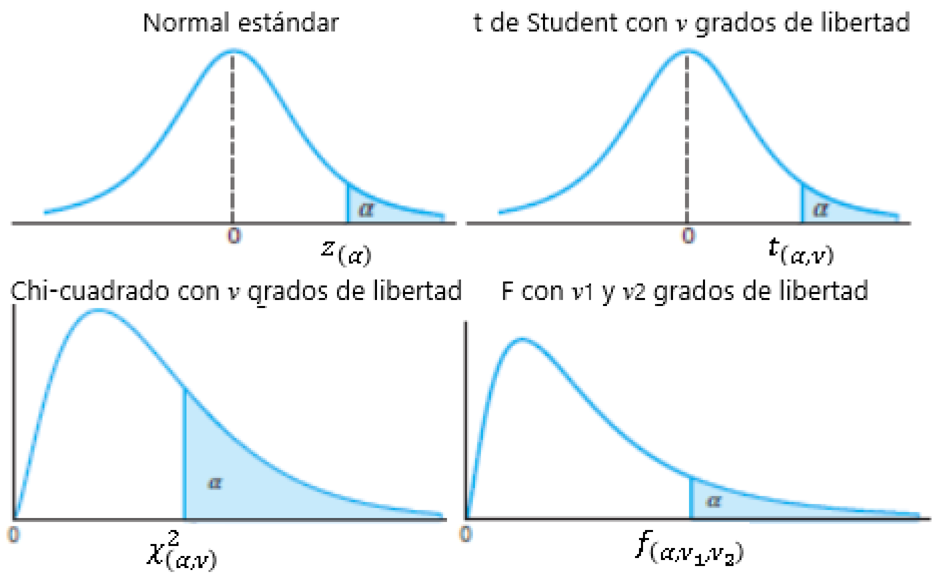
$$P\left[a < Q < b\right] = 1-\alpha$$

para Q con distribución conocida que no depende de θ.

6 Bajo normalidad

En esta sección se hará una revisión de algunos resultados relacionadas con la estimación por intervalo (intervalos de confianza) cuando se supone que las muestras aleatorias provienen de poblaciones con distribución normal.

En adelante se utilizará la siguiente notación, tal y como se representa en el siguiente gráfico.



Note la relación entre la notación que se suele utilizar para los cuantiles y esta notación adicional (mi intención es tener una pequeña variante en la notación que permita diferenciar y no confundir una cosa con otra):

$$
\begin{aligned}
z_{1-\alpha} &= z_{(\alpha)} \\
t_{1-\alpha,\nu} &= t_{(\alpha,\nu)} \\
\chi^2_{1-\alpha,\nu} &= \chi^2_{(\alpha,\nu)} \\
f_{1-\alpha,\nu_1,\nu_2} &= f_{(\alpha,\nu_1,\nu_2)}
\end{aligned}
$$

Adicionalmente, recordemos que:

- Si $Z_1, \ldots, Z_\nu$ son variables aleatorias normales estándar independientes, entonces

$$\sum_{i=1}^{\nu} Z_i^2 \sim \chi_\nu^2$$

- Si Z es una variable aleatoria normal estándar, V es una variable aleatoria chi-cuadrado con ν grados de libertad, y Z y V son independientes, entonces

$$\frac{Z}{\sqrt{V/\nu}} \sim t_\nu$$

- Si W es una variable aleatoria chi-cuadrado con ν_1 grados de libertad, V es una variable aleatoria chi-cuadrado con ν_2 grados de libertad, y W y V son independientes, entonces

$$\frac{W/\nu_1}{V/\nu_2} \sim F_{\nu_1,\nu_2}$$

- Si $X \sim \mathcal{N}(\mu, \sigma^2)$, entonces,
  - $\bar{X}$ y S^2 son variables aleatorias independientes (naturalmente, $\bar{X}$ y $M^{(2)}$ también lo son).
  - $$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$
  - $$\frac{\sum_{i=1}^n \left(X_i - \mu\right)^2}{\sigma^2} \sim \chi_n^2$$
  - $$\frac{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}{\sigma^2} = \frac{nM^{(2)}}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

6.1 Media

6.1.1 Varianza conocida

Para una muestra aleatoria de una población Normal de media $\mu = \theta$ y varianza σ^2 conocida; usando el método de la variable pivote, se deduce que,

$$\left(\bar{X} - (b)\left(\frac{\sigma}{\sqrt{n}}\right), \bar{X} + (-a)\left(\frac{\sigma}{\sqrt{n}}\right)\right)$$

es un intervalo de confianza bilateral del $100(1-\alpha)\%$ para θ, donde a y b serían cuantiles de la distribución normal estándar tales que $\Phi(b) - \Phi(a) = 1 - \alpha$, siendo $\Phi(.)$ la función de distribución acumulativa normal estándar (a partir de la realización del Ejercicio 5.1 y del Ejercicio 5.2).

Note que:

- Un estimador puntual de θ $\left(\bar{X}\right)$ es un punto de referencia para el intervalo (el intervalo está definido alrededor de ese punto).
- El límite superior está a una distancia de $-a$ veces la desviación estándar del mencionado estimador puntual (desviación estándar del estimador puntual = **error estándar** del estimador: $SE\left[\bar{X}\right] = SD\left[\bar{X}\right] = \sigma/\sqrt{n}$) y el límite inferior está a una distancia de b veces ese error estándar.
- Si se utiliza $\bar{x}$ para estimar θ, se tiene un $100(1-\alpha)\%$ de confianza en que la distancia entre $\bar{x}$ y θ no excederá $\max\left\{(-a)\frac{\sigma}{\sqrt{n}}, (b)\frac{\sigma}{\sqrt{n}}\right\}$, a veces denominado **error máximo admisible** o **margen de error**.
- Además, se tiene un $100(1-\alpha)\%$ de confianza en que el margen de error no excederá una cantidad dada e, cuando el tamaño de muestra es por lo menos de,

$$n = \left\lceil \left(\max\left\{(-a)\frac{\sigma}{e}, (b)\frac{\sigma}{e}\right\}\right)^2 \right\rceil.$$

  Claramente, para calcular el anterior tamaño de muestra se requiere un valor para σ. En la práctica, usualmente, dicho valor se desconoce por completo y en ese caso la recomendación es usar una desviación estándar estimada proveniente de un estudio previo o de una muestra piloto o preliminar.

Por otra parte, al tomar diferentes valores para a y b obtenemos infinitos intervalos de confianza del $100(1-\alpha)\%$ para θ. Si se desea el intervalo de confianza de menor longitud posible entonces,

$$\begin{aligned}
\min_{a<b}\left\{T^{(2)} - T^{(1)}\right\} &= \min_{a<b}\left\{\bar{X} + (-a)\frac{\sigma}{\sqrt{n}} - \left(\bar{X} - (b)\frac{\sigma}{\sqrt{n}}\right)\right\} \\
&= \min_{a<b}\left\{(-a)\frac{\sigma}{\sqrt{n}} + (b)\frac{\sigma}{\sqrt{n}}\right\} \\
&= \min_{a<b}\left\{(b-a)\frac{\sigma}{\sqrt{n}}\right\} = \min_{a<b}\left\{b-a\right\}
\end{aligned}$$

es decir, minimizar la longitud del intervalo es equivalente a minimizar $b-a$, con la condición o restricción de que $\Phi(b) - \Phi(a) - 1 + \alpha = 0$ (y de que $a < b$).

Derivando la función a minimizar con respecto a b, e igualando a cero se tiene que,

$$\frac{\partial}{\partial b}(b-a) = 1 - \frac{\partial a}{\partial b} = 0,$$

al derivar la restricción con respecto a b se tiene que,

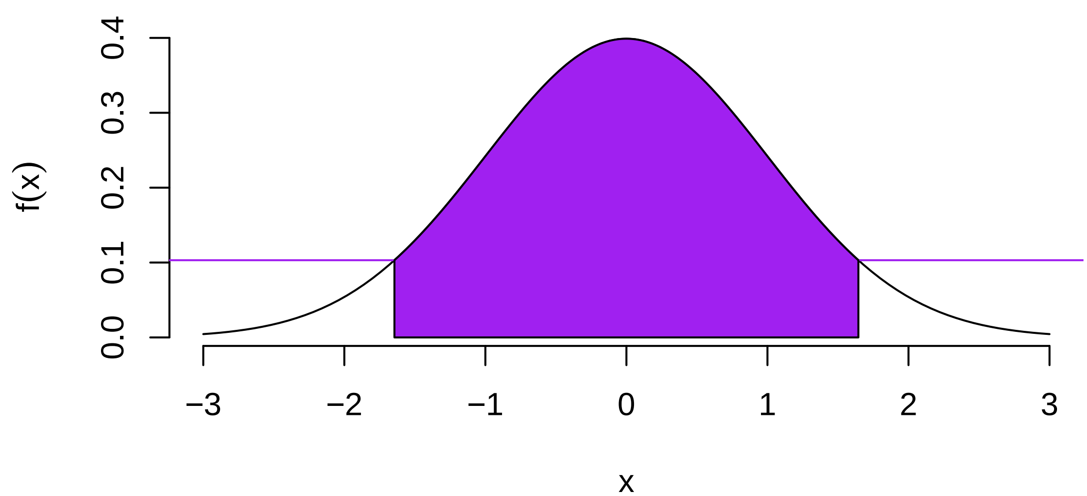
$$\begin{aligned}
\frac{\partial}{\partial b}(\Phi(b)-\Phi(a)-1+\alpha) &= \frac{\partial}{\partial b}0 \\
\phi(b)-\phi(a)\frac{\partial a}{\partial b} &= 0 \\
\frac{\partial a}{\partial b} &= \frac{\phi(b)}{\phi(a)}
\end{aligned}$$

y sustituyendo,

$$\begin{aligned}
1-\frac{\partial a}{\partial b} &= 0 \\
1-\frac{\phi(b)}{\phi(a)} &= 0 \\
\frac{\phi(b)}{\phi(a)} &= 1 \\
\phi(b) &= \phi(a)
\end{aligned}$$

Como la función de densidad normal estándar ($\phi(.)$) es simétrica, los únicos a y b tales que $a < b$ y $\phi(b) = \phi(a)$ son $b = -a = z_{(\alpha/2)} = \Phi^{-1}(\alpha/2)$ $\left(P\left[Z > z_{(\alpha/2)}\right] = \alpha/2\right)$.

```
alpha = 0.1; z <- qnorm(1-alpha/2); f.z <- dnorm(z)
curve(dnorm(x), -3, 3, bty="n", ylab = expression(f(x)))
abline(h = f.z, col = "purple")
x <- seq(-z, z, 0.01); y <- dnorm(x)
polygon(c(z,z,-z,-z,x), c(f.z,0,0,f.z,y), col = "purple")
```



En conclusión, el intervalo de confianza bilateral, de longitud mínima, del $100(1-\alpha)\%$ para

θ, bajo el supuesto de que la varianza σ^2 es conocida, es,

$$\left(\bar{X} - z_{(\alpha/2)}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{(\alpha/2)}\frac{\sigma}{\sqrt{n}}\right)$$

Ejercicio 6.1. Para el intervalo de confianza bilateral de longitud mínima:

- ¿Cuál sería la fórmula del respectivo margen de error?
- ¿Cuál sería la fórmula para calcular el menor tamaño de muestra que se requeriría para no pasarme de un cierto margen de error dado?

Ejercicio 6.2. (*D. Anderson, D. Sweeney, T. Williams, J. Camm - Estadística para Negocios y Economía. 11ra Edición, Cengage Learning (2012). Capítulo 8. Ejercicio 6.*) The Wall Street Journal informó que en 2008 los accidentes automovilísticos le costaron \$162 mil millones a Estados Unidos (The Wall Street Journal, 5 de marzo de 2008). El costo promedio por persona de los accidentes automovilísticos en el área de Tampa, Florida, fue considerado de \$1599. Suponga que este costo promedio se basó en una muestra de 50 personas que estuvieron involucradas en dichos percances y que la desviación estándar poblacional es \$600.

- Proporcione una estimación mediante un intervalo de confianza de 90% para el costo promedio por persona de los accidentes automovilísticos en el área de Tampa, Florida.
- ¿Cuál es el margen de error para un intervalo de 90% de confianza?
- ¿Qué recomendaría si el estudio requiriera un margen de error de \$50 o menos?
- ¿Qué le sucede a la amplitud del intervalo de confianza a medida que el nivel de confianza aumenta (90%, 95%, 99%)?
- ¿Qué le sucede a la amplitud del intervalo de confianza a medida que el tamaño de muestra aumenta (50, 100, 500)?

Ejercicio 6.3. (*D. Anderson, D. Sweeney, T. Williams, J. Camm - Estadística para Negocios y Economía. 11ra Edición, Cengage Learning (2012). Capítulo 8. Ejercicio 50.*) Se efectúan pruebas de rendimiento de gasolina con un determinado modelo de automóvil. Si se desea dar un intervalo de confianza de 98% con un margen de error de 1 milla por galón, ¿cuántos automóviles deberán usarse? Suponga que por pruebas anteriores se sabe que la desviación estándar del rendimiento es 2.6 millas por galón.

Ejercicio 6.4. Obtener los dos intervalos de confianza unilaterales del $100(1-\alpha)\%$ para θ, bajo el supuesto de que la varianza σ^2 es conocida.

6.1.2 Varianza desconocida

Ejercicio 6.5. Para una muestra aleatoria de una población Normal de media $\mu = \theta$ y varianza σ^2 desconocida, demuestre que,

$$\frac{\bar{X} - \theta}{S/\sqrt{n}} = \frac{\sqrt{n}(\bar{X} - \theta)}{S}$$

es una variable pivote para θ.

Ejercicio 6.6. A partir de la variable pivote del anterior ejercicio obtenga los respectivos intervalos de confianza bilateral de longitud mínima y unilaterales del $100(1-\alpha)\%$ para θ.

- ¿Cuál sería la fórmula para el margen de error asociado al intervalo bilateral de longitud mínima?
- ¿Qué problema se presenta al tratar de obtener una fórmula para el tamaño de muestra asociado al intervalo bilateral de longitud mínima?

Ejercicio 6.7. (*R. Walpole, R. Myers, S. Myers, K. Ye - Probabilidad y Estadística para Ingeniería y Ciencias. 9na Edición, Pearson (2012). Capítulo 9. Ejercicio 11.*) Una máquina produce piezas metálicas de forma cilíndrica. Se toma una muestra de las piezas y los diámetros son 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01 y 1.03 centímetros.

- ¿Cuál es la estimación puntual de la media poblacional?
- ¿Cuál es la estimación puntual de la desviación estándar poblacional?
- Con 95% de confianza, ¿cuál es el margen de error para la estimación de la media poblacional?
- ¿Cuál es el intervalo de confianza de 95% para la media poblacional?

6.2 Varianza

6.2.1 Media conocida

Para una muestra aleatoria de una población Normal de media μ y varianza $\sigma^2 = \theta$, bajo el supuesto de que la media μ es conocida, sabemos que,

$$\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\theta} \sim \chi^2_n$$

es una variable pivote para θ.

Ejercicio 6.8. A partir de la anterior variable pivote obtenga intervalos de confianza bilaterales del $100(1-\alpha)\%$ para la varianza σ^2 y para la desviación estándar σ.

160

⚠ ¿Qué condición se debe cumplir para que el intervalo de confianza bilateral de la desviación estándar sea de longitud mínima? ¿es la misma condición que se debe cumplir para que el intervalo de confianza bilateral de la varianza sea de longitud mínima?

Ejercicio 6.9. A partir de la variable pivote dada, obtenga los respectivos intervalos de confianza unilaterales del $100(1-\alpha)\%$ para la varianza σ^2 y para la desviación estándar σ.

6.2.2 Media desconocida

Para una muestra aleatoria de una población Normal de media μ y varianza θ, bajo el supuesto de que la media μ es desconocida,

$$\frac{(n-1)S^2}{\theta} \sim \chi^2_{n-1}$$

es una variable pivote para θ.

Ejercicio 6.10. A partir de la anterior variable pivote obtenga los respectivos intervalos de confianza bilateral y unilaterales del $100(1-\alpha)\%$ para θ.

Ejercicio 6.11. (*R. Walpole, R. Myers, S. Myers, K. Ye - Probabilidad y Estadística para Ingeniería y Ciencias. 9na Edición, Pearson (2012). Capítulo 9. Ejercicio 71.*) Un fabricante de baterías para automóvil afirma que sus baterías durarán, en promedio, 3 años con una varianza de 1 año. Suponga que 5 de estas baterías tienen duraciones de 1.9, 2.4, 3.0, 3.5 y 4.2 años y con base en esto construya un intervalo de confianza del 95% para la desviación estándar poblacional, después decida si es válida la afirmación del fabricante de que dicha desviación estándar poblacional es igual a 1. Suponga que la población de duraciones de las baterías se distribuye de forma aproximadamente normal.

6.3 Diferencia de medias

6.3.1 Muestras pareadas

Cuando las variables aleatorias X, Y representan variables medidas en las mismas unidades y que cuantifican el mismo aspecto de la unidad estadística sólo que en circunstancias distintas, y cuando $D_i = X_i - Y_i$ tiene sentido, la muestra aleatoria $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ se denomina **muestra pareada o dependiente**.

Sea $D_1, \dots, D_n$ una muestra aleatoria de una población Normal de media θ ($\mu_D = \mu_X - \mu_Y = \theta$) y varianza σ_D^2 ($\sigma_D^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y$), bajo el supuesto de varianza desconocida,

$$\frac{\sqrt{n}(\bar{D} - \theta)}{S_D} \sim t_{n-1}$$

es una variable pivote para θ.

Ejercicio 6.12. A partir de la anterior variable pivote obtenga los respectivos intervalos de confianza bilateral y unilaterales, de longitud mínima, del $100(1-\alpha)\%$ para θ.

- ¿Cuál sería la fórmula para el margen de error asociado al intervalo bilateral?
- ¿Qué problema se presenta al tratar de obtener una fórmula para el tamaño de muestra asociado al intervalo bilateral?

Ejercicio 6.13. (*D. Anderson, D. Sweeney, T. Williams, J. Camm - Estadística para Negocios y Economía. 11ra Edición, Cengage Learning (2012). Capítulo 10. Ejercicio 27.*) Un fabricante produce dos modelos de una lijadora automática, uno de lujo y otro estándar, diseñado para uso doméstico. Los precios de venta de una muestra de distribuidores minoristas se presentan a continuación.

Minorista	1	2	3	4	5	6	7
Lujo	39	39	45	38	40	39	35
Estándar	27	28	35	30	30	34	29

¿Cuál es el intervalo de 95% de confianza para la diferencia entre la media de los precios de ambos modelos?

6.3.2 Muestras independientes

6.3.2.1 Varianzas conocidas

Sean $X_1, \dots, X_n$ y $Y_1, \dots, Y_m$ muestras aleatorias independientes de poblaciones con distribución Normal de medias μ_X y μ_Y y varianzas conocidas σ_X^2 y σ_Y^2, respectivamente,

$$\frac{\left(\bar{X} - \bar{Y}\right) - \theta}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim \mathcal{N}(0, 1)$$

es una variable pivote para $\mu_X - \mu_Y = \theta$.

Ejercicio 6.14. A partir de la anterior variable pivote obtenga los respectivos intervalos de confianza bilateral y unilaterales, de longitud mínima, del $100(1-\alpha)\%$ para θ.

- ¿Cuál sería la fórmula para el margen de error asociado al intervalo bilateral?
- ¿Suponga que $n = m$ y obtenga una fórmula para el tamaño de muestra $n = m$ asociado al intervalo bilateral?

Ejercicio 6.15. (*D. Anderson, D. Sweeney, T. Williams, J. Camm - Estadística para Negocios y Economía. 11ra Edición, Cengage Learning (2012). Capítulo 10. Ejercicio 5.*) Se esperaba que el Día de San Valentín el gasto promedio fuera de \$100.89 (USA Today, 13 de febrero de 2006). ¿Hay diferencia en las cantidades que desembolsan los hombres y las mujeres? El gasto promedio en una muestra de 40 hombres fue de \$135.67 y en una muestra de 30 mujeres fue de \$68.64. Por estudios anteriores se sabe que la desviación estándar poblacional en el consumo de los hombres es \$35 y en el de las mujeres es \$20.

- ¿Cuál es la estimación puntual de la diferencia entre el gasto medio poblacional de los hombres y el gasto medio poblacional de las mujeres?
- Con 99% de confianza, ¿cuál es el margen de error?
- Elabore un intervalo de confianza de 99% para la diferencia entre las dos medias poblacionales.

6.3.2.2 Varianzas desconocidas, suponiendo igualdad

Sean $X_1, \ldots, X_n$ y $Y_1, \ldots, Y_m$ muestras aleatorias independientes de poblaciones con distribución Normal de medias μ_X y μ_Y, respectivamente, y varianzas desconocidas pero iguales $\sigma_X^2 = \sigma_Y^2$,

$$\frac{\left(\bar{X} - \bar{Y}\right) - \theta}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2},$$

con $S_p = \sqrt{\frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}}$, es una variable pivote para $\mu_X - \mu_Y = \theta$.

Ejercicio 6.16. A partir de la anterior variable pivote obtenga los respectivos intervalos de confianza bilateral y unilaterales, de longitud mínima, del $100(1-\alpha)\%$ para θ.

- ¿Cuál sería la fórmula para el margen de error asociado al intervalo bilateral?

Ejercicio 6.17. (*R. Walpole, R. Myers, S. Myers, K. Ye - Probabilidad y Estadística para Ingeniería y Ciencias. 9na Edición, Pearson (2012). Capítulo 9. Ejercicio 46.*) Los siguientes datos representan el tiempo de duración de películas producidas por dos empresas cinematográficas.

Empresa							
I	103	94	110	87	98		
II	97	82	123	92	175	88	118

Calcule un intervalo de confianza del 90% para la diferencia entre la duración promedio de las películas que producen las dos empresas. Suponga que las diferencias en la duración se distribuyen de forma aproximadamente normal y que tienen varianzas iguales.

6.3.2.3 Varianzas desconocidas, sin suponer igualdad

Sean $X_1, \dots, X_n$ y $Y_1, \dots, Y_m$ muestras aleatorias independientes de poblaciones con distribución Normal de medias μ_X y μ_Y y varianzas desconocidas y distintas σ_X^2 y σ_Y^2, respectivamente,

$$\frac{\left(\bar{X} - \bar{Y}\right) - \theta}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \sim t_v,$$

con $v \approx \frac{\left(s_X^2/n + s_Y^2/m\right)^2}{\frac{\left(s_X^2/n\right)^2}{n-1} + \frac{\left(s_Y^2/m\right)^2}{m-1}}$, es una variable pivote para $\mu_X - \mu_Y = \theta$.

Ejercicio 6.18. A partir de la anterior variable pivote obtenga los respectivos intervalos de confianza bilateral y unilaterales, de longitud mínima, del $100(1-\alpha)\%$ para θ.

- ¿Cuál sería la fórmula para el margen de error asociado al intervalo bilateral?

Ejercicio 6.19. (*R. Walpole, R. Myers, S. Myers, K. Ye - Probabilidad y Estadística para Ingeniería y Ciencias. 9na Edición, Pearson (2012). Capítulo 9. Ejercicio 46.*) Los siguientes datos representan el tiempo de duración de películas producidas por dos empresas cinematográficas.

Empresa							
I	103	94	110	87	98		
II	97	82	123	92	175	88	118

Calcule un intervalo de confianza del 90% para la diferencia entre la duración promedio de las películas que producen las dos empresas. Suponga que las diferencias en la duración se distribuyen de forma aproximadamente normal y que tienen varianzas distintas.

6.4 Cociente de varianzas

6.4.1 Muestras independientes

6.4.1.1 Medias conocidas

Para $X_1, \ldots, X_n$ y $Y_1, \ldots, Y_m$ muestras aleatorias independientes de poblaciones con distribución Normal de medias conocidas μ_X y μ_Y y varianzas σ_X^2 y σ_Y^2, respectivamente,

$$\frac{\sum_{i=1}^{m} \left(Y_i - \mu_Y\right)^2 / m}{\sum_{i=1}^{n} \left(X_i - \mu_X\right)^2 / n} \theta \sim F_{m,n}$$

es una variable pivote para $\sigma_X^2 / \sigma_Y^2 = \left(\sigma_X / \sigma_Y\right)^2 = \theta$.

Ejercicio 6.20. A partir de la anterior variable pivote obtenga un intervalo de confianza bilateral del $100(1-\alpha)\%$ para el cociente de varianzas θ.

Solución Ejercicio 6.20

Transformando la correspondiente variable pivote tenemos que,

$$P\left[a < \frac{\sum_{i=1}^{m}\left(Y_i-\mu_Y\right)^2/m}{\sum_{i=1}^{n}\left(X_i-\mu_X\right)^2/n}\theta < b\right] = 1-\alpha$$

Ejercicio 6.21. A partir de la variable pivote dada, obtenga los respectivos intervalos de confianza unilaterales de $100(1-\alpha)\%$ para el cociente de varianzas θ.

6.4.1.2 Medias desconocidas

Sean $X_1, \dots, X_n$ y $Y_1, \dots, Y_m$ muestras aleatorias independientes de poblaciones con distribución Normal de medias desconocidas μ_X y μ_Y y varianzas σ_X^2 y σ_Y^2, respectivamente.

$$\frac{\sum_{i=1}^{m}\left(Y_i-\bar{Y}\right)^2/(m-1)}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2/(n-1)}\theta=\frac{S_Y^2}{S_X^2}\theta\sim F_{m-1,n-1}$$

es una variable pivote para $\sigma_X^2/\sigma_Y^2=\left(\sigma_X/\sigma_Y\right)^2=\theta$.

Ejercicio 6.22. A partir de la anterior variable pivote obtenga los respectivos intervalos de confianza bilateral y unilaterales del $100(1-\alpha)\%$ para el cociente de varianzas σ_X^2/σ_Y^2.

- ¿Cuáles serían los intervalos de confianza bilateral y unilaterales del $100(1-\alpha)\%$ para el cociente de desviaciones estándar σ_X/σ_Y?

Ejercicio 6.23. (*R. Walpole, R. Myers, S. Myers, K. Ye - Probabilidad y Estadística para Ingeniería y Ciencias. 9na Edición, Pearson (2012). Capítulo 9. Ejercicio 79.*) Construya un intervalo de confianza del 90% para σ_1/σ_2 en el Ejercicio 6.19.

- ¿De acuerdo al intervalo de confianza resultante, al construir un intervalo de confianza para la diferencia de medias, lo correcto sería suponer que las varianzas / desviaciones estándar poblacionales son iguales o que no lo son?

7 Bajo otras distribuciones

En esta sección se hará una revisión de algunos resultados relacionadas con la estimación por intervalo, para muestras aleatorias provenientes de una población, para la cual se supone una cierta distribución dada, diferente a la distribución normal.

En términos generales, para aplicar el método de la variable pivote, se necesita una variable que sea función de la muestra y del parámetro, en donde se pueda despejar $\mathrm{g}(\theta)$, y cuya distribución debe permitir la obtención de los cuantiles necesarios (obviamente dicha distribución no puede depender de θ).

7.1 Distribución asintótica (opción 1)

Bajo un caso regular de estimación (condiciones de regularidad), si el estimador T es insesgado para $\mathrm{g}(\theta)$ y tiene una varianza que coincide con la cota de Cramer - Rao, entonces se tiene que,

$$T \xrightarrow[n\to\infty]{d} W \sim \mathcal{N}\left(\mathrm{g}(\theta), \frac{\left(\frac{\partial}{\partial\theta}\mathrm{g}(\theta)\right)^2}{n\,\mathcal{I}(\theta)}\right).$$

donde $I(\theta)$ es la información de Fisher y n es el tamaño de muestra. Lo que quiere decir que para un tamaño de muestra suficientemente grande,

$$\frac{T - \mathrm{g}(\theta)}{\left|\frac{\partial}{\partial\theta}\mathrm{g}(\theta)\right| \sqrt{\frac{1}{n\mathcal{I}(\theta)}}} \sim \mathcal{N}(0,1)$$

es una variable pivote **asintótica / aproximada** para $\mathrm{g}(\theta)$.

Ejercicio 7.1. A partir de la anterior variable pivote asintótica y la muestra de una población con $X \sim Exp(\theta)$ $\left(\mathrm{f}(x,\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}\,\mathrm{I}_{(0,\infty)}(x)\right)$, obtenga un intervalo de confianza asintótico / aproximado bilateral del $100(1-\alpha)\%$ para θ.

Solución Ejercicio 7.1

Para X se tiene que $E[X] = \theta$, $Var[X] = \theta^2$ y $I(\theta) = \frac{1}{\theta^2}$. Por otra parte, para $\bar{X}$ se tiene que $E\left[\bar{X}\right] = \theta$ (insesgado) y $Var\left[\bar{X}\right] = \frac{\theta^2}{n} = \frac{1}{n\mathcal{I}(\theta)}$ (cota de Cramer - Rao para θ).

Transformando la correspondiente variable pivote asintótica se tiene que,

$$P\left[a < \frac{\sqrt{n}\left(T - \theta\right)}{\sqrt{\frac{1}{I(\theta)}}} < b\right] = 1 - \alpha$$

$$P\left[a < \frac{\sqrt{n}\left(\bar{X} - \theta\right)}{\theta} < b\right] = 1 - \alpha$$

$$P\left[a < \sqrt{n}\left(\frac{\bar{X}}{\theta} - 1\right) < b\right] = 1 - \alpha$$

$$P\left[a\frac{1}{\sqrt{n}} < \frac{\bar{X}}{\theta} - 1 < b\frac{1}{\sqrt{n}}\right] = 1 - \alpha$$

$$P\left[1 + a\frac{1}{\sqrt{n}} < \frac{\bar{X}}{\theta} < 1 + b\frac{1}{\sqrt{n}}\right] = 1 - \alpha$$

$$P\left[\frac{\sqrt{n} + a}{\sqrt{n}\bar{X}} < \frac{1}{\theta} < \frac{\sqrt{n} + b}{\sqrt{n}\bar{X}}\right] = 1 - \alpha$$

$$P\left[\frac{\sqrt{n}\bar{X}}{\sqrt{n} + b} < \theta < \frac{\sqrt{n}\bar{X}}{\sqrt{n} + a}\right] = 1 - \alpha$$

y si por facilidad tomamos $b = -a = z_{(\alpha/2)}$, entonces,

$$P\left[\frac{\sqrt{n}\bar{X}}{\sqrt{n} + z_{(\alpha/2)}} < \theta < \frac{\sqrt{n}\bar{X}}{\sqrt{n} - z_{(\alpha/2)}}\right] = 1 - \alpha$$

Por lo tanto,

$$\left(\frac{\bar{X}}{1 + \frac{z_{(\alpha/2)}}{\sqrt{n}}}, \frac{\bar{X}}{1 - \frac{z_{(\alpha/2)}}{\sqrt{n}}}\right)$$

es un intervalo de confianza bilateral del $100(1 - \alpha)\%$ para θ.

Con el objetivo de facilitar las transformaciones requeridas para llegar al intervalo de confianza, y teniendo en cuenta que $T \to \mathrm{g}(\theta)$ cuando $n \to \infty$, cada $\mathrm{g}(\theta)$ en el denominador podría ser reemplazado por el estimador T (o de manera equivalente cada θ puede ser reemplazado por

$\mathrm{g}^{-1}(T)$). Es decir, para un tamaño de muestra suficientemente grande, si

$$W = \left[\left|\frac{\partial}{\partial\theta}\mathrm{g}(\theta)\right|\sqrt{\frac{1}{n\mathcal{I}(\theta)}}\right]\Bigg|_{\mathrm{g}(\theta)=T \text{ o } \theta=\mathrm{g}^{-1}(T)}$$

entonces

$$\frac{T-\mathrm{g}(\theta)}{W} \sim \mathcal{N}(0,1)$$

también es una variable pivote asintótica / aproximada para $\mathrm{g}(\theta)$. Transformando dicha variable pivote asintótica,

$$\begin{aligned}
P\left[a < \frac{T-\mathrm{g}(\theta)}{W} < b\right] &= 1-\alpha \\
P\left[aW < T-\mathrm{g}(\theta) < bW\right] &= 1-\alpha \\
P\left[-T+aW < -\mathrm{g}(\theta) < -T+bW\right] &= 1-\alpha \\
P\left[T-bW < \mathrm{g}(\theta) < T+(-a)W\right] &= 1-\alpha
\end{aligned}$$

de donde se tiene que,

$$\Big(T-bW\,,\,T+(-a)W\Big)$$

con

$$W = \left[\left|\frac{\partial}{\partial\theta}\mathrm{g}(\theta)\right|\sqrt{\frac{1}{n\mathcal{I}(\theta)}}\right]\Bigg|_{\mathrm{g}(\theta)=T \text{ o } \theta=\mathrm{g}^{-1}(T)}$$

es un intervalo de confianza asintótico / aproximado bilateral del $100(1-\alpha)\%$ para $\mathrm{g}(\theta)$, y de longitud mínima cuando $b=-a=z_{(\alpha/2)}$ (**de longitud mínima para** $\mathrm{g}(\theta)$).

Ejercicio 7.2. A partir del anterior resultado, obtenga los respectivos intervalos de confianza bilateral y unilaterales del $100(1-\alpha)\%$ para θ $\left(\mathrm{f}(x,\theta)=\frac{1}{\theta}e^{-\frac{x}{\theta}}\,\mathrm{I}_{(0,\infty)}(x)\right)$.

- ¿Cuál sería la formula para el margen de error asociado al intervalo de confianza bilateral?
- ¿Cuál sería la formula para el tamaño de muestra asociado al intervalo de confianza bilateral?

Ejercicio 7.3. Usando cada una de las dos variables pivote asintóticas / aproximadas para θ, obtener los intervalos de confianza aproximados bilateral y unilaterales del $100(1-\alpha)\%$ para θ, asociados a una muestra aleatoria de una población con,

- $X \sim Weibull(\theta, k \text{ conocido})$ es decir $\mathrm{f}(x,\theta)=\frac{k}{\theta}\left(\frac{x}{\theta}\right)^{k-1}\exp\left(-\left(\frac{x}{\theta}\right)^{k}\right)\,\mathrm{I}_{(0,\infty)}(x)$, además tenga en cuenta que si $X \sim Weibull(\theta,k)$ entonces $E\left[X^r\right]=\theta^r\,\Gamma\left(1+\frac{r}{k}\right)$ y por ende $E[T]=E\left[\frac{1}{n}\sum_{i=1}^{n}X_i^k\right]=\theta^k=g(\theta)$.

- $X \sim Pareto(k \text{ conocido}, \theta)$ es decir $\mathrm{f}(x,\theta) = \theta k^{\theta} x^{-(\theta+1)} \,\mathrm{I}_{(k,\infty)}(x)$, además tenga en cuenta que si $X \sim Pareto(k,\theta)$ entonces $\log\left(\frac{X}{k}\right) \sim Exp(1/\theta)$ y por ende $E[T] = E\left[\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{X_i}{k}\right)\right] = \frac{1}{\theta} = g(\theta)$.
- $X \sim Pois(\lambda = \theta)$.
- $X \sim Bern(p = \theta)$ (*intervalos de confianza para la proporción*).

En cada uno de los anteriores casos, obtenga las fórmulas correspondientes al margen de error y al tamaño de muestra de los respectivos intervalos bilaterales.

Ejercicio 7.4. (*D. Anderson, D. Sweeney, T. Williams, J. Camm - Estadística para Negocios y Economía. 11ra Edición, Cengage Learning (2012). Capítulo 8. Ejercicio 54.*) En un estudio de USA Today/CNN/Gallup realizado con 369 padres que trabajan, se encontró que 200 consideran que pasan muy poco tiempo con sus hijos debido a sus compromisos laborales.

- Proporcione una estimación puntual de la proporción poblacional de padres que trabajan y piensan que pasan muy poco tiempo con sus hijos debido a sus compromisos laborales.
- ¿Cuál es el error estándar de la estimación puntual anterior?
- ¿Cuál es el margen de error para 95% de confianza?
- ¿Cuál es el intervalo de confianza de 95% para la proporción poblacional de padres que trabajan y piensan que pasan muy poco tiempo con sus hijos debido a sus compromisos ocupacionales?

Ejercicio 7.5. (*D. Anderson, D. Sweeney, T. Williams, J. Camm - Estadística para Negocios y Economía. 11ra Edición, Cengage Learning (2012). Capítulo 8. Ejercicio 58.*) Una firma de tarjetas de crédito de un conocido banco desea estimar la proporción de tarjetahabientes que al final del mes tienen un saldo distinto de cero que ocasiona cargos. Suponga que el margen de error deseado es 0.03 con 98% de confianza.

- ¿De qué tamaño deberá tomarse la muestra si se cree que 70% de los tarjetahabientes de la firma tienen un saldo distinto de cero al final del mes?
- ¿De qué tamaño deberá tomarse la muestra si no se puede especificar ningún valor planeado para la proporción?

Ejercicio 7.6. (*Intervalos de confianza para la diferencia de proporciones de dos poblaciones independientes*) Para $X_1, \ldots, X_n$ y $Y_1, \ldots, Y_m$ muestras aleatorias independientes de poblaciones con distribución Bernoulli de parámetros (proporciones) p_X y p_Y, respectivamente,

$$\frac{\left(\hat{P}_X - \hat{P}_Y\right) - \theta}{\sqrt{\frac{\hat{P}_X\left(1-\hat{P}_X\right)}{n} + \frac{\hat{P}_Y\left(1-\hat{P}_Y\right)}{m}}} \sim \mathcal{N}(0,1)$$

es una variable pivote para $\theta = p_X - p_Y$.

- A partir de la anterior variable pivote obtenga los respectivos intervalos de confianza bilateral y unilaterales, de longitud mínima, del $100(1-\alpha)\%$ para θ.
- ¿Cuál sería la formula para el margen de error asociado al intervalo bilateral?

Ejercicio 7.7. (*D. Anderson, D. Sweeney, T. Williams, J. Camm - Estadística para Negocios y Economía. 11ra Edición, Cengage Learning (2012). Capítulo 10. Ejercicio 30.*) En una encuesta de BusinessWeek/Harris se pidió a los ejecutivos de empresas grandes su opinión acerca de cómo veían las perspectivas económicas para el futuro. Una de las preguntas era: ¿Piensa usted que en los próximos 12 meses aumentará en su empresa el número de empleados de tiempo completo? En la encuesta actual, 220 de 400 ejecutivos respondieron Sí, mientras que en la realizada el año anterior, 192 de 400 respondieron en el mismo sentido.

- Calcule la proporción muestral de los que respondieron Sí en la encuesta actual y la proporción de los que respondieron Sí en la encuesta anterior.
- ¿Cuál es la estimación puntual de la diferencia entre las proporciones de las dos poblaciones? ¿Qué indica tal estimación? ¿Cuál es el error estándar de dicha estimación puntual?
- Encuentre un intervalo de 95% de confianza para estimar la diferencia entre las proporciones en estas dos encuestas. ¿Cuál es su interpretación de la estimación por intervalo?

7.2 Parámetro localización o escala (opción 2)

El rol del parámetro puede ayudar a identificar la variable pivote junto con su respectiva distribución.

Sea $\left\{\mathrm{f}_X(x,\theta)\,\middle|\,\theta \in \Theta \subseteq \mathbb{R}^k\right\}$ una familia de densidades. El componente θ_j de θ se denomina **componente o parámetro de localización**, si y sólo si la distribución de

$$X-\theta_j \quad \text{o} \quad X+\theta_j$$

no depende de θ_j.

Sea $\left\{\mathrm{f}_X(x,\theta)\,\middle|\,\theta \in \Theta \subseteq \mathbb{R}^k\right\}$ una familia de densidades. El componente θ_j de θ se denomina **componente o parámetro de escala**, si y sólo si la distribución de

$$\theta_j X \quad \text{o} \quad \frac{X}{\theta_j}$$

no depende de θ_j.

De lo anterior, si θ es un parámetro de localización, entonces

$$\sum_{i=1}^{n}(X_i-\theta) \quad \text{o} \quad \sum_{i=1}^{n}(X_i+\theta)$$

es una variable pivote para θ, y si θ es un parámetro de escala, entonces

$$\sum_{i=1}^{n} \theta X_i \quad \text{o} \quad \sum_{i=1}^{n} \frac{X_i}{\theta}$$

es una variable pivote para θ.

Además, si θ_1 es un parámetro de localización y si $T^{(1)}$ es un estimador máximo verosímil de θ_1, entonces

$$T^{(1)} - \theta_1 \quad \text{o} \quad T^{(1)} + \theta_1$$

es una variable pivote para θ_1; si θ_2 es un parámetro de escala y si $T^{(2)}$ es un estimador máximo verosímil de θ_2, entonces

$$\theta_2 T^{(2)} \quad \text{o} \quad \frac{T^{(2)}}{\theta_2}$$

es una variable pivote para θ_2; y

$$\frac{T^{(1)} - \theta_1}{T^{(2)}}$$

es una variable pivote para θ_1, si esta no depende de los demás componentes de θ, o si éstos son considerados conocidos.

Ejercicio 7.8. A partir de los resultados anteriormente expuestos, dada una muestra de una población con $X \sim \mathcal{U}(0, \theta)$, obtenga un intervalo de confianza bilateral del $100(1-\alpha)\%$ para θ.

Solución Ejercicio 7.8

Ejercicio 7.9. A partir de los resultados expuestos en esta sección, obtener una variable pivote y los respectivos intervalos de confianza bilateral y unilaterales del $100(1-\alpha)\%$ para θ, asociados a una muestra aleatoria de una población con:

- $X \sim Pareto(\theta, k \text{ conocido})$ es decir $\mathrm{f}(x,\theta) = k\theta^k x^{-(k+1)}\, \mathrm{I}_{(\theta,\infty)}(x)$.
- $X \sim Exp(\theta)$ es decir $\mathrm{f}(x,\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}\, \mathrm{I}_{(0,\infty)}(x)$.
- $X \sim Weibull(\theta, k \text{ conocido})$ es decir $\mathrm{f}(x,\theta) = \frac{k}{\theta}\left(\frac{x}{\theta}\right)^{k-1} \exp\left(-\left(\frac{x}{\theta}\right)^k\right) \mathrm{I}_{(0,\infty)}(x)$.

Ejercicio 7.10. Para todas las distribuciones que conoce, incluidas todas las mencionadas en el curso, identificar qué parámetros son de localización, qué parámetros son de escala y qué parámetros no son ni de localización ni de escala (por ejemplo, el parámetro λ de la distribución Poisson).

7.3 Probability Inverse Transform (opción 3)

Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con función de distribución acumulativa continua $\mathrm{F}_X(x,\theta)$. Se tiene que,

$$\mathrm{F}_X(X_i,\theta) = U_i \sim \mathcal{U}(0,1)$$

y también que,

$$1 - \mathrm{F}_X(X_i,\theta) = 1 - U_i \sim \mathcal{U}(0,1)$$

Por lo tanto,

$$\mathrm{F}_Y^{-1}(U_i) = \mathrm{F}_Y^{-1}(\mathrm{F}_X(X_i,\theta)) = Y_i \sim \mathcal{Y}$$

y

$$\mathrm{F}_Y^{-1}(1-U_i) = \mathrm{F}_Y^{-1}(1-\mathrm{F}_X(X_i,\theta)) = Y_i \sim \mathcal{Y}$$

para una distribución $\mathcal{Y}$ con función de distribución acumulativa continua $\mathrm{F}_Y(x)$, que no depende de θ. Lo que quiere decir que $\sum_{i=1}^n Y_i$ sería una variable pivote para θ.

La distribución exponencial sería una muy buena elección para las Y_i, ya que si se toma,

$$\mathrm{F}_Y(x) = 1 - e^{-x},$$

entonces,

$$\mathrm{F}_Y^{-1}(u) = -\log(1-u),$$

Es decir que,

$$-\log(1-U) \sim Exp(1)$$

y

$$-\log(U) \sim Exp(1),$$

Por lo tanto,

$$\sum_{i=1}^{n} -\log(1-U_i) = -\sum_{i=1}^{n} \log(1-\mathrm{F}_X(X_i,\theta)) \sim Gamma(n,1)$$

y

$$\sum_{i=1}^{n} -\log(U_i) = -\sum_{i=1}^{n} \log(\mathrm{F}_X(X_i,\theta)) \sim Gamma(n,1)$$

Lo que quiere decir que $-\sum_{i=1}^{n} \log(1-\mathrm{F}_X(X_i,\theta))$ y $-\sum_{i=1}^{n} \log(\mathrm{F}_X(X_i,\theta))$ son variables pivote para θ.

Teniendo en cuenta que $2Y \sim Exp(2)$ y $Gamma(v/2,2) \equiv \chi^2_v$, se tiene que,

$$-\sum_{i=1}^{n} 2\log(\mathrm{F}_X(X_i,\theta)) \sim Gamma(n,2) \equiv \chi^2_{2n}$$

y

$$-\sum_{i=1}^{n} 2\log(1-\mathrm{F}_X(X_i,\theta)) \sim Gamma(n,2) \equiv \chi^2_{2n}$$

son otras variables pivote para θ.

Es decir que para toda distribución continua, asociada a la muestra $X_1,\ldots,X_n$, existen dos variables pivote con distribución chi-cuadrado (distribución para la que se tienen tablas estadísticas). De lo cual concluimos que, si la distribución de la población tiene asociada una inversa explícita, con respecto a θ, de su función de distribución acumulativa continua (para poder “despejar” θ), entonces es relativamente fácil obtener un intervalo de confianza a partir de una de dichas variables pivote.

Ejercicio 7.11. A partir de la anterior variable pivote y de la muestra de una población con $X \sim Exp(\theta)$ $\left(\mathrm{f}(x,\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}\mathrm{I}_{(0,\infty)}(x)\right)$, obtenga un intervalo de confianza bilateral del $100(1-\alpha)\%$ para θ.

Solución Ejercicio 7.11

En este caso,

$$
\begin{aligned}
-\sum_{i=1}^{n} 2\log\left(1-\mathrm{F}_X(X_i,\theta)\right) &= -\sum_{i=1}^{n} 2\log\left(1-\left(1-\exp\left(-\frac{X_i}{\theta}\right)\right)\right) \\
&= -\sum_{i=1}^{n} 2\log\left(\exp\left(-\frac{X_i}{\theta}\right)\right) \\
&= \sum_{i=1}^{n} 2\frac{X_i}{\theta} \\
&= \frac{2\sum_{i=1}^{n} X_i}{\theta} \\
&= \frac{2n\bar{X}}{\theta} \sim \chi^2_{2n}
\end{aligned}
$$

Transformando esa variable pivote (“despejando” θ) tenemos que,

$$
\begin{aligned}
P\left[a < \frac{2n\bar{X}}{\theta} < b\right] &= 1-\alpha \\
P\left[\frac{1}{b} < \frac{\theta}{2n\bar{X}} < \frac{1}{a}\right] &= 1-\alpha \\
P\left[\frac{2n\bar{X}}{b} < \theta < \frac{2n\bar{X}}{a}\right] &= 1-\alpha
\end{aligned}
$$

Por lo tanto,

$$
\left(\frac{2n\bar{X}}{b}, \frac{2n\bar{X}}{a}\right)
$$

es un intervalo de confianza bilateral del $100(1-\alpha)\%$ para θ, donde a y b son cuantiles tales que $\mathrm{F}_{\chi^2_{2n}}(b) - \mathrm{F}_{\chi^2_{2n}}(a) = 1-\alpha$, y $\mathrm{F}_{\chi^2_{2n}}(.)$ es la función de distribución acumulativa chi-cuadrado con $2n$ grados de libertad. La condición de longitud mínima es $b^2\mathrm{f}_{\chi^2_{2n}}(b) - a^2\mathrm{f}_{\chi^2_{2n}}(a) = 0$ (es necesario resolver numéricamente). Si se toma $a = \chi^2_{(1-\alpha/2,2n)}$ y $b = \chi^2_{(\alpha/2,2n)}$, entonces el intervalo de confianza queda así,

$$
\left(\frac{2n\bar{X}}{\chi^2_{(\alpha/2,2n)}}, \frac{2n\bar{X}}{\chi^2_{(1-\alpha/2,2n)}}\right)
$$

Ejercicio 7.12. A partir de lo anterior, obtener una variable pivote y los respectivos intervalos de confianza bilateral y unilaterales del $100(1-\alpha)\%$ para θ, asociados a una muestra aleatoria de una población con:

- $X \sim \mathcal{U}(0, \theta)$.
- $X \sim Pareto(\theta, k$ conocido$)$, es decir $\mathrm{f}(x, \theta) = k\theta^k x^{-(k+1)}\, \mathrm{I}_{(\theta,\infty)}(x)$.

Ejercicio 7.13. Para una muestra aleatoria de una población con $X \sim Bern(\theta)$, leer acerca de otros intervalos de confianza para θ que hayan sido propuestos (*intervalos de confianza para la proporción*). Por ejemplo, ver Wikipedia: Binomial proportion confidence interval.

Ejercicio 7.14. Para una muestra aleatoria de una población con $X \sim Pois(\theta)$, leer acerca de otros intervalos de confianza para θ que hayan sido propuestos. Por ejemplo, ver Patil V.V., Kulkarni H.V. (2012) Comparison of confidence intervals for the Poisson mean: some new aspects. REVSTAT – Statistical Journal. Volume 10, Number 2, June 2012, 211–227.

Parte V

Juzgamiento de hipótesis

8 Aspectos generales

En esta sección se hará una revisión de algunos aspectos generales relacionados con el juzgamiento (la prueba) de hipótesis, específicamente en cuanto a conceptos y definiciones iniciales.

8.1 Conceptos y definiciones iniciales

Una **hipótesis estadística** es una aseveración o conjetura acerca de la distribución de una población, afirmación que generalmente queda determinada por la pertenencia de los parámetros a un subconjunto dado del espacio de parámetros.

El **juzgamiento de una hipótesis estadística** es un proceso que culmina con una decisión de rechazar o no rechazar una hipótesis, con base en la información de una muestra aleatoria $X_1, \ldots, X_n$ (de una población con función de densidad $\mathrm{f}_X(x, \theta)$, es decir, el modelo probabilístico que se ha asumido para la población).

8.1.1 Hipótesis

La hipótesis sobre la cual se estructura el proceso de juzgamiento se denomina **hipótesis nula**, se denota H_0 y se enuncia como,

$$H_0 : \theta \in \Theta_0 \subset \Theta.$$

La hipótesis elegida como contraste a la hipótesis nula se denomina **hipótesis alterna**, se denota H_1 (o H_a) y se enuncia como,

$$H_1 : \theta \in \Theta_1 \subset \Theta,$$

donde $\Theta_0 \cap \Theta_1 = \emptyset$.

La díada de hipótesis nula y alterna constituye el **sistema de hipótesis** del proceso de juzgamiento de la hipótesis nula, sistema que se enuncia como,

$$H_0 : \theta \in \Theta_0 \text{frente a} H_1 : \theta \in \Theta_1$$

Una hipótesis se denomina **hipótesis simple** si con dicha aseveración queda plenamente especificada la distribución de la población. En caso contrario se denomina **hipótesis compuesta**.

8.1.2 Test y región crítica

El proceso de juzgamiento de la hipótesis nula conlleva un procedimiento, regla o norma, denominado **test**, que permite tomar la decisión a que haya lugar.

El test utilizado dentro del proceso de juzgamiento de la hipótesis nula H_0, tiene vinculado un subconjunto del espacio de las observaciones. Este subconjunto denotado por C_τ queda establecido por su respectivo test así:

$$\tau: \text{ Rechazar } H_0 \text{ si } (x_1, \dots, x_n) \in C_\tau$$

El conjunto C_τ se denomina **región crítica** o **región de rechazo** del test (para juzgar H_0) y el test así definido se denomina **test no aleatorizado**. El conjunto C_τ^C, complemento del conjunto C_τ, recibe el nombre de **región de aceptación** del test (para juzgar a H_0).

Ejemplo 8.1. (*D. Anderson, D. Sweeney, T. Williams, J. Camm - Estadística para Negocios y Economía. 11ra Edición, Cengage Learning (2012). Capítulo 9. Ejercicio 60.*) En una línea de producción, el peso promedio con que se llena cada recipiente es 16 onzas. Un llenado insuficiente ocasiona problemas serios y, cuando es detectado, es necesario que el operador detenga la línea de producción para reajustar el mecanismo de llenado. Con base en datos anteriores, se supone que la desviación estándar poblacional es 0.8 onzas. Cada hora, un inspector de control de calidad toma una muestra de 30 recipientes y decide si es necesario detener la producción y hacer un reajuste.

Se idealiza el llenado como una variable aleatoria y se adopta un modelo probabílistico para que determine su comportamiento. Por ejemplo, se puede suponer que el peso promedio con que se llena cada recipiente es una variable aleatoria normal con media θ y varianza conocida 0.8^2.

La hipótesis $H_0 : \theta = 16$ indica que el llenado está centrado en 16 onzas como se requiere. Además, claramente esta es una hipótesis simple. La afirmación de que el llenado es insuficiente se podría representar con la siguiente hipótesis compuesta $H : \theta < 16$. Estas dos afirmaciones se podrían utilizar dentro del siguiente sistema de hipótesis:

$$H_0 : \theta = 16 \text{frente a} H_1 : \theta < 16$$

el cual representa la situación descrita.

Supongamos que un "experto" propone el siguiente test para el juzgamiento de H_0 dentro del anterior sistema de hipótesis.

$$\tau: \text{ Rechazar } H_0 \text{ si } \bar{x} < 15.5$$

Lo que quiere decir que, según el "experto", se debe detener la producción si el promedio de la muestra que toma el inspector es menor a 15.5. Naturalmente, la producción no se detiene en caso contrario.

La región crítica asociada al anterior test sería,

$$C_\tau = \{(x_1, \dots, x_n) : \bar{x} < 15.5\}$$

8.1.3 Errores

Suponga que está en un juicio. Bajo la idea de que un acusado es inocente hasta que se demuestre lo contrario (el jurado debe encontrarlo culpable "más allá de la duda razonable") entonces esta situación se puede representar con el siguiente sistema de hipótesis:

$$\begin{aligned} H_0 &: \text{ El acusado es inocente} \\ &\text{frente a} \\ H_1 &: \text{ El acusado es culpable} \end{aligned}$$

La evidencia (la información en la muestra) tiene que ser suficiente para condenar al acusado (rechazar H_0). En caso contrario, se dice que no hay suficiente evidencia para declararlo culpable (no se rechaza H_0 a partir de lo observado en la muestra).

El jurado puede tomar una decisión ajustada al sistema legal y a las evidencias, y aún así podría no coincidir con la realidad de los hechos ocurridos, hechos que el jurado no tiene forma de conocer más allá que por medio de la evidencia que se le presenta. Es decir, siguiendo únicamente la evidencia, el jurado podría declarar inocente a alguien que en realidad es culpable y viceversa. Así mismo, en el juzgamiento de hipótesis estadísticas se puede errar, rechazando la hipótesis nula cuando, sin saberlo, en realidad es verdadera o no rechazándola cuando, sin saberlo, en realidad es falsa.

Dentro del proceso de juzgamiento de la hipótesis nula, se denomina **error de tipo I** a la decisión de rechazar H_0, siendo verdadera la hipótesis; asimismo, se designa como **error de tipo II** a la decisión de no rechazar la hipótesis nula siendo ella falsa.

	Rechazar H_0	No rechazar H_0
H_0 es cierta	**Error de tipo I**	Sin Error
H_0 es falsa	Sin Error	**Error de tipo II**

Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con función de densidad $\mathrm{f}_X(x, \theta)$ y sea además τ un test no aleatorizado para el juzgamiento de la hipótesis nula $H_0 : \theta \in \Theta_0$ definido como

$$\tau : \text{ Rechazar } H_0 \text{ si } (x_1, \ldots, x_n) \in C_\tau$$

La función

$$\varphi_\tau(\mathbf{x}) = \begin{cases} 1 & \text{si } \mathbf{x} = (x_1, \ldots, x_n) \in C_\tau \\ 0 & \text{si } \mathbf{x} \in C_\tau^C \end{cases}$$

se denomina **función crítica** del test no aleatorizado τ.

El **tamaño del test** τ, el **nivel del test** τ, el **tamaño de la región crítica** C_τ o la **probabilidad de error de tipo I** se refieren a lo mismo y suelen denotarse por α, el cual está definido como,

$$\alpha = \max_{\theta \in \Theta_0} P_\theta \left[\varphi_\tau(\mathbf{X}) = 1 \right].$$

La probabilidad de error tipo II se suele denotar β.

Ejercicio 8.1. Sea $X_1, \dots, X_n$ una muestra aleatoria de tamaño 10 de una población con distribución Bernoulli de parámetro θ. Para juzgar la hipótesis nula $H_0 : \theta \leq 0.75$ dentro del sistema de hipótesis,

$$\begin{gathered} H_0 : \theta \leq 0.75 \\ \text{frente a} \\ H_1 : \theta > 0.75 \end{gathered}$$

se propone el test,

$$\begin{aligned} \tau : \text{Rechazar } H_0 \text{ si } \sum_{i=1}^{10} x_i &\geq 9 \\ \frac{1}{10}\sum_{i=1}^{10} x_i &\geq \frac{9}{10} \\ \bar{x} &\geq 0.9 \\ \hat{p} &\geq 0.9 \end{aligned}$$

Calcular la probabilidad de error del tipo I (nivel del test).

Solución

Teniendo en cuenta que $\sum_{i=1}^{10} x_i \sim Bin(10, \theta)$ entonces,

$$\begin{aligned} P_\theta\left[\varphi_\tau(\mathbf{X}) = 1\right] &= P\left[\sum_{i=1}^{10} X_i \geq 9\right] \\ &= \binom{10}{9}\theta^9(1-\theta)^1 + \binom{10}{10}\theta^{10}(1-\theta)^0 \\ &= 10\theta^9 - 10\theta^{10} + \theta^{10} \\ &= 10\theta^9 - 9\theta^{10} \end{aligned}$$

y

$$\begin{aligned} \alpha &= \max_{\theta \in (0, 0.75]} P_\theta\left[\varphi_\tau(\mathbf{X}) = 1\right] \\ &= \max_{\theta \in (0, 0.75]} \left\{10\theta^9 - 9\theta^{10}\right\} \\ &= 10(0.75)^9 - 9(0.75)^{10} \\ &= \frac{255879}{1048576} \\ &\approx 0.2440 \end{aligned}$$

Ejercicio 8.2. Retomando el Ejemplo 8.1 calcule el error de tipo I (α) y de tipo II (β) para cada uno de los siguientes test y para varios valores de θ:

(*D. Anderson, D. Sweeney, T. Williams, J. Camm - Estadística para Negocios y Economía. 11ra Edición, Cengage Learning (2012). Capítulo 9. Ejercicio 60.*) En una línea de producción, el peso promedio con que se llena cada recipiente es 16 onzas. Un llenado insuficiente ocasiona problemas serios y, cuando es detectado, es necesario que el operador detenga la línea de producción para reajustar el mecanismo de llenado. Con base en datos anteriores, se supone que la desviación estándar poblacional es 0.8 onzas. Cada hora, un inspector de control de calidad toma una muestra de 30 recipientes y decide si es necesario detener la producción y hacer un reajuste.

El sistema de hipótesis es,

$$H_0 : \theta = 16 \text{frente a} H_1 : \theta < 16$$

y supongamos los siguientes tests,

$$\tau_1 : \text{ Rechazar } H_0 \text{ si } \bar{x} < 15.81$$

$$\tau_2 : \text{ Rechazar } H_0 \text{ si } \bar{x} < 15.77$$

$$\tau_3 : \text{ Rechazar } H_0 \text{ si } \bar{x} < 15.64$$

Solución

La probabilidad de error tipo I para el primer test τ_1 sería (usando las tablas estadísticas),

$$\begin{aligned}\alpha &= P_{\theta_0}\left[\varphi_{\tau_1}(\mathbf{X}) = 1\right] \\ &= P_{\theta=16}\left[\bar{X} < 15.81\right] \\ &= P\left[Z < \frac{15.81 - 16}{\frac{0.8}{\sqrt{30}}}\right] \\ &\approx P\left[Z < -1.30\right] \\ &\approx 0.0968\end{aligned}$$

De la misma forma, se obtendría $\alpha = 0.0582$ y $\alpha = 0.0069$ para los test τ_2 y τ_3, respectivamente. Los tres test tendrían un tamaño o nivel inferior a 0.1.

```r
n <- 30 # Tamaño de la muestra
sigma <- 0.8 # Desviación estándar poblacional
theta_0 <- 16 # Hipótesis nula simple
bar.x.tau <- c(15.81, 15.77, 15.64) # medias muestrales de los tests
# P[ \bar{X} < bar.x.tau | theta = theta_0 ] y varianza conocida
alpha <- pnorm(bar.x.tau, mean = theta_0, sd = sigma/sqrt(n))
names(alpha) <- paste0("alpha.tau_", 1:3)
print(round(alpha, 6))
```

```
alpha.tau_1 alpha.tau_2 alpha.tau_3
   0.096656    0.057663    0.006855
```

La probabilidad de error tipo II para el primer test τ_1, por ejemplo con $\theta = 15.9$, sería,

$$\begin{aligned}\beta &= P_{\theta=15.9}\left[\bar{X} \geq 15.81\right] \\ &= P\left[Z \geq \frac{15.81 - 15.9}{\frac{0.8}{\sqrt{30}}}\right] \\ &\approx P\left[Z \geq -0.62\right] \\ &\approx 1 - 0.2676 \\ &\approx 0.7324\end{aligned}$$


```r
theta.f <- c(15.3, 15.5, 15.7, 15.9) # Algunos valores con H_0 falsa
i <- length(theta.f)
# P[ \bar{X} > bar.x.tau | theta = theta.f ] y varianza conocida
beta <- 1 - pnorm(rep(bar.x.tau, each = i), mean = theta.f, sd = sigma/sqrt(n))
```

En los anteriores ejercicios, los test fueron dados, y a partir de ellos, calculamos las probabilidades de errores tipo I y II. En particular, calculamos el nivel del test (α) (probabilidad de error tipo I).
¿Será que es posible partir de un nivel de test α dado y mediante el "proceso inverso" obtener o construir un test?

Ejercicio 8.3. Para cada uno de los parámetros θ y las distribuciones $f_X(x;\theta)$ de la población de donde provienen las muestras que se trabajaron en el capítulo anterior (Estimación por intervalo), realice la deducción analítica completa que permita obtener un test, determinado por un estadístico de prueba y un(os) valor(es) crítico(s), para juzgar cada uno de los siguientes tres sistemas de hipótesis,

(i).

$$H_0 : \theta = \theta_0 \quad \text{frente a} \quad H_1 : \theta > \theta_0,$$

(ii).

$$H_0 : \theta = \theta_0 \quad \text{frente a} \quad H_1 : \theta < \theta_0,$$

(iii).

$$H_0 : \theta = \theta_0 \quad \text{frente a} \quad H_1 : \theta \neq \theta_0.$$

8.1.4 Valor p o p-valor

Expresado de forma sencilla, un p-valor es la probabilidad, bajo un modelo estadístico especificado, de que un estadístico que sintetiza alguna característica de los datos sea igual o más extremo que su valor observado.

! Lea acerca de la declaración de la ASA sobre la significancia estadística y los p-valores (por ejemplo en: https://soce.iec.cat/wp-content/uploads/2016/04/efba1672f6161e2bbedc6acf6a6f8d45.pdf)

8.1.5 Relación entre pruebas de hipótesis e intervalos de confianza

¿Existe una relación entre las pruebas de hipótesis y los intervalos de confianza?

Una prueba de hipótesis a dos colas ($H_a : \theta \neq \theta_0$) con un nivel de significancia α es equivalente a calcular un intervalo de confianza bilateral del $100(1-\alpha)\%$ para θ y rechazar H_0 si θ_0 está por fuera del intervalo.

En general, la región de "no rechazo" es equivalente a (es la transformación de) un intervalo de confianza del parámetro de interés.

8.2 Ejercicios

Para cada uno de los siguientes ejercicios: plantee el juzgamiento de hipótesis respectivo, y a partir de una variable pivote que corresponda a la situación dada, responda a lo solicitado.

Ejercicio 8.4.

La cámara de comercio de una comunidad de la costa del Golfo en Florida anuncia en su publicidad que hay disponibilidad de propiedades en el área residencial a un costo medio de \$125000 o menos por lote. Suponga que en una muestra de 32 propiedades se encuentra una media muestral de \$130000 por terreno y una desviación estándar muestral es \$12500. Use 0.05 como nivel de significancia para probar la validez de lo que se dice en la publicidad.

Anderson, Sweeney, Williams & Camm (2016). Estadística para Negocios y Economía. (12a. ed.) Cengage Learning. Capítulo 9. Ejercicio 66.

Ejercicio 8.5.

Una estación de radio de Myrtle Beach anuncia que, por lo menos, 90% de los hoteles y moteles estarán llenos el fin de semana en que se conmemora el Día de los Caídos. La radiodifusora aconseja a sus oyentes hacer sus reservaciones con anticipación si piensan pasar ese fin de semana en esa localidad vacacional. La noche del sábado, una muestra de 58 hoteles y moteles, indicó que 49 estaban completamente llenos y 9 aún tenían habitaciones libres. ¿Cuál es su reacción ante lo anunciado por la estación de radio después de ver la evidencia muestral? Use $\alpha = 0.05$ al realizar el estadístico de prueba. ¿Cuál es el valor-p?

Anderson, Sweeney, Williams & Camm (2016). Estadística para Negocios y Economía. (12a. ed.) Cengage Learning. Capítulo 9. Ejercicio 72.

Ejercicio 8.6.

En una muestra de 9 días de los últimos seis meses se encontró que un dentista había tratado a los siguientes números de pacientes: 22, 25, 20, 18, 15, 22, 24, 19 y 26. Si el número de sujetos atendidos por día tiene una distribución normal, ¿un análisis de estos datos muestrales permitiría rechazar la hipótesis de que la varianza de la cantidad de pacientes atendidos por día es 10? Use un nivel de significancia de 0.10. ¿Cuál es su conclusión?

Anderson, Sweeney, Williams & Camm (2016). Estadística para Negocios y Economía. (12a. ed.) Cengage Learning. Capítulo 11. Ejercicio 29.

Ejercicio 8.7.

En los últimos años prolifera una cantidad cada vez mayor de opciones de entretenimiento que compiten por el tiempo de los consumidores. En 2004 la televisión por cable y la radio superaron a la televisión abierta, la música grabada y los periódicos, convirtiéndose en los medios de entretenimiento más usados (The Wall Street Journal, 26 de enero de 2004). Con una muestra de 15 individuos, los investigadores obtienen los datos de las horas por semana que destinan a ver televisión por cable y de las horas por semana en que escuchan la radio.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
TV	22	8	25	22	12	26	22	19	21	23	14	14	14	16	24
Radio	25	10	29	19	13	28	23	21	21	23	15	18	17	15	23

Use como nivel de significancia 0.05 y haga una prueba para la diferencia entre las medias poblacionales de la cantidad de horas destinadas a la televisión por cable y la cantidad de horas destinadas a la radio. ¿Cuál es el valor-p?

Anderson, Sweeney, Williams & Camm (2016). Estadística para Negocios y Economía. (12a. ed.) Cengage Learning. Capítulo 10. Ejercicio 25.

Ejercicio 8.8.

Con cierta periodicidad, Merrill Lynch solicita a sus clientes evaluaciones sobre los consultores y los servicios financieros que les proporciona. Las puntuaciones más altas en la encuesta de satisfacción del cliente indican mejor servicio con 7 como la puntuación más alta. A continuación se presentan en forma resumida las puntuaciones otorgadas a dos consultores financieros por los miembros de dos muestras aleatorias independientes. El consultor A tiene 10 años de experiencia, mientras que el consultor B tiene sólo 1 año. Use $\alpha = 0.05$ y realice una prueba para determinar si el consultor con más experiencia tiene la media poblacional más alta en la evaluación del servicio.

Consultor	n	$\bar{x}$	s
A	16	6.82	0.64
B	10	6.25	0.75

Suponga varianzas iguales y repita suponiendo varianzas diferentes.

Anderson, Sweeney, Williams & Camm (2016). Estadística para Negocios y Economía. (12a. ed.) Cengage Learning. Capítulo 10. Ejercicio 17.

Ejercicio 8.9.

En un estudio de la Asociación Estadounidense de Automovilistas (AAA, por sus siglas en inglés) se investigó si era más probable que conductores de género masculino o femenino se detuvieran para solicitar indicaciones sobre cómo llegar a una dirección (AAA, enero de 2006). Se preguntaba a los conductores: "Si usted y su cónyuge van en su automóvil y se pierden, ¿se detiene para preguntar por el domicilio que busca?" En una muestra representativa se encontró que 300 de 811 mujeres dijeron que sí se detenían para preguntar, mientras que 255 de 750 hombres dijeron que también lo hacían. La hipótesis de investigación de AAA afirmaba que era más probable que las mujeres se detuvieran para preguntar por el domicilio. Pruebe la hipótesis usando $\alpha = 0.05$.

Anderson, Sweeney, Williams & Camm (2016). Estadística para Negocios y Economía. (12a. ed.) Cengage Learning. Capítulo 10. Ejercicio 32.

Ejercicio 8.10.

El Amstat News (diciembre de 2004) lista los sueldos medios de profesores asociados de estadística en instituciones de investigación, en escuelas de humanidades y en otras instituciones en Estados Unidos. Suponga que una muestra de 200 profesores asociados de instituciones de investigación tiene un sueldo promedio de \$70750 anuales con una desviación estándar de \$6000. Suponga también que una muestra de 200 profesores asociados de otros tipos de instituciones tienen un sueldo promedio de \$65200 con una desviación estándar de \$5000. Pruebe la hipótesis de que el sueldo medio de profesores asocia dos de instituciones de investigación es \$2000 más alto que el de los profesores de otras instituciones. Utilice un nivel de significancia de 0.01. Para determinar si debe usar varianzas iguales o distintas, realice la prueba de hipótesis respectiva.

Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación. Ejercicio 10.32.

9 Tests más potentes

En esta sección se hará una revisión de algunos aspectos generales relacionados con la potencia de un test, los tests más potentes, y los tests de razón de verosimilitudes.

9.1 Función de potencia de un test

Sea τ un test no aleatorizado para el juzgamiento de H_0 con función crítica $\varphi_\tau(\mathbf{x})$. La **función de potencia** denotada como $\pi_\tau(\theta)$ es una función con dominio Θ y recorrido el intervalo $(0,1)$, definida como,

$$\pi_\tau(\theta) = P_\theta\left[\varphi_\tau(\mathbf{X}) = 1\right]$$

Si $\Theta_1 = \Theta_0^C$, la función

$$\beta_\tau(\theta) = 1 - \pi_\tau(\theta)$$

es denominada **curva característica de operación** o **curva CO del test** τ.

La función de potencia de un test en teoría perfecto τ sería,

$$\pi_\tau(\theta) = 1 - \mathrm{I}_{\Theta_0}(\theta)$$

o equivalentemente la curva CO para un test en teoría perfecto τ se establecería como,

$$\beta_\tau(\theta) = 1 - \mathrm{I}_{\Theta_1}(\theta)$$

Ejercicio 9.1. El tiempo que una persona requiere para comprar una tarjeta de ingreso en el sistema de Transmilenio en la estación de Alcala durante el 2002, ha mostrado un comportamiento que sugiere el modelo Uniforme en el intervalo $(0,\theta)$ para su descripción. Se afirma que el tiempo máximo de permanencia en la fila está entre dos y tres minutos. Para evaluar la afirmación y tomar los correctivos del caso, se va a registrar el tiempo empleado por n personas que serán elegidas por medio de un procedimiento especial de muestreo en la rampa de ingreso, y se propone la utilización del test,

$$\tau: \text{ Rechazar } H_0 \text{ si } x_{(n)} \leq 1.9 \text{ o si } x_{(n)} > 3.1,$$

para el juzgamiento de la hipótesis nula H_0 en el sistema,

$$H_0 : \theta \in [2,3] \text{frente a} H_1 : \theta \notin [2,3].$$

Determinar la función de potencia del test propuesto.

Solución

La función de potencia del test propuesto es

$$\begin{aligned}
\pi_\tau(\theta) &= P\left[\varphi_\tau(\mathbf{X}) = 1\right] \\
&= P\left[X_{(n)} \leq 1.9 \text{ o } X_{(n)} > 3.1\right] \\
&= P\left[X_{(n)} \leq 1.9\right] + P\left[X_{(n)} > 3.1\right] \\
&= P\left[X_{(n)} \leq 1.9\right] + 1 - P\left[X_{(n)} \leq 3.1\right] \\
&= \mathrm{F}_{X_{(n)}}(1.9;\theta) + 1 - \mathrm{F}_{X_{(n)}}(3.1;\theta)
\end{aligned}$$

donde,

$$\mathrm{F}_{X_{(n)}}(x;\theta) = \left[\mathrm{F}_X(x)\right]^n = \frac{x^n}{\theta^n}\mathrm{I}_{[0,\theta)}(x) + \mathrm{I}_{[\theta,\infty)}(x),$$

por lo tanto

9.2 Test más potente

Definición 9.1. Sea el sistema de hipótesis,

$$H_0 : \theta = \theta_0 \text{frente a} H_1 : \theta = \theta_1$$

el test τ se denomina **test más potente** para H_0, con nivel α, si,

$$\pi_\tau(\theta_0) = \alpha$$

y

$$\pi_\tau(\theta_1) \leq \pi_{\tau^*}(\theta_1)$$

para todo test τ^* de nivel menor o igual a α.

9.3 Test uniformemente más potente (UMP)

Definición 9.2. Sea el sistema de hipótesis,

$$H_0 : \theta \in \Theta_0 \text{frente a} H_1 : \theta \in \Theta_1$$

el test τ se denomina **test uniformememte más potente (test UMP)**, para H_0, con nivel α, si,

$$\sup_{\theta \in \Theta_0} \pi_\tau(\theta) = \alpha$$

y

$$\pi_\tau(\theta) \leq \pi_{\tau^*}(\theta)$$

para todo $\theta \in \Theta_1$ y para todo test τ^* de nivel menor o igual a α.

> ⚠ ¿existe algún mecanismo para obtener o construir el mejor test posible para cada situación (el test más potente para cada sistema de hipótesis y cada α dados)?

9.4 Razón simple de verosimilitudes

Definición 9.3. Sea el sistema de hipótesis simples,

$$H_0 : \theta = \theta_0 \text{frente a} H_1 : \theta = \theta_1,$$

el siguiente test se denomina **test de razón simple de verosimilitudes** de nivel α:

$$\tau : \text{ Rechazar } H_0 \text{ si } W\left(x_1, \ldots, x_n\right) < k,$$

donde,

$$W\left(x_1, \dots, x_n\right) = \frac{\mathrm{L}\left(\theta_0; x_1, \dots, x_n\right)}{\mathrm{L}\left(\theta_1; x_1, \dots, x_n\right)},$$

y k es una constante tal que,

$$P_{\theta_0}\left[W\left(X_1, \dots, X_n\right) < k\right] = \alpha$$

9.5 Lema de Neyman Pearson

Teorema 9.1. *El* ***test de razón simple de verosimilitudes*** *es un* ***test más potente*** *para* H_0.

9.6 Razón generalizada de verosimilitudes

Definición 9.4. Sea el sistema de hipótesis,

$$H_0 : \theta \in \Theta_0 \text{frente a} H_1 : \theta \in \Theta_1,$$

con $\Theta_1 \cup \Theta_0 = \Theta$ (hipótesis antitéticas), el siguiente test se denomina **test de razón simple de verosimilitudes** de nivel α:

$$\tau : \text{ Rechazar } H_0 \text{ si } W\left(x_1, \dots, x_n\right) < k,$$

donde,

$$W\left(x_1, \dots, x_n\right) = \frac{\sup\limits_{\theta \in \Theta_0} \mathrm{L}\left(\theta; x_1, \dots, x_n\right)}{\sup\limits_{\theta \in \Theta} \mathrm{L}\left(\theta; x_1, \dots, x_n\right)},$$

y k es una constante tal que,

$$\max_{\theta \in \Theta_0} P_\theta\left[W\left(X_1, \dots, X_n\right) < k\right] = \alpha.$$

Note que el denominador de $W\left(x_1, \dots, x_n\right)$ es la función de verosimilitud evaluada en el estimador máximo verosímil de θ.

Por otra parte, debemos conocer la distribución de $W = W\left(X_1, \dots, X_n\right)$, para poder establecer la constante k, la cual necesitamos para terminar de formular por completo el test. Con frecuencia tendremos que recurrir a tests equivalentes derivados de W (tests a partir de variables aleatorias resultantes de aplicar una serie de operaciones a W) en donde la distribución sea conocida. Otra alternativa es usar distribuciones asintóticas (aproximaciones), en aquellos casos en donde se considere que la muestra es suficientemente grande.

9.7 Razón monótona de verosimilitudes (MLR)

Definición 9.5. Una familia de densidades $\{\mathrm{f}_X(x;\theta) : \theta \in \Theta \subseteq \mathbb{R}\}$ se dice que tiene una **razón monótona de verosimilitudes (MLR)** en la estadística $T = T(X_1, \ldots, X_n)$, si para dicha estadística, el cociente,

$$\frac{\mathrm{L}(\theta_1; x_1, \ldots, x_n)}{\mathrm{L}(\theta_2; x_1, \ldots, x_n)}$$

es una función no creciente o no decreciente de $T(x_1, \ldots, x_n)$, para cada $\theta_1 < \theta_2$.

Ejercicio 9.2. Realice la deducción analítica completa que muestra que la familia de densidades Poisson tiene razón monótona de verosimilitudes en la estadística $T = \bar{X}$.

9.8 UMP bajo MLR

Teorema 9.2. *Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con densidad perteneciente a una familia de densidades con **MLR** en la estadística $T = T(X_1, \ldots, X_n)$. Si dentro del juzgamiento de la hipótesis nula H_0 se considera el sistema de hipótesis,*

$$H_0 : \theta \leq \theta_0 \textit{frente a} H_1 : \theta > \theta_0,$$

y si la razón MLR es:

a. ***no decreciente**, entonces,*

$$\tau : \text{ Rechazar } H_0 \text{ si } T(x_1, \ldots, x_n) < k$$

*es un test **uniformemente más potente (UMP)** para H_0, donde k es tal que,*

$$P_{\theta_0}[T(X_1, \ldots, X_n) < k] = \alpha$$

b. ***no creciente**, entonces,*

$$\tau : \text{ Rechazar } H_0 \text{ si } T(x_1, \ldots, x_n) > k$$

*es un test **uniformemente más potente (UMP)** para H_0, donde k es tal que,*

$$P_{\theta_0}[T(X_1, \ldots, X_n) > k] = \alpha$$

9.9 UMP bajo familia exponencial

Teorema 9.3. *Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución perteneciente a la familia exponencial unidimensional. Si dentro del juzgamiento de la hipótesis nula H_0 se considera el sistema de hipótesis,*

$$H_0 : \theta \leq \theta_0 \ (o\ \theta = \theta_0)\, frente\ a H_1 : \theta > \theta_0,$$

si T es la estadística natural de la familia $\left(T = \sum_{i=1}^{n} \mathrm{T}(X_i)\right)$, y si $\mathrm{c}(\theta)$ es una función monótona:

a. ***creciente*** *de θ, entonces,*

$$\tau : \ \textit{Rechazar } H_0 \textit{ si } T > k$$

es un test ***uniformemente más potente (UMP)*** *para H_0, donde k es tal que,*

$$P_{\theta_0}[T > k] = \alpha$$

b. ***decreciente*** *de θ, entonces,*

$$\tau : \ \textit{Rechazar } H_0 \textit{ si } T < k$$

es un test ***uniformemente más potente (UMP)*** *para H_0, donde k es tal que,*

$$P_{\theta_0}[T < k] = \alpha$$

10 Algunas pruebas adicionales

EN CONSTRUCCION.

Ejercicio 10.1. Investigue acerca de la denominada "prueba de signos" (describa en qué consiste, en qué casos se puede utilizar, qué supuestos requiere, y cómo se obtendrían estadístico de prueba, valor(es) crítico(s) y p-valor). A partir de lo anterior resuelva el siguiente ejercicio:

> Los siguientes datos representan el tiempo, en minutos, que un paciente tiene que esperar durante 12 visitas al consultorio de un médico antes de ser atendido:
>
> | | | | | | |
> |---|---|---|---|---|---|
> | 17 | 15 | 20 | 20 | 32 | 28 |
> | 12 | 26 | 25 | 25 | 35 | 24 |
>
> Utilice la ***prueba de signo*** a un nivel de significancia de 0.05 para probar la afirmación del médico de que la mediana del tiempo de espera de sus pacientes no es mayor de 20 minutos.
>
> Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación. Ejercicio 16.1.

Ejercicio 10.2. Investigue acerca de la denominada "prueba de rangos con signo de Wilcoxon" (describa en qué consiste, en qué casos se puede utilizar, qué supuestos requiere, y cómo se obtendrían estadístico de prueba, valor(es) crítico(s) y p-valor). A partir de lo anterior resuelva el siguiente ejercicio:

> Se afirma que una nueva dieta reducirá el peso de una persona en 4.5 kilogramos, en promedio, en un periodo de dos semanas. Se registran los pesos de 10 mujeres que siguen esta dieta, antes y después de un periodo de dos semanas, y se obtienen los siguientes datos:

	1	2	3	4	5	6	7	8	9	10
Antes	58.5	60.3	61.7	69	64	62.6	56.7	63.6	68.2	59.4
Después	60	54.9	58.1	62.1	58.5	59.9	54.4	60.2	62.3	58.7

Utilice la ***prueba de rango con signo*** a un nivel de significancia de 0.05 para probar la hipótesis de que la dieta reduce la mediana del peso en 4.5 kilogramos, en comparación con la hipótesis alternativa de que la mediana de la pérdida de peso es menor que 4.5 kilogramos.

Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación. Ejercicio 16.11 (Ejercicio 16.5).

Ejercicio 10.3. Investigue acerca de la denominada "prueba U de Mann-Withney (suma de rangos de Wilcoxon)" (describa en qué consiste, en qué casos se puede utilizar, qué supuestos requiere, y cómo se obtendrían estadístico de prueba, valor(es) crítico(s) y p-valor). A partir de lo anterior resuelva el siguiente ejercicio:

Un fabricante de cigarrillos afirma que el contenido de alquitrán de la marca de cigarrillos B es menor que la de la marca A. Para probar esta afirmación se registraron las siguientes medidas del contenido de alquitrán, en miligramos:

Marca A	1	12	9	13	11	14
Marca B	8	10	7			

Utilice la ***prueba de suma de rangos*** con un nivel de significancia de 0.05 para probar si la afirmación es válida.

Walpole, Myers & Myers (2012). Probabilidad y estadística para ingeniería y ciencias. (9a. ed.) Pearson Educación. Ejercicio 16.15.

EN CONSTRUCCION.

A Distribuciones muestrales

En esta sección se hará una revisión de algunos aspectos relacionados con algunas distribuciones continuas que suelen ser llamadas distribuciones muestrales.

A.1 Distribución Ji/Chi-cuadrado

Características:

Si

$$Z_1 \sim \mathcal{N}(0,1), \ldots, Z_\nu \sim \mathcal{N}(0,1)$$

entonces

$$\sum_{i=1}^{\nu} Z_i^2 \sim \chi_\nu^2$$

$\Big($Si $Z_1, \ldots, Z_\nu$ son variables aleatorias normales estándar independientes, entonces la variable aleatoria $\sum_{i=1}^{\nu} Z_i^2$ es chi-cuadrado de parámetro ν.$\Big)$

Dominio (valores que puede tomar la variable aleatoria):

$x \in \mathbb{R}^+$

Parámetros:

υ, grados de libertad de la variable aleatoria. $\upsilon \in \mathbb{Z}^+$

Notación:

$X \sim \chi_\upsilon^2$, esto se lee así: la variable X tiene una distribución chi-cuadrado con υ grados de libertad.

Función de densidad:

$$f_X(x) = \begin{cases} \frac{1}{2^{\upsilon/2}\Gamma(\upsilon/2)} x^{\upsilon/2-1} e^{-x/2} & \text{Si } x \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

Función de distribución acumulativa:

$F_X(x)$ no tiene una expresión analítica cerrada simple.

Valor esperado y varianza:

- $\mu_X = E[X] = \upsilon$
- $\sigma_X^2 = Var[X] = 2\upsilon$

Función característica:

$$\varphi_X(t) = (1 - 2\mathrm{i}t)^{-\upsilon/2}$$

Para ilustrar:

> Efecto de los parámetros
>
> https://cjtorresj.shinyapps.io/introChiSqDistribution/

- Chi-Squared Dist
- Chi-Square Distribution Applet/Calculator

¿Cómo hallar o calcular probabilidades?:

Use aproximaciones numéricas tabuladas (ver libro de Anderson. Página 983) o herramientas tecnológicas que calculen las aproximaciones numéricas.

Ejercicio A.1.

- Sea $X \sim \chi^2_{\nu=14}$

  a) $P[X > 23.685] = ¿?$
  b) $P[X < 6.571] = ¿?$
  c) $P[4.075 < X < 21.064] = ¿?$

- Sea $X \sim \chi^2_{\nu=6}$

  a) $P[X > ¿?] = 0.95$
  b) $P[X < ¿?] = 0.9$
  c) $P[¿? < X < 16.812] = 0.98$

- Sea $X \sim \chi^2_{\nu=23}$

  a) $P[X > ¿?] = 0.05$
  b) $P[X > ¿?] = 0.975$
  c) $P[X > ¿?] = 0.025$

A.2 Distribución t de Student

Características:

Si

$$Z \sim \mathcal{N}(0,1)$$

y

$$V \sim \chi^2_\nu$$

son independiente, entonces,

$$\frac{Z}{\sqrt{V/\nu}} \sim t_\nu$$

(Si Z es una variable aleatoria normal estándar, V es una variable aleatoria chi-cuadrado con ν grados de libertad, y Z y V son independientes, entonces la variable aleatoria $\frac{Z}{\sqrt{V/\nu}}$ es t de Student con ν grados de libertad.)

Dominio (valores que puede tomar la variable aleatoria):

$x \in \mathbb{R}$

Parámetros:

υ, grados de libertad de la variable aleatoria. $\upsilon \in \mathbb{Z}^+$

Notación:

$X \sim t_\upsilon$, esto se lee así: la variable X tiene una distribución t de Student con υ grados de libertad.

Función de densidad:

$$f_X(x) = \frac{\Gamma((\upsilon+1)/2)}{\Gamma(\upsilon/2)\sqrt{\pi\,\upsilon}}\left(1+\frac{x^2}{\upsilon}\right)^{-(\upsilon+1)/2}$$

Función de distribución:

$F_X(x)$ no tiene una expresión analítica cerrada simple.

Valor esperado y varianza:

- $\mu_X = E[X] = 0$ para $\upsilon > 1$.
- $\sigma^2_X = Var[X] = \frac{\upsilon}{\upsilon-2}$ para $\upsilon > 2$.

Para ilustrar:

Efecto de los parámetros

https://cjtorresj.shinyapps.io/introtDistribution/

- t Distribution
- Student's t-Distribution Applet/Calculator

¿Cómo hallar o calcular probabilidades?:

Use aproximaciones numéricas tabuladas (ver libro de Anderson. Página 980) o herramientas tecnológicas que calculen las aproximaciones numéricas.

Ejercicio A.2.

- Sea $X \sim t_{\nu=8}$

  a) $P[X > 2.306] = ¿?$
  b) $P[X < 0.889] = ¿?$
  c) $P[-1.397 < X < 2.896] = ¿?$

- Sea $X \sim t_{\nu=16}$

  a) $P[X > ¿?] = 0.95$
  b) $P[X < ¿?] = 0.1$
  c) $P[¿? < X < ¿?] = 0.98$

- Sea $X \sim t_{\nu=32}$

  a) $P[X > ¿?] = 0.05$
  b) $P[X > ¿?] = 0.975$
  c) $P[X > ¿?] = 0.025$

A.3 Distribución F de Fisher-Snedecor

Características:

Si

$$U \sim \chi^2_{\nu_1}$$

y

$$V \sim \chi^2_{\nu_2}$$

son independientes, entonces,

$$\frac{U/\nu_1}{V/\nu_2} \sim f_{\nu_1,\nu_2}$$

$\Big($Si U es una variable aleatoria chi-cuadrado con ν_1 grados de libertad, V es una variable aleatoria chi-cuadrado con ν_2 grados de libertad, y U y V son independientes, entonces la variable aleatoria $\frac{U/\nu_1}{V/\nu_2}$ es F de Fisher con ν_1 grados de libertad en el numerador y ν_2 grados de libertad en el denominador.$\Big)$

Dominio (valores que puede tomar la variable aleatoria):

$x \in [0, \infty)$

Parámetros:

υ_1, grados de libertad en el numerador de la variable aleatoria, y υ_2, grados de libertad en el denominador de la variable aleatoria. $\upsilon_1, \upsilon_2 \in \mathbb{Z}^+$

Notación:

$X \sim f_{\upsilon_1, \upsilon_2}$, lo cual se lee así: la variable X tiene una distribución F de Fisher con υ_1 y υ_2 grados de libertad.

Función de densidad:

$$f_X(x) = \frac{\sqrt{\frac{(\upsilon_1\, x)^{\upsilon_1}\, \upsilon_2^{\upsilon_2}}{(\upsilon_1\, x + \upsilon_2)^{\upsilon_1 + \upsilon_2}}}}{x\, B(\upsilon_1/2, \upsilon_2/2)}$$

Función de distribución:

$F_X(x)$ no tiene una expresión analítica cerrada simple.

Valor esperado y varianza:

- $\mu_x = E[X] = \frac{\upsilon_2}{\upsilon_2 - 2}$ para $\upsilon_2 > 2$.
- $\sigma_X^2 = Var[X] = \frac{2\,\upsilon_2^2(\upsilon_1 + \upsilon_2 - 2)}{\upsilon_1(\upsilon - 2)^2(\upsilon - 4)}$ para $\upsilon_2 > 4$.

Para ilustrar:

Efecto de los parámetros

https://cjtorresj.shinyapps.io/introFDistribution/

- F-Distribution
- F-Distribution Applet/Calculator

¿Cómo hallar o calcular probabilidades?:

Use aproximaciones numéricas tabuladas o herramientas tecnológicas que calculen las aproximaciones numéricas.

Cuando se usan aproximaciones numéricas tabuladas, suele ser necesario usar la siguiente propiedad:

Si

$$Y \sim f_{\upsilon_2,\upsilon_1}$$

entonces

$$\frac{1}{Y} = X \sim f_{\upsilon_1,\upsilon_2}$$

Lo que quiere decir que,

$$P[Y < y] = P\left[\frac{1}{Y} > \frac{1}{y}\right] = P\left[X > \frac{1}{y}\right]$$

Ejercicio A.3.

- Sea $X \sim F_{\nu_1=10,\nu_2=15}$

  a) $P[X > 2.06] = ¿?$
  b) $P[X < 3.80] = ¿?$
  c) $P[2.54 < X < 3.06] = ¿?$
  d) $P[X > ¿?] = 0.025$
  e) $P[X > ¿?] = 0.95$
  f) $P[X > ¿?] = 0.975$

- Sea $X \sim F_{\nu_1=15,\nu_2=10}$

  a) $P[X > 2.24] = ¿?$
  b) $P[X < 4.56] = ¿?$
  c) $P[2.85 < X < 3.52] = ¿?$
  d) $P[X > ¿?] = 0.025$
  e) $P[X > ¿?] = 0.95$
  f) $P[X > ¿?] = 0.975$

B Familia exponencial

En esta sección se hará una revisión de algunos aspectos relacionados con la familia exponencial.

Definición B.1 (Familia exponencial uniparamétrica)**.** Una distribución de probabilidad, con parámetro θ, pertenece a la **familia exponencial uniparamétrica (univariada/unidimensional)**, si la función de masa de probabilidad / función de densidad se puede escribir de la forma,

$$\mathrm{f}_X(x;\theta) = \exp\Big(\mathrm{c}(\theta)\,\mathrm{T}(x) + d(\theta) + S(x)\Big)\, I_A(x)$$

donde $\mathrm{c}(\theta)$ y $\mathrm{d}(\theta)$ son funciones que dependen únicamente de θ, $\mathrm{S}(x)$ y $\mathrm{T}(x)$ son funciones que dependen únicamente de x, y A no depende de θ.

Naturalmente, las distribuciones con más de un parámetro pueden considerarse distribuciones uniparamétricas, si todos sus parámetros, excepto uno, se consideran conocidos.

Definición B.2 (Familia exponencial multiparamétrica)**.** Una distribución de probabilidad pertenece a la **familia exponencial multiparamétrica (multivariada/multidimensional)** si la función de masa de probabilidad / función de densidad se puede escribir de la forma,

$$\mathrm{f}_X(\mathbf{x};\theta) = \exp\Big(\mathrm{c}(\theta)'\,\mathrm{T}(\mathbf{x}) + \mathrm{d}(\theta) + S(\mathbf{x})\Big)\, I_A(\mathbf{x})$$

donde $\mathrm{T}(\mathbf{x})$ y $\mathrm{c}(\theta)$ son funciones vectoriales, y $\mathrm{S}(\mathbf{x})$ y $\mathrm{d}(\theta)$ son funciones reales.

Ejercicio B.1. Determine si la distribución binomial $(X \sim \mathcal{B}inomial(n = n_0 \text{ conocido}, p = \theta))$ pertenece a la familia exponencial uniparamétrica.

Ejercicio B.2. Para $X_1, \dots, X_n$ son variables aleatorias idénticamente distribuidas con función de densidad común perteneciente a la familia exponencial uniparamétrica, determine si la función de densidad conjunta $\mathrm{f}_X(x_1, \dots, x_n)$ también pertenece a la familia exponencial uniparamétrica.

⚠ No todos los libros usan una misma notación para la definición de familia exponencial, sin embargo utilizan ecuaciones equivalentes que expresan lo mismo:

E. Cepeda (2015). Estadística Matemática. Universidad Nacional de Colombia.:

$$\mathrm{P}_X(\mathbf{x}, \theta) = \exp\left(\mathrm{c}(\theta)'\,\mathrm{T}(\mathbf{x}) + \mathrm{d}(\theta) + S(\mathbf{x})\right) I_A(\mathbf{x})$$

J. H. Mayorga (2004). Inferencia Estadística. Universidad Nacional de Colombia.:

$$\mathrm{f}(\mathbf{x}, \theta) = \exp\left(\sum_{i=1}^{k} \mathrm{d}_i(\mathbf{x})\mathrm{c}_i(\theta) + \mathrm{a}(\theta) + \mathrm{b}(\mathbf{x})\right)$$

G. Casella, R. L. Berger (2001). Statistical Inference. Brooks/Cole.:

$$\mathrm{f}(\mathbf{x}|\theta) = \mathrm{h}(\mathbf{x})\mathrm{c}(\theta)\exp\left(\sum_{i=1}^{k} \mathrm{w}_i(\theta)\mathrm{t}_i(\mathbf{x})\right)$$

J. Shao (2003). Mathematical Statistics. Springer.:

$$\frac{dP_\theta}{d\nu}(\mathbf{w}) = \exp\left(\left[\eta(\theta)\right]^T \mathrm{T}(\mathbf{w}) - \xi(\theta)\right)\mathrm{h}(\mathbf{w})$$

Si una función de densidad $\mathrm{f}_X(x;\theta)$ pertenece a la familia exponencial uniparámetrica y si $c()$ es una función uno a uno, haciendo $\eta = \mathrm{c}(\theta)$, y por ende $\theta = \mathrm{c}^{-1}(\eta)$, entonces dicha función de densidad se puede escribir de la siguiente manera,

$$\mathrm{f}_X(x;\eta) = \exp\left(\eta\,\mathrm{T}(x) + \mathrm{d}_0(\eta) + \mathrm{S}(x)\right) I_A(x)$$

donde $\mathrm{d}_0(\eta) = \mathrm{d}\left(\mathrm{c}^{-1}(\eta)\right)$. Esta forma de escribir la familia exponencial uniparamétrica es denominada la **forma natural de la familia exponencial**, η es conocido como el **parámetro natural** de la distribución y $T(x)$ ha sido llamada la **estadística natural** por algunos autores.

Si c no es una función uno a uno, $\mathrm{d}_0(\eta)$ se puede determinar fácilmente.

$$\mathrm{d}_0(\eta) = -\log\int_A \exp\left(\eta\,\mathrm{T}(x) + \mathrm{S}(x)\right)\,dx$$

Teorema B.1. *Si X tiene distribución en la familia exponencial uniparamétrica y η es un punto interior de $H = \{\eta : \mathrm{d}_0(\eta)$ es finito$\}$, la función generadora de momentos de $\mathrm{T}(X)$ está* *dada por*

$$\Psi(s) = \exp\left(\mathrm{d}_0(\eta) - \mathrm{d}_0(\eta + s)\right)$$

A partir de lo anterior, se tiene que,

- $E[\mathrm{T}(X)] = -\mathrm{d}_0'(\eta)$.
- $Var[\mathrm{T}(X)] = -\mathrm{d}_0''(\eta)$

Ejercicio B.3. Sea $X_1, \dots, X_n$ una muestra aleatoria de una población con distribución binomial de parámetros $n = n_0 \in \mathbb{N}$ conocido y $p = \theta \in (0, 1)$. **Utilizando el anterior resultado**,

- Determine la media y la varianza de la distribución poblacional (ya se sabe que tienen que dar $\mu_X = np$ y $\sigma_X^2 = np(1 - p)$).
- Determine la función generadora de momentos, la media y la varianza del correspondiente estadístico natural de la muestra ($\mathrm{T}(X)$).

⚠ En general, en la familia exponencial, el conjunto de valores para los cuales $\mathrm{f}_X(x; \theta) > 0$ no puede depender de θ ¿por qué?

Las estimaciones de máxima verosimilitud de $\eta = (\eta_1, \dots, \eta_k)$ se pueden obtener solucionando el sistema de ecuaciones dado por,

$$\frac{\partial \mathrm{d}_0}{\partial \eta_i} = -\mathrm{T}_i(x)$$

para $i = 1, \dots, k$.

B.1 Ejercicios

1. Verifique que las siguientes distribuciones pertenecen a la familia exponencial uniparamétrica. Encuentre su forma canónica y el conjunto de valores posibles del parámetro natural. Encuentre función generadora de momentos, valor esperado y varianza utilizando los resultados expuestos en esta sección.

    - Poisson.
    - Binomial Negativa con el parámetro número de éxitos conocido.
    - Exponencial.
    - Weibull con el parámetro de forma conocido.
    - Laplace con media conocida
    - Normal con varianza conocida.
    - Normal con media conocida.

2. Verifique que las siguientes distribuciones pertenecen a la familia exponencial multiparamétrica

    - Normal
    - Gamma
    - Beta
    - Lognormal