Apuntes para el curso de
Cálculo de Probabilidades I

Ernesto Barrios Zamudio

2 de enero de 2024

Versión 0.24

Índice

Prefacio

Las condiciones en que continuamos este año 2021 ha motivado el trabajo. La imposibilidad de compartir mis notas personales por su desorden me llevó a hacer manuscritos con la mayoría del material del temario de Cálculo de Probabilidades I. En paralelo, comencé a pasar las notas a una presentación más formal usando LaTeX. La mayoría de las gráficas están hechas con R y algunos de los dibujos con LaTeX mismo o *Mayura*. Este documento es el resultado.

Estas *apuntes* son precisamente eso, unos apuntes o notas para apoyar el curso Cálculo de Probabilidades I que ofrezco anualmente en ITAM.

Las notas de apoyo para el curso de Cálculo de Probabilidades II las encuentra en Barrios (2024).

Durante el curso es mi responsabilidad motivar y ligar los distintos temas y en este sentido las notas son de apoyo al desarrollo teórico y técnico de los mismos. No se pretende que los temas sean autocontenidos ni son una versión muy preliminar de algo más elaborado y formal. No es material para ser referenciado.

Cualquier error que identifique, comentario y/o sugerencia serán bienvenido. Diríjalo a Ernesto Barrios `<ebarrios at itam.mx>`.

Ciudad de México, 28 de julio de 2021

1. Espacios de Probabilidad

1.1. Probabilidad y axiomas de probabilidad

Definición : Un **experimento** o **ensayo aleatorio** $\mathcal{E}$ (EA) es aquel cuya salida varía de manera impredecible. Experimentos o ensayos bajo "las mismas condiciones" podrán arrojar distintas salidas o resultados.

Ejemplo :

- La cara hacia arriba en el lanzamiento de un dado.
- El número de personas que responden una encuesta de interés en un día.
- El tiempo de vida útil de las lámparas de los salones de clase.
- La proporción de enfermos curados bajo cierto tratamiento.
- El rendimiento diario de cierta inversión.
- La cantidad que ha de pagar una aseguradora por accidente automovilístico.

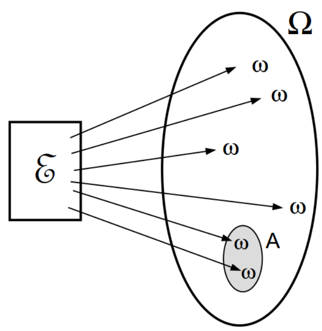


Figura 1: Experimento aleatorio $\mathcal{E}$ arroja salidas impredecibles ω. El espacio muestral Ω es el conjunto de todas las posibles resultados ω. El conjunto A de salidas de interés denota un evento.

Definición : Sea Ω el conjunto de todas las posibles realizaciones, salidas o resultados del experimento aleatorio $\mathcal{E}$, $\Omega = \{\omega : \omega \text{ es una posible salida del } \mathcal{E}\}$. Ω se dice el **espacio muestral** o **espacio muestra** (EM).

La figura 1 ilustra el experimento aleatorio $\mathcal{E}$ y espacio muestral Ω.

Nota: El espacio muestral Ω puede ser finito o infinito, numerable o no numerable. Refiéranse a los ejemplos antes citados.

Definición : Sea $\mathcal{S}$ una familia no vacía de subconjuntos de Ω. Si $A \subseteq \Omega$ y es miembro de la familia $\mathcal{S}$, A se dice que es un **evento**.

Si la salida del experimento aleatorio $\mathcal{E}$ es $\omega \in \Omega$ y $\omega \in A$, se dice que *el evento A ha ocurrido*. De otra forma, si $\omega \notin A$, se dice que A *no ocurrió*, por lo que A^C también es un evento y deberá ser miembro de la familia $\mathcal{S}$.

Sean $A, B \in \mathcal{S}$, eventos. Si $\mathcal{E} \to \omega$, el experimento aleatorio arroja ω y $\omega \in A \cup B$, se dice que *A o B han ocurrido*. Así, también $A \cup B$ deberá ser evento y $A \cup B \in \mathcal{S}$. De manera similar, se dice que *A y B han ocurrido* si $\omega \in A$ y $\omega \in B$. Esto es, $\omega \in A \cap B$, por lo que nuevamente $A \cap B$ deberá ser también un evento y por lo mismo $A \cap B \in \mathcal{S}$. Así, si $A_1, A_2, \ldots, A_n \in \mathcal{S}$, se puede decir si $\omega \in \cup_{i=1}^n A_i$, o bien, si $\omega \in \cap_{i=1}^n A_i$. Luego, resulta natural pedir que la familia de eventos $\mathcal{S}$, subconjuntos de Ω, sea *cerrada bajo complementos, uniones e intersecciones* finitas. Aún más, puesto que siempre que $\mathcal{E} \to \omega$, con $\omega \in \Omega$, el espacio muestral Ω *siempre ocurre*, por lo que él mismo es un evento, llamado en ocasiones **evento seguro** y en consecuencia el conjunto vacío es un evento pues $\emptyset = \Omega^C \in \mathcal{S}$ y es llamado el **evento imposible**.

Por razones técnicas que se estudian en **Teoría de la Medida** y cuyos detalles van más allá de los alcances de este curso es necesario extender la familia de eventos $\mathcal{S}$ para que incluya también las uniones y en consecuencia intersecciones numerables de eventos. (Para elementos de Teoría de la Medida, vea por ejemplo, Bartle (1966), Grabinsky (2000) y Ash and Doléans-Dade (2000).)

Definición : Sea Ω un espacio muestral y $\mathcal{S}$ una familia no vacía de subconjuntos de Ω. Se dice que $\mathcal{S}$ es un **álgebra (campo)** de subconjuntos de Ω, si satisface

i) Si $A \in \mathcal{S}$, entonces $A^C \in \mathcal{S}$.

ii) Si $A_1, \ldots, A_n \in \mathcal{S}$, entonces $\cup_{i=1}^n A_i \in \mathcal{S}$.

$\mathcal{S}$ se dice que es una **σ-álgebra (σ-campo)** de subconjuntos de Ω, si además la familia es cerrada bajo uniones numerables

ii') Si $A_1, A_2, \cdots \in \mathcal{S}$, entonces $\cup_{i=1}^\infty A_i \in \mathcal{S}$.

Nota:

1. Si $\mathcal{S}$ es una σ-álgebra de subconjuntos de Ω. Entonces,

   i) Ω y $\emptyset \in \mathcal{S}$.

   ii) Si $A_1, A_2, \cdots \in \mathcal{S}$, entonces $\cap_{i=1}^\infty A_i \in \mathcal{S}$.

2. Sea Ω un conjunto no vacío y $A \subset \Omega$. La σ-álgebra más pequeña que contiene a A es $\mathcal{S} = \{\emptyset, A, A^C, \Omega\}$

3. Sea $\Omega = \{\omega_1, \ldots, \omega_n\}$ un conjunto finito. Entonces el conjunto potencia de Ω, familia de todos los suibconjuntos de Ω,

   $$\mathcal{P}(\Omega) = \{\emptyset, \{\omega_1\}, \ldots \{\omega_n\}, \{\omega_1\omega_2\}, \ldots, \Omega\}$$

   es una σ-álgebra de subconjuntos de Ω. Recordar que si $\#(\cdot)$ denota la cardinalidad de un conjunto, $\#(\Omega) = n$ y $\#(\mathcal{P}(\Omega)) = 2^n$.

4. Considere $\Omega = \mathbb{R}$. Los **conjuntos de Borel** o **Borelianos** de $\mathbb{R}$, denotados por $\mathcal{B}(\mathbb{R})$, constituye la σ-álgebra más pequeña que contiene todos los intervalos de la forma $(a, b]$, para $a < b$. De igual forma, note por ejemplo que

   $$(a, b) = \bigcup_{n=1}^{\infty} (a, b - 1/n]$$

   Así, los conjuntos Borelianos de $\mathbb{R}$, es la σ-álgebra generada por los intervalos abiertos (cerrados, semiabiertos) de $\mathbb{R}$.

5. Conjuntos como el *conjunto de Cantor* es un conjunto de Borel pero no todos los subconjuntos de $\mathbb{R}$ son miembros de $\mathcal{B}(\mathbb{R})$. La construcción de subconjuntos no Borelianos va más allá del interés de este curso.

Definición : Sea Ω uan colección de objetos (espacio muestral) y $\mathcal{S}$ una σ-álgebra de subconjuntos de Ω. Luego $(\Omega, \mathcal{S})$ se dice que es un **espacio medible** y los elementos de $\mathcal{S}$, **conjuntos medibles** o para los propósitos de este curso, **eventos**.

Ejemplo : Considere el experimento aleatorio $\mathcal{E}$, "lanzar un dado honesto" y se fija uno en la cara que queda hacia arriba. El conjunto de posibles salidas, el espacio muestral, $\Omega = \{1,2,3,4,5,6\}$. Sea el conjunto potencia de Ω como la familia de eventos. Esto es, $\mathcal{S} = \mathcal{P}(\Omega)$ es la σ-álgebra de subconjuntos de Ω. Considere los eventos $A = \{\text{múltiplo de } 3\} = \{3.6\}$ y $B = \{\text{par}\} = \{2,4,6\}$ y $A, B \subseteq \Omega$. Entonces, *"A o B ocurre"* cuando $A \cup B = \{2,3,4,6\}$ ocurre. Y similarmente, *"A y B ocurre"* cuando $A \cap B = \{6\}$ ocurre.

Ejemplo : Probabilidad clásica. Bajo la definición de *probabilidad clásica*, cada punto es igualmente posible. En el ejemplo del dado honesto, cada uno de las caras tienen la misma probabilidad asociada 1/6. Así, si $C \subseteq \Omega$,

$$\mathbb{P}(C) = \sum_{\omega_i \in C} \mathbb{P}(\{\omega_i\}) = \frac{n(C)}{n(\Omega)}$$

donde $n(\cdot)$ es el número de elementos en el conjunto. Luego,

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(\{3,6\}) = \mathbb{P}(\{3\}) + \mathbb{P}(\{6\}) = \frac{2}{6} \\
\mathbb{P}(A \cup B) &= \mathbb{P}(\{2,3,4,6\}) = \frac{n(A \cup B)}{n(\Omega)} = \frac{4}{6} \\
\mathbb{P}(A \cap B) &= \mathbb{P}(\{6\}) = \frac{n(A \cap B)}{n(\Omega)} = \frac{1}{6}
\end{aligned}$$

Ejemplo : Probabilidad como frecuencia relativa. Este ejemplo fue tomado de Hoel et al. (1971).

Se considera un material radioactivo y se está interesado en el decaimiento de un átomo a un isótopo estable ($\mathcal{E}$). Leyes físicas indican que es imposible saber el tiempo en que un átomo en particular habrá en decaer. Una vez que el átomo decae se queda en ese estado, por lo que el experimento no se puede repetir. Sin embargo es posible observar un conjunto de átomos y considerarlos cada uno de ellos como la repetición del experimento $\mathcal{E}$.

Sea $N_0 = N(0)$ el número de átomos inicial y $N(t)$ el número de átomos que no ha decaído al tiempo $t(> 0)$. Se está interesado en predecir la proporción de átomos $N(t)/N_0$ que no habría decaído al tiempo t.

Suponga que la tasa de decaimiento es proporcional al número de átomos que no ha decaído

$$\frac{dN(t)}{dt} = -\lambda N(t)$$

con $N_0 = N(0)$, $\lambda > 0$ fija, la constante de proporcionalidad. Le ecuación diferencial tiene como solución (única)

$$N(t) = N_0 e^{-\lambda t}, \text{ para } t > 0$$

Por lo que la proporción de átomos que no ha decaído es $\dfrac{N(t)}{N_0} = e^{-\lambda t}$. Luego, si $0 < t_1 < t_2$, la proporción de átomos que decayó en el intervalo (t_1, t_2) es $(e^{-\lambda t_1} - e^{-\lambda t_2})$.

En la construcción del *modelo probabilístico* se consideró el espacio de posibles resultados $\Omega = [0, \infty)$. En este caso, el espacio muestral Ω es infinito no numerable. Además, si $t_1 = t = t_2$, la probabilidad de que un átomo decaiga exactamente al tiempo t es, $e^{-\lambda t_1} - e^{-\lambda t_2} = 0$. Esto es, la probabilidad del evento $\{\mathrm{t}\}$ es cero. Note que $\mathbb{P}(\Omega) = e^{-\lambda(0)} - e^{-\lambda(\infty)} = 1 - 0 = 1$.

Finalmente, si A y B son dos intervalos disjuntos (ajenos), la proporción de átomos que decaen en A o B es la suma de las proporciones que caen en A o los que decaen en B. Esto es, se espera la aditividad de la probabilidad en conjuntos (eventos) ajenos. Así, si $A, B \in \mathcal{S}$, tales que $A \cap B = \emptyset$, entonces $\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Uno o una está interesado en determinar o medir la *incertidumbre*, *posibilidad* o *probabilidad* de la ocurrencia o no de eventos.

Definición : Sea Ω un espacio muestral, $\mathcal{S}$ una σ-álgebra de subconjuntos de Ω. Una **medida de probabilidad** $\mathbb{P}$ sobre $\mathcal{S}$ es una función real, con dominio en $\mathcal{S}$, tal que:

$\mathbf{K_1}$: $\mathbb{P}(B) \geq 0$, para todo $B \in \mathcal{S}$.

$\mathbf{K_2}$: $\mathbb{P}(\Omega) = 1$.

$\mathbf{K_3}$: Si $A_1, A_2, \dots \in \mathcal{S}$, tales que $A_i \cap A_j = \emptyset$, para $i \neq j$, entonces,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Definición : La triada $(\Omega, \mathcal{S}, \mathbb{P})$ se dice un **espacio de probabilidad** y las propiedades K_1, K_2 y K_3 anteriores se conocen como los **axiomas de probabilidad** o **axiomas de Kolmogorov** (Andréi Kolmogórov, 1933).

El problema de dado un espacio muestral Ω, la construcción de una σ-álgebra de conjuntos $\mathcal{S}$ y una medida de probabilidad $\mathbb{P}$ es materia de **Teoría de la Medida**, más allá de las pretensiones de este curso. La Teoría axiomática de la probabilidad parte de espacios abstractos de probabilidad para su desarrollo. En este curso y en la mayoría de las aplicaciones los espacios de probabilidad son poco complicados, bien conocidos: $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, y las familias de eventos, $\mathcal{P}(\mathbb{N})$, $\mathcal{P}(\mathbb{Z})$ y $\mathcal{B}(\mathbb{R})$, los conjuntos borelianos como familia de eventos.

Ejemplo :

a). Considere nuevamente el ejemplo del lanzamiento de un dado. Ahí, $\Omega = \{1,\dots,6\}$, $\mathcal{S} = \mathcal{P}(\Omega)$, la familia de todos los subconjuntos de Ω ($2^6 = 64$ eventos) y utilizando la probabilidad clásica, para todo $A \in \mathcal{S}$ se define

$$\mathbb{P}(A) = \frac{n(A)}{n(\Omega)}$$

b). En el ejemplo del decaimiento radioactivo, $\Omega = (0, \infty)$. Hay varias σ-álgebras posibles, pero cualquiera de ellas que se elija debe cumplir que para cualquier intervalo (t_1, t_2),

$$\mathbb{P}\big((t_1, t_2)\big) = e^{-\lambda t_1} - e^{-\lambda t_2}, \quad \text{donde } 0 \leq t_1 < t_2$$

1.2. Espacios de probabilidad uniforme (*"puntos equiprobables"*)

La siguiente idea de espacios uniformes se presenta conforme a Hoel et al. (1971). En palabras, considerar situaciones donde *todas las salidas ω's son igualmente posibles.*

- **Caso discreto.** $\Omega = \{\omega_1, \ldots, \omega_n\}$. *"Elegir un punto al azar"* , es suponer que la salida de todos y cada una de las salidas ω es igualmente posible, por lo que $\mathbb{P}(\{\omega_i\}) = 1/n$, para todo $\omega_i \in \Omega$. Sea $\mathcal{S} = \mathcal{P}(\Omega)$, $A \in \mathcal{S}$. Luego,

$$\mathbb{P}(A) = \frac{\#(A)}{\#(\Omega)} = \frac{n(A)}{n(\Omega)}$$

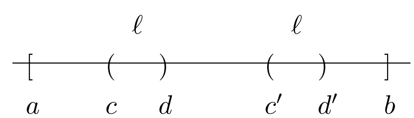
  donde $\#(A)$ denota la cardinalidad del evento, en este caso el número (finito) de elementos de A $n(A)$. Entonces, *eventos con igual número de puntos tienen la misma probabilidad.*

- **Caso continuo unidimensional.** Suponga $a < b$ y $\Omega = [a, b]$. Se puede pensar *"el tamaño de Ω"* como la longitud de Ω, luego $|\Omega| = \text{longitud}([a, b]) = b - a$. Entonces cualesquiera dos subintervalos A y B tendrán el mismo tamaño si tienen la misma longitud, digamos, ℓ y

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\ell}{b-a}$$

  Así subintervalos de $[a, b]$ con *la misma longitud tendrán la misma probabilidad.* Si $d - c = \ell = d' - c'$,

$$\mathbb{P}\big((c, d)\big) = \frac{\ell}{b-a} = \mathbb{P}\big((c', d')\big)$$



  Esto es, en los espacios uniformes, intervalos de igual longitud tienen asociadas la misma probabilidad.

- **Caso continuo r-dimensional.** Suponga $\Omega \subset \mathbb{R}^r$, tal que $|\Omega| = \text{volumen}(\Omega) < \infty$. Para $A \subseteq \Omega$, por ejemplo, $A \in \mathcal{B}(\mathbb{R}^r)$,

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\text{vol}\,(A)}{\text{vol}\,\Omega)}$$

  Así, *regiones con el mismo volumen tiene la misma probabilidad.*

Ejemplo : Considere la lotería con 20 boletos igualmente probables: $\Omega = \{1, 2, \ldots, 20\}$ y los eventos $A = \{\text{Números menores a } 5\}$, $B = \{\text{Números primos}\}$ y $C = \{\text{Múltiplos de } 5\}$. Entonces, $A = \{1, 2, 3, 4\}$, con $n(A) = 4$, $B = \{2, 3, 5, 7, 11, 13, 17, 19\}$ con $n(B) = 8$ y $C = \{5, 10, 15, 20\}$ con $n(C) = 4$.

Bajo la *probabilidad uniforme*: $\mathbb{P}(E) = n(E)/n(\Omega)$, para todo $E \subseteq \Omega$. Entonces, $\mathbb{P}(\Omega) = n(\Omega)/n(\Omega) = 1$; $\mathbb{P}(A) = n(A)/n(\Omega) = 4/20$; $\mathbb{P}(B) = n(B)/n(\Omega) = 8/20$; $\mathbb{P}(C) = n(C)/n(\Omega) = 4/20$.

Además, $A \cup B = \{1, 2, 3, 4, 5, 7, 11, 13, 17, 19\}$, con $n(A \cup B) = 10$ y $A \cap B = \{2, 3\}$, con $n(A \cap B) = 2$; $A \cup C = \{1, 2, 3, 4, 5, 10, 15, 20\}$, con $n(A \cup C) = 8$ y $A \cap C = \emptyset$ con $n(A \cap C) = 0$; $B \cup C = \{2, 3, 5, 7, 10, 11, 13, 15, 17, 19, 20\}$ con $n(B \cup C) = 11$ y $B \cap C = \{5\}$ con $n(B \cap C) = 1$. Luego, se satisface

$$\begin{aligned}
\mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 4/20 + 8/20 - 2/10 = 10/20 \\
\mathbb{P}(A \cup C) &= \mathbb{P}(A) + \mathbb{P}(C) - \mathbb{P}(A \cap C) = 4/20 + 4/20 - 0 = 8/20 \\
\mathbb{P}(B \cup C) &= \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(B \cap C) = 8/10 + 4/10 - 1/10 = 11/10
\end{aligned}$$

Ejemplo : Considere ahora el intervalo $\Omega = [0, 1]$ y los subintervalos $A = (1/10, 4/10)$, $B = (3/10, 8/10)$ y $C = (7/10, 9/10)$. Si consideramos la medida de los subintervalos como su longitud, $|\Omega| = \ell(\Omega) = 1$ y, nuevamente, bajo la *probabilidad uniforme*, la probabilidad del subintervalo $\mathbb{P}(I) = |I|/|\Omega| = \ell(I)$. Entonces, $\mathbb{P}(\Omega) = \ell([0,1]) = 1$, $\mathbb{P}(A) = \ell(A) = 3/10$, $\mathbb{P}(B) = \ell(B) = 5/10$ y $\mathbb{P}(C) = \ell(C) = 2/10$. Entonces,

$$\begin{aligned}
\mathbb{P}(A \cup B) &= \ell\big((1/10, 8/10)\big) = 7/10 \\
\mathbb{P}(A \cap B) &= \ell\big((3/10, 4/10)\big) = 1/10 \\
\mathbb{P}(A \cup C) &= \ell\big((1/10, 4/10)\big) + \ell\big((7/10, 9/10)\big) = 3/10 + 2/10 = 5/10 \\
\mathbb{P}(A \cap C) &= \mathbb{P}(\emptyset) = 0 \\
\mathbb{P}(B \cup C) &= \ell\big((3/10, 9/10)\big) = 6/10 \\
\mathbb{P}(B \cap C) &= \ell\big((8/10, 9/10)\big) = 1/10
\end{aligned}$$

Y se cumple que

$$\begin{aligned}
\mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 3/10 + 5/10 - 1/10 = 7/10 \\
\mathbb{P}(A \cup C) &= \mathbb{P}(A) + \mathbb{P}(C) - \mathbb{P}(A \cap C) = 3/10 + 2/10 - 0 = 5/10 \\
\mathbb{P}(B \cup C) &= \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(B \cap C) = 5/10 + 2/10 - 1/10 = 6/10
\end{aligned}$$

1.3. Corolarios a los axiomas de probabilidad

Sean Ω un espacio muestral, $\mathcal{S}$, σ-álgebra de subconjuntos de Ω, la familia de eventos y $\mathbb{P}$ una medida de probabilidad de manera que $(\Omega, \mathcal{S}, \mathbb{P})$ es un espacio de probabilidad.

Para los siguientes resultados considere $A, B, A_1, A_2, \cdots \in \mathcal{S}$ y note que para todo evento $A \in \mathcal{S}$, $\Omega = A \cup A^C$, unión de eventos ajenos.

Proposición :

1. Si $A \subseteq B$,
$$\mathbb{P}(A) \leq \mathbb{P}(B)$$
$\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c) \geq \mathbb{P}(B \cap A) + 0 = \mathbb{P}(A)$, pues $\mathbb{P}(B \cap A^C) \geq 0$, por K_1.

2. $$\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$$
$1 \overset{K_2}{=} \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^C) \overset{K_3}{=} \mathbb{P}(A) + \mathbb{P}(A^C)$, y se sigue el resultado.

3. Para todos A y B,
$$\mathbb{P}(B \backslash A) = \mathbb{P}(B) - \mathbb{P}(B \cap A)$$
donde $B \backslash A = (B \cap A^C)$ denota la diferencia de conjuntos. $B = B \cap \Omega = B \cap (A \cup A^C) = (B \cap A) \cup (B \cap A^C)$, unión de eventos disjuntos. Luego, se sigue del tercer axioma de probabilidad, K_3, que $\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \backslash A)$ y de ahí el resultado.

4. $$\mathbb{P}(\emptyset) = 0$$
$1 \overset{K_2}{=} \mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \emptyset) \overset{K_3}{=} \mathbb{P}(\Omega) + \mathbb{P}(\emptyset)$, y se tienes el resultado.

5. Para todo A,
$$0 \leq \mathbb{P}(A) \leq 1$$
se sigue K_1, K_2 y de el resultado C_1 anterior pues $\emptyset \subseteq A \subseteq \Omega$.

6. Sean $A_1, \ldots, A_n \in \mathcal{S}$, eventos mutuamente ajenos, es decir, $A_i \cap A_j = \emptyset$, para $i \neq j$. Entonces,
$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$
Para ver esto, defina $\emptyset = A_{n+1} = A_{n+2}, \ldots$ y aplique K_3 el resultado 4.

   Nota: De las leyes de De Morgan se sigue que $A_1, A_2, \cdots \in \mathcal{S}$, entonces

   i) $\left(\cup_{i=1}^{\infty} A_i\right)^C = \cap_{i=1}^{\infty} A_i^C \in \mathcal{S}$
   ii) $\left(\cap_{i=1}^{\infty} A_i\right)^C = \cup_{i=1}^{\infty} A_i^C \in \mathcal{S}$

   En efecto,

   i) Si $\omega \in \left(\cup_{i=1}^{\infty} A_i\right)^C$, $\omega \notin \cup_{i=1}^{\infty} A_i \Longrightarrow \omega \notin A_i$ para ningún $i = 1, 2, \ldots$. Luego, $\omega \in A_i^C$ para todo i y por lo tanto, $\omega \in \cap_{i=1}^{\infty} A_i^C$.
   ii) Si $\omega \in \left(\cap_{i=1}^{\infty} A_i\right)^C \Longrightarrow \omega \notin \left(\cap_{i=1}^{\infty} A_i\right)$. Luego ω no está en al menos un A_j, por lo que $\omega \in A_j^C$, para algún A_j. Por lo tanto $\omega \in \cup_{i=1}^{\infty} A_j^C$.

7.
$$\mathbb{P}\left(\text{Al menos uno del los } A_i \text{ ocurre}\right) = 1 - \mathbb{P}\left(\text{Ninguno de los } A_i \text{ ocurre}\right)$$
$\cup_{i=1}^{\infty} A_i = \{\text{Al menos uno del los } A_i \text{ ocurre}\}$, $\cap_{i=1}^{\infty} A_i^C =$ Ninguno de los A_i ocurre. Se sigue entonces de las leyes de De Morgan que $\mathbb{P}\left(\cup_{i=1}^{\infty} A_i\right) = 1 - \mathbb{P}\left(\cap_{i=1}^{\infty} A_i^C\right)$.

8.
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$
$A \cup B = (A \backslash B) \cup (A \cap B) \cup (B \backslash A)$, eventos ajenos. Luego,
$$\begin{aligned} \mathbb{P}(A \cup B) &\overset{K_3}{=} \mathbb{P}(A \backslash B) + \mathbb{P}(A \cap B) + \mathbb{P}(B \backslash A) \\ &\overset{C_3}{=} \left[\mathbb{P}(A) - \mathbb{P}(A \cap B)\}\right] + \mathbb{P}(A \cap B) + \left[\mathbb{P}(B) - \mathbb{P}(A \cap B)\right] \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \end{aligned}$$

9. **Desigualdad de Bonferroni**.
$$\mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1$$
Se sigue del punto anterior que $\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B)$ y puesto que $\mathbb{P}(A \cup B) \leq 1$, se tiene la desigualdad.

10.
$$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$$

11. **Desigualdad de Boole**. Para todo $A_1, A_2, \cdots \in \mathcal{S}$,
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$
Para ver el resultado, sean $B_1 = A_1, B_2 = A_2 \backslash A_1, \ldots, B_n = A_n \backslash \cup_{i=1}^{n-1} A_i$. Entonces, para todo $n > 0$, los eventos B_n son ajenos y $\cup_{n=1}^{\infty} B_n = \cup_{i=1}^{\infty} A_i$, por lo que de K_3 se sigue que $\mathbb{P}\left(\cup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\cup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(B_n)$. Pero $B_n = A_n \backslash \cup_{i=1}^{n-1} \subseteq A_n \Longrightarrow \mathbb{P}(B_n) \leq \mathbb{P}(A_n)$. Por lo tanto,
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

12. **Principio de inclusión–exclusión en probabilidad**

$$\mathbb{P}\left(\cup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1\leq i<j\leq n} \mathbb{P}(A_i \cdot A_j) + \sum_{1\leq i<j<k\leq n} \mathbb{P}(A_i \cdot A_j \cdot A_k) + \cdots + (-1)^n \mathbb{P}(A_1 \cdots A_n)$$

donde los puntos $\cdot$ representan intersecciones de conjuntos.

$k = 2$: Se sigue de C_8.

$k = 3$:

$$\begin{aligned}
\mathbb{P}(A_1 \cup A_2 \cup A_3) =& \mathbb{P}\left((A_1 \cup A_2) \cup A_3)\right) \\
=& \mathbb{P}(A_1 \cup A_2) + \mathbb{P}(A_3) - \mathbb{P}\left((A_1 \cup A_2) \cap A_3\right) \\
=& \left[\mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3)\right] - \left[\mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_1 \cap A_3) + \mathbb{P}(A_2 \cap A_3)\right] \\
& + \mathbb{P}(A_1 \cap A_2 \cap A_3)
\end{aligned}$$

$k = n$: Por inducción.

Teorema de la continuidad de la probabilidad

Recuerde, h función real se dice una **función continua** en $x_0 \in \mathbb{R}$ si $\lim_{x\to x_0} h(x) = h(x_0)$. Que podría escribirse $\lim_{x\to x_0} h(x) = h(\lim_{x\to x_0}\ x)$. De manera similar, se verá que si $A, A_1, A_2, \cdots \in \mathcal{S}$ y $A_n \nearrow A$, $(A_n \searrow A)$, $\mathbb{P}(A_n) \longrightarrow \mathbb{P}(A)$.

Teorema : Sea $(\Omega, \mathcal{S}, \mathbb{P})$ un espacio de probabilidad.

i) Si $A_1 \subseteq A_2 \subseteq \cdots \in \mathcal{S}$ y $A = \cup_{i=1}^{\infty} A_i \in \mathcal{S}$, entonces $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

ii) Si $A_1 \supseteq A_2 \supseteq \cdots \in \mathcal{S}$ y $A = \cap_{i=1}^{\infty} A_i \in \mathcal{S}$, entonces $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

Demostración:

i) Sean $A_1, A_2, \cdots \in \mathcal{S}$, con $A_1 \subseteq A_2 \subseteq \cdots$ y $\cup_{i=1}^{\infty} A_i = A$. Sea $B_1 = A_1$ y para $n \geq 2$, $B_n = A_n \backslash A_{n-1} (= A_n \cap A_{n-1}^C)$. Note que $\omega \in B_n$ si y solo si $\omega \in A_n$ y A_n es el primer evento que contiene a ω de la sucesión $\{A_i\}$. Por construcción los B_n son eventos disjuntos y $A_n = \cup_{i=1}^{n} B_i$. Luego, $\mathbb{P}(A_n) = \mathbb{P}(\cup_{i=1}^{n} B_i) = \sum_{i=1}^{n} \mathbb{P}(B_i)$ por C_6, y como $A = \cup_{n=1}^{\infty} B_n$, se sigue $\mathbb{P}(A) = \mathbb{P}\left(\cup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(B_n)$. Por lo tanto, $\lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(B_i) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) = \mathbb{P}(A)$. Esto es,

$$\lim_{n\to\infty} \mathbb{P}(\cup_{i=1}^{n} A_i) = \mathbb{P}\left(\lim_{n\to\infty} \cup_{i=1}^{n} A_i\right)$$

ii) Suponga ahora que $A_1 \supseteq A_2 \supseteq \cdots \in \mathcal{S}$ y $A = \cap_{n=1}^{\infty} A_n$. Luego, $A_1^C \subseteq A_2^C \subseteq \cdots$ y $A^C = \cup_{n=1}^{\infty} A_n^C$. Se sigue del inciso anterior que $\lim_{n\to\infty} \mathbb{P}(A_n^C) = \mathbb{P}(A^C) = 1 - \mathbb{P}(A)$, Luego $\lim_{n\to\infty} \left[1 - \mathbb{P}(A_n)\right] = 1 - \mathbb{P}(A)$ y por lo tanto $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$. Así,

$$\lim_{n\to\infty} \mathbb{P}\left(\cap_{i=1}^{n} A_i\right) = \mathbb{P}\left(\lim_{n\to\infty} \cap_{i=1}^{n} A_i\right)$$

1.4. Espacios de probabilidad no uniforme

Caso discreto.

Se vio ya que en el caso del espacio uniforme de probabilidad discreto necesariamente se tiene que $\#(\Omega) = n < \infty$ (vea problema 1.1). Si Ω tienen un número infinito de elementos el espacio no puede ser uniforme pues los elementos muestrales no pueden tener la misma probabilidad de ocurrencia.

Casos de espacios no uniformes son los más comunes. Considere por ejemplo el lanzamiento de una moneda cargado, por ejemplo en favor del *sol*. $\Omega = \{\text{sol, águila}\}$, con $\mathbb{P}(\text{sol}) = 2\mathbb{P}(\text{águila})$. Luego $\mathbb{P}(\text{sol}) = 2/3$ y $\mathbb{P}(\text{águila}) = 1/3$.

Considere ahora el lanzamiento de un *dado honesto*. Experimento aleatorio $\mathcal{E}$ consiste en lanzar un dado hasta que cae la cara con el 5 hacia arriba. Sea $\omega_i = \{\text{El primer 5 cae en el } i\text{-ésimo lanzamiento}\}$, con $i = 1, 2, \ldots$. Luego, $\Omega = \{\omega_1, \omega_2, \ldots\}$. Note, que $\mathcal{E}$ tiene un número infinito de salidas posibles y $\#(\Omega) = \infty$. En el lanzamiento del dado $\mathbb{P}\left(\text{sale 5}\right) = 1/6$ y $\mathbb{P}\left(\text{no sale 5}\right) = 5/6$. Luego, para $r = 1, 2, \ldots,$

$$\begin{aligned}
\mathbb{P}(\{w_r\}) =& \mathbb{P}\left(\text{sale 5 hasta el } r\text{-ésimo lanzamiento}\right) \\
=& \underbrace{\frac{5}{6}\cdot\frac{5}{6}\cdots\frac{5}{6}}_{r-1 \text{ términos}}\cdot\frac{1}{6} \\
=& \left(\frac{5}{6}\right)^{r-1}\left(\frac{1}{6}\right)
\end{aligned}$$

Nota:

i) $\mathbb{P}(\{\omega_i\}) \geq 0$, $i = 1, 2, \ldots$.

ii) $\mathbb{P}(\Omega) = 1$. A saber,

$$\mathbb{P}(\Omega) = \mathbb{P}\left(\cup_{i=1}^{\infty}\{\omega_i\}\right) \overset{K_3}{=} \sum_{i=1}^{\infty}\mathbb{P}\left(\{\omega_i\}\right) = \sum_{i=1}^{\infty}\left(\frac{5}{6}\right)^{i-1}\frac{1}{6} = \frac{1}{6}\sum_{i=0}^{\infty}\left(\frac{5}{6}\right)^{i} = 1$$

pues la suma geométrica $\displaystyle\sum_{i=0}^{\infty}\left(\frac{5}{6}\right)^{i} = \frac{1}{1-5/6} = 6.$

Que $\mathbb{P}(\Omega) = \mathbb{P}\left(\cup_{i=1}^{\infty}\{\omega_i\}\right) = 1$, se interpreta que con certeza caerá el 5 eventualmente.

Caso continuo.

Recuerde el ejemplo del decaimiento de partículas. Si $t_1 < t_2$, la probabilidad de emisión en el intervalo $[t_1, t_2]$ es $\mathbb{P}\left([t_1, t_2]\right) = e^{-\lambda t_1} - e^{-\lambda t_2}$. En este ejemplo, $\Omega = [0, \infty)$.

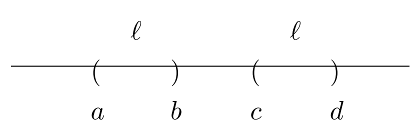
Nota:

i) Para todo $a < b$, $\mathbb{P}\left([a, b]\right) = e^{-\lambda a} - e^{-\lambda b} > 0$.

ii) $\mathbb{P}(\Omega) = \mathbb{P}(\mathbb{R}^+) = \text{lím}_{b\to\infty}\mathbb{P}\left([0, b]\right) = \text{lím}_{b\to\infty}\left(e^{-\lambda 0} - e^{-\lambda b}\right) = 1$.

iii) Para $a < b$ y $c < d$ con $b - a = \ell = d - c$, se tiene que

$$\begin{aligned}
\mathbb{P}\left((a, b)\right) &> \mathbb{P}\left((c, d)\right) \\
e^{-\lambda a} - e^{-\lambda b} &> e^{-\lambda c} - e^{-\lambda d}
\end{aligned}$$



Esto es, a diferencia de los espacios uniformes, intervalos de la misma longitud tienen asociadas distintas probabilidades.

Ejemplo : El siguiente es el problema 2.119 de Wackerly et al. (2008). Suponga que se lanzan repetidamente dos dados y se cuenta la suma de las caras hacia arriba en cada lanzamiento. Determine la probabilidad de los siguientes eventos:

a). La suma 3 sale antes que la suma 7.

b). La suma 4 sale antes que la suma 7.

Solución:

a). Considere los eventos:

$$\begin{aligned} S_j &= \{\text{Suma es j}\} \\ T_k &= \{\text{El juego termina en el ensayo } k\} \\ G_3 &= \{\text{Sale el 3 antes que el 7}\} \end{aligned}$$

Entonces,

$$G_3 = \bigcup_{k=1}^{\infty} (S_3 \cap T_k)$$

con los eventos $(S_3 \cap T_k)$ ajenos. Luego,

$$\mathbb{P}(G_3) = \sum_{k=1}^{\infty} \mathbb{P}(S_3 \cap T_k)$$

Ahora bien, $\mathbb{P}(S_3) = 2/36$ y $\mathbb{P}(S_7) = 6/36$. Entonces, $\mathbb{P}(\{\text{Suma 3 ó 7}\}) = 2/36+6/36 = 8/36$, por lo que la probabilidad de que el juego no se decida en un lanzamiento de los dados es $1 - 8/36 = 28/36$. Entonces,

$$\mathbb{P}(S_3 \cap T_k) = \underbrace{\left(\frac{28}{36}\right) \cdots \left(\frac{28}{36}\right)}_{k-1 \text{ términos}} \frac{2}{36} = \frac{2}{36} \left(\frac{28}{36}\right)^{k-1}$$

y por lo tanto, la probabilidad de que salga la suma 3 antes que la suma 7 es

$$\mathbb{P}(G_3) = \frac{2}{36} \sum_{k=1}^{\infty} \left(\frac{28}{36}\right)^{k-1} = \frac{2}{36} \cdot \frac{1}{1 - \frac{28}{36}} = \frac{2}{36} \frac{36}{8} = \frac{1}{4}$$

b). De manera similar, la probabilidad de que salga primero S_4 que suma S_7 es

$$\mathbb{P}(G_4) = \frac{3}{36} \sum_{k=1}^{\infty} \left(\frac{27}{36}\right)^{k-1} = \frac{3}{36} \cdot \frac{1}{1 - \frac{27}{36}} = \frac{3}{36} \cdot \frac{36}{9} = \frac{1}{3}$$

1.5. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 1, Barrios and Heiras (2024).

2. Técnicas de Conteo

Esta sección se basa fundamentalmente en el capítulo 2 de Hoel et al. (1971).

Sea $(\Omega, \mathcal{S}, \mathbb{P})$ un espacio de probabilidad uniforme discreto. Luego Ω tiene un número finito de puntos, por ejemplo, $\Omega = \{\omega_1, \ldots, \omega_n\}$. Sea además $\mathcal{S} = \mathcal{P}(\Omega)$, el conjunto potencia, σ-álgebra de subconjuntos de Ω y $\mathbb{P}$, la medida de probabilidad clásica con $\mathbb{P}(\{\omega_i\}) = 1/n$, $i = 1, \ldots, n$. Entonces, para el evento $B = \left\{\omega'_1, \ldots, \omega'_k\right\}$,

$$\mathbb{P}(B) = \sum_{i=1}^{k} \mathbb{P}(\left\{\omega'_i\right\}) = k \cdot \frac{1}{n} = \frac{k}{n}$$

Así, para todo evento $A \in \mathcal{S}$,

$$\mathbb{P}(A) = \frac{n(A)}{n}$$

donde $n(A)$ denota el número (finito) de elementos del conjunto A. Luego, el cálculo de la probabilidad de cualquier evento se reduce a contar el número de elementos del conjunto, de ahí el título de la sección.

2.1. Muestras ordenadas

Suponga que se tienen r conjuntos S_i, con n_i elementos respectivamente, $i = 1, \ldots, r$. Luego, el número de r-úplas $(s_1, \ldots, s_r)$, con $s_i \in S_i$, es

$$\#\left\{(s_1, \ldots, s_r) : s_i \in S_i\right\} = n_1 \cdots n_r$$

En particular, cuando $S_i = S$, para $i = 1, \ldots, r$, entonces el número de r-uplas es n^r, donde $n = \#(S)$.

Ejemplo : Considere 3 monedas $S_i = \{a, s\}$, $i = 1, 2, 3$. $\Omega = \{aaa, aas, \ldots, sss\}$, $n(\Omega) = 2^3 = 8$. Ω se puede ver alternativamente como el resultado de lanzar 3 veces una moneda.

Suponga una urna con n pelotas numeradas $1, 2, \ldots, n$. Se extrae una pelota, se registra su número y es devuelta a la urna. El número total de extracciones es r. Luego $\Omega = \left\{(x_1, \ldots, x_r) : x_i \in S\right\}$, $S = \{s_1, \ldots, s_n\}$. La r-upla $(x_1, \ldots, x_r)$ se dice una *muestra de tamaño r de una población de n objetos*. Si se supone que cada una de las n^r r-uplas son igualmente probables, el procedimiento se conoce como **muestro aleatorio con reemplazo**.

Note que el número de ordenamientos de r elementos de n posibles es n^r, en ocasiones denotado por ${}_n\mathrm{O}_r$, esto es, ${}_n\mathrm{O}_r = n^r$.

Ejemplo : Se lanza una moneda n veces. ¿Cuál es la probabilidad de obtener al menos un águila?

Solución: Sean los eventos $A_i = \left\{\text{Águila en el } i\text{-ésimo lanzamiento}\right\}$,
$A = \{\text{Al menos un águila}\} = \cup_{i=1}^{n} A_i$ y $A^C = \{\text{Todos son sol}\} = \{ss\cdots s\} = \cap_{i=1}^{n} A_i^C$. Luego,

$$\mathbb{P}(A) = \mathbb{P}(\cup_{i=1}^{n} A_i) = 1 - \mathbb{P}(\cap_{i=1}^{n} A_i^C) = 1 - (1/2)^n$$

pues $\mathbb{P}(\{ss\cdots s\}) = (1/2)^n$.

Suponga ahora que se tiene $S = \{s_1, \ldots, s_n\}$. Se extrae una pelota pero no se regresa a la urna. El procedimiento se repite hasta haber extraído r bolas, $r = 1, \ldots, n$. Al final se tiene una r-upla, $(x_1, \ldots, x_r)$, $x_i \in S$, $x_i \neq x_j$, para $i \neq j$. Luego,

$$\Omega = \{(x_1, \ldots, x_r), x_i \neq x_j\} \quad y \quad \#(\Omega) = n(n-1)\cdots(n-r+1)$$

pues la primera bola se elige de n posibles; la segunda de $(n-1)$; $\ldots$; la r-ésima se elige de $(n-r+1)$ bolas posibles. Si todas las r-uplas tienen la misma probabilidad de ser elegidas el procedimiento se conoce como **muestreo aleatorio sin reemplazo** de muestras de tamaño r de una población de n objetos. Notación:

$$(n)_r \equiv \underbrace{n(n-1)\cdots(n-r+1)}_{r \text{ factores}}$$

Note que $(n)_n = n(n-1)\cdots 2\,1 = n!$.

La muestra de tamaño r de una población de n objetos equivale a escribir $n\;(n-1)\cdots(n-r+1)$ en cualquier orden. Esto es, es el número de distintas permutaciones (ordenaciones) de r objetos de n posibles, denotado en ocasiones por ${}_n\mathrm{P}_r$. Luego,

$$\begin{aligned}(n)_r &= n(n-1)\cdots(n-r+1)\\ &= \frac{n(n-1)\cdots(n-r+1)(n-r)(n-r-1)\cdots 2\,1}{(n-r)!}\\ &= \frac{n!}{(n-r)!}\end{aligned}$$

Suponga ahora una muestra con reemplazo de tamaño r de una población de n elementos. Sea $A = \{\text{muestra sin objetos repetidos}\}$. Entonces, $n(A) = {}_n\mathrm{P}_r = n(n-1)\cdots(n-r+1)$, mientras que el número total de muestras aleatorias con reemplazo es $n(\Omega) = {}_n\mathrm{O}_r = n^r$. Así,

$$\begin{aligned}\mathbb{P}(A) =& \frac{(n)_r}{n^r}\\ =& \frac{n}{n}\,\frac{n-1}{n}\cdots\frac{n-r+1}{n}\\ =& 1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{r-1}{n}\right)\end{aligned}$$

Ejemplo : Fechas no repetidas de cumpleaños de r personas.

Sea $B_r = \{\text{Cumpleaños no repetidos en } r \text{ personas}\}$, con $r = 1, 2, \ldots, n$ y $n = 365$. Entonces,

$$\begin{aligned}n(B_r) =& (n)_r = n(n-1)\cdots(n-r+1)\\ \mathbb{P}(B_r) =& \frac{(n)_r}{n^r} = 1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{r-1}{n}\right)\\ \mathbb{P}(B_r^C) =& 1-\mathbb{P}(B_r)\end{aligned}$$

con $B^C = \{\text{Al menos dos personas de las } r \text{ tienen el mismo cumpleaños}\}$.

La siguiente tabla muestra las probabilidades de coincidan o no al menos dos personas con el mismo cumpleaños.

r	5	10	15	20	25	30
$\mathbb{P}(B_r^C)$	0.027	0.117	0.253	0.411	0.569	0.706
$\mathbb{P}(B_r)$	0.973	0.883	0.747	0.589	0.431	0.294

Nota: Si el número o tamaño de la población es grande, no tiene mucha diferencia práctica si el muestreo se hace con o sin reemplazo. En efecto,

$$\lim_{n\to\infty} \frac{(n)_r}{n^r} = \lim_{n\to\infty} \left[1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{r-1}{n}\right)\right] = 1$$

Ejercicio : Muestre que

$$\left(1-\frac{r-1}{n}\right)^{r-1} \leq \frac{(n)_r}{n^r} \leq \left(1-\frac{1}{n}\right)^{r-1}$$

2.2. Permutaciones (muestras ordenadas)

Suponga n cajas y n pelotas distintas.

- ¿De cuántas formas diferentes se pueden acomodar una bola en cada caja? Respuesta: de $n! = n(n-1)\cdots 21$ maneras distintas. Decir que se colocan *"al azar"* es suponer que del total de $n!$ formas, todas ellas son igualmente posibles, con probabilidad $1/n!$.
- ¿Cuál es la probabilidad de $E = \{\text{La bola } i \text{ vaya a dar al cajón } j\}$? La respuesta es acomodar $(n-1)$ bolas en $(n-1)$ cajones. Luego, hay $(n-1)!$ maneras distintas. Por lo que el evento E se da con una probabilidad

  $$\mathbb{P}(E) = \frac{(n-1)!}{n!} = \frac{1}{n}$$

  Lo anterior es equivalente a acomodar n objetos y nos preguntamos por el objeto i en la posición j.
- ¿Cuál es la probabilidad de k bolas en k cajones específicos? Respuesta: $(n-k)!/n!$.
- Por ejemplo, considere un mazo de cartas $1, 2, \ldots n$ revueltas. Se extrae una carta a la vez. Determine la probabilidad de que la i-ésima carta salga en la j-ésima extracción.

Ejercicio : 10 parejas llegan a un salón y se mezclan al azar. ¿Cuál es la probabilidad de que la chica i quede con el j muchacho?

2.3. Combinaciones (muestras no ordenadas)

Considere una mano de *poker*. Hay 52 cartas distintas (9 números: 2,3,...,10; 4 letras: J, Q, K, A; y cada número o letra de 4 distintas figuras: $\clubsuit, \diamondsuit, \heartsuit, \spadesuit$). Una *"mano de poker"* consiste en 5 cartas $(x_1, \ldots, x_5)$. Suponer que hay $(52)_5$ distintas manos es considerar, por ejemplo, que $(A_\clubsuit, K_\diamondsuit, Q_\heartsuit, J_\clubsuit, 10_\diamondsuit) \neq (J_\clubsuit, K_\diamondsuit, 10_\diamondsuit, Q_\heartsuit, A_\clubsuit)$. Para el propósito del juego, son la misma mano y de hecho, para cada mano hay 5!=120 maneras distintas de ordenar las cartas. Así, el número total de manos diferentes en un juego de poker es $\frac{(52)_5}{5!} = 2\,598\,960$.

Nota:

$$\frac{(52)_5}{5!} = \frac{52\cdot 51\cdot 50\cdot 49\cdot 48}{1\cdot 2\cdot 3\cdot 4\cdot 5} = \frac{(52)_r\, 47!}{5!\, 47!} = \frac{52!}{5!\,(52-5)!}$$

El cociente anterior se conoce como **coeficiente binomial** y se denota por

$$\binom{52}{5} = \frac{52!}{5!\,(52-5)!} = 2\,598\,960$$

En general, si se tiene un conjunto S con n elementos hay $(n)_r$ maneras distintas de ordenar r elementos a la vez. Cada una de estas r-uplas se puede ordenar de $r!$ formas diferentes. Entonces hay $(n)_r/r!$ maneras diferentes de *"combinar"* r elementos de n objetos distintos.

$$\frac{(n)_r}{r!} = \frac{(n)_r}{r!}\frac{(n-r)!}{(n-r)!} = \frac{n!}{r!\,(n-r)!} = \binom{n}{r} \equiv \quad \text{coeficiente binomial}$$

pues del **teorema binomial**

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Esto es, hay ${}_n\mathrm{C}_r = \binom{n}{r} = \dfrac{n!}{r!\,(n-r)!}$ distintas muestras sin reemplazo no ordenadas de tamaño r de una población de n objetos.

Nota: El coeficiente $\binom{a}{r}$ está bien definido para todos $a \in \mathbb{R}$ y $r \in \mathbb{N}_0 = \{0\} \cup \mathbb{N} = \{0,1,2,\ldots\}$. A saber,

$$\binom{a}{r} = \frac{(a)_r}{r!} = \frac{a(a-1)\cdots(1-r+1)}{r!}$$

con $0! = 1$ y $(a)_0 = 1$. Se tiene también, $\binom{a}{r} = 0$, si $r > a$.

Ejemplo : Sea $S = \{1, 2, \ldots, n\}$. Entonces, para $r = 1, 2, \ldots, n$ hay $\binom{n}{r}$ formas de elegir $i_1, \ldots, i_r \in S$ tales que $1 \leq i_1 < i_2 < \cdots < i_r \leq n$.

Solución:

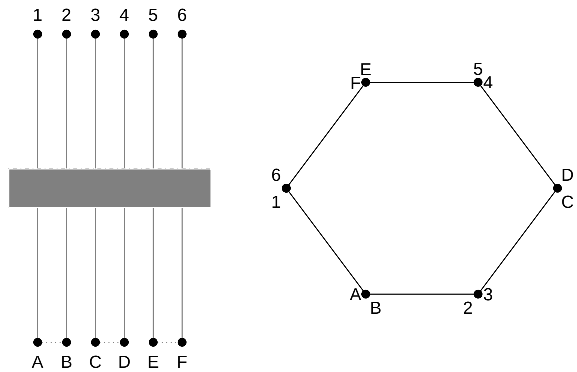
Solo hay $(n)_r$ maneras de elegir los r distintos números i_j y $r!$ formas distintas de ordenarlas pero solamente una de ellas es tal que $i_1 < i_2 < \cdots < i_r$. Entonces hay $\binom{n}{r} = \frac{(n)_r}{r!}$ distintas selecciones.

Ejemplo : Suponga que hay 25 profesores de tiempo completo, 15 de medio tiempo y 35 profesores por asignatura. Se forma un comité de 6 miembros seleccionados al azar. Determine la probabilidad de que todos los profesores sean de asignatura.

Solución: Hay un total de $n = 25 + 15 + 35 = 75$ profesores. Los miembros del comité se pueden elegir de $\binom{75}{6}$ maneras distintas. Ahora, hay 35 profesores de asignatura y $\binom{35}{6}$ formas distintas de elegir a 6 de ellos y $\binom{40}{0}$ entre los otros. Por lo tanto, la probabilidad de que todos los miembros del comité sean de asignatura es

$$\frac{\binom{35}{6}\binom{40}{0}}{\binom{75}{6}} = \frac{35!\;69!}{75!\;29!} = 0.0081$$

Ejercicio : El problema de los estambres rusos consiste, como yo lo entendí, en lo siguiente: se tienen 6 tiras de estambre indistinguibles y se sujetan todos ellos juntos con el puño cerrado de manera que no se sabe qué estambre es cuál. Simplemente se ven 6 extremos de tiras de estambre por arriba y por abajo del puño. Esta situación trata de representarse en la figura de abajo a la izquierda, aunque se han identificados los distintos estambres por número y letra.



Ahora bien, hay que anudar los estambres a pares, arriba y abajo como se muestra en la imagen de la izquierda de manera que al abrir el puño los estambres formen un único ciclo, por ejemplo, como el mostrado arriba a la derecha. Así, determine la probabilidad de que al anudar los extremos quede formado un solo ciclo.

Ejemplo : Suponga n bolas distribuidas aleatoriamente en n cajas. Es posible más de una bola por caja. Hay n^n posibles arreglos. Determine la probabilidad de que solamente C_1 la caja 1 quede vacía.

Solución: Sea Ω el espacio muestral de todos los posibles arreglos. Luego, $n(\Omega) = n^n$. Si solamente una caja queda vacía, hay una de las $(n-1)$ cajas restantes con dos bolas y $(n-2)$ restante con una sola bola. Sea $B_j = \{\text{Caja } j,\ C_j, \text{ con 2 bolas y la caja 1 vacía}\}$. Sea $A_1 = \cup_{j=2}^{n} B_j$, con los $B_j's$ siendo eventos ajenos. Ahora, $n(B_j) = \binom{n}{2}(n-2)!$, el número de maneras de cómo seleccionar las dos bolas que van a la caja j y las $(n-2)$ bolas que no van a C_i ni C_j. Luego, $\mathbb{P}(B_j) = \binom{n}{2}(n-2)!/n^n$, y por lo tanto

$$p_1 = \mathbb{P}(A_1) = \mathbb{P}\left(\cup_{j=2}^{n} B_j\right) = \frac{(n-1)\binom{n}{2}(n-2)!}{n^n}$$

¿Cuál es la probabilidad de que solamente una caja quede vacía?.

Si $A = \{\text{Una sola caja quede vacía}\} = \cup_{i=1}^{n} A_i$, con los A_i's eventos ajenos, entonces

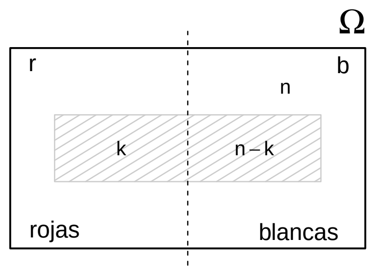
$$\mathbb{P}(A) = \frac{n(n-1)\binom{n}{2}(n-2)!}{n^n} = \frac{n!}{n^n}\binom{n}{2}$$

2.4. Particiones

Problema : Una urna tiene r bolas rojas y b bolas blancas. Se toma una muestra no ordenada sin reemplazo de tamaño n. ¿Cuál es la probabilidad p de que haya k bolas rojas y $(n-k)$ blancas en la muestra?

Solución: Sin importar el orden, una muestra de tamaño n sin reemplazo de $(r+b)$ objetos se puede tomar de $\binom{r+b}{n}$ posibles. Cada una de ellas con probabilidad $\binom{r+b}{n}^{-1}$ de ser elegida. Ahora, hay $\binom{r}{k}$ maneras de elegir las k bolas rojas y $\binom{b}{n-k}$ formas de elegir las blancas. Entonces, la probabilidad p pedida es

$$p = \frac{\binom{r}{k}\binom{b}{n-k}}{\binom{r+b}{n}}$$



En esta situación lo característico es reconocer que el espacio muestral Ω puede ser particionado en dos clases, en este ejemplo, bolas rojas y blancas.

Ejemplo : Suponga una caja con 50 tornillos, 10 de los cuales son defectuosos. De una muestra aleatoria de 5 tornillo, ¿cuál es la probabilidad que $k(= 0, 1, \ldots, 5)$ de ellos sean defectuosos? ¿De que haya al menos dos defectuosos?

Solución:

Hay $\binom{50}{5} = 2\,118\,760$ distintas muestras.
Para $k = 0, 1, \ldots, 5$,

$$\mathbb{P}\left(\{\text{k defectuos}\}\right) = \frac{\binom{10}{k}\binom{40}{5-k}}{\binom{50}{5}}$$

k	Prob
0	0.31056
1	0.43134
2	0.20984
3	0.04418
4	0.00396
5	0.00012

$$\mathbb{P}\left(\{\text{Al menos 2 defectuos}\}\right) = 1 - \Big[\mathbb{P}\left(\{\text{k=0}\}\right) + \mathbb{P}\left(\{\text{k=1}\}\right)\Big] = 0.2581$$

Ejemplo : Considere una partida de poker. Como se ha visto ya hay $\binom{52}{5} = 2\,598\,960$ distintas manos de poker. Sea $p = 1/2598960$

a). Determine la probabilidad de que en una mano haya 2 reyes (K) exactamente. – Se seleccionan 2 reyes de 4 disponibles y de las 48 cartas que no son reyes se elijen 3. Luego

$$\mathbb{P}\left(\{\text{Dos reyes exactamente}\}\right) = p\binom{4}{2}\binom{48}{3} = 103776p = 0.03993$$

b). *i)* Determine la probabilidad de que en una mano haya 3 tréboles (♣) exactamente. Hay $\binom{13}{3}\binom{39}{2} = 211926$ maneras de elegir tres tréboles y cualquier otro par de cartas. Luego,

$$\mathbb{P}\left(\{\text{Tres tréboles exactamente}\}\right) = p\binom{13}{3}\binom{39}{2} = 57798p = 0.08154$$

ii) Exactamente 3 cartas del mismo palo (misma figura) exactamente.

$$\mathbb{P}\left(\{\text{Tres cartas del mismo palo}\}\right) = p\binom{4}{1}\binom{13}{3}\binom{39}{2} = 847704p = 0.32617$$

c). Determine la probabilidad de una mano con dos pares, (x_1, x_2, y_1, y_2, z): Note que (x_1, x_2, y_1, y_2, z) es la misma mano que (y_1, y_2, x_1, x_2, z), por lo que considerar $\binom{13}{2}$ ignora el orden en que aparecen los dos pares. Luego,

$$\mathbb{P}\left(\{\text{Dos pares}\}\right) = p\binom{13}{2}\binom{4}{2}\binom{4}{2}\binom{11}{1}\binom{4}{1} = 123552p = 0.04754$$

Ejercicio : Considere un juego de poker. Hay cuatro figuras: corazones $\heartsuit$, diamantes $\diamondsuit$, tréboles $\clubsuit$ y espadas $\spadesuit$. Hay 9 números: $2, 3, \ldots, 9, 10$ y 4 letras J, Q, K, A. Una corrida de 5 cartas puede ser: $(A, 2, 3, 4, 5)$; $(2, 3, 4, 5, 6)$;$\ldots(10, J, Q, K, A)$.

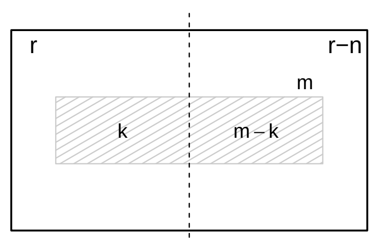
Calcule la probabilidad de cada una de las siguientes manos:

1. *Royal flush*: $(10, J, Q, K, A)$ del mismo tipo.
2. *Straight flush*: 5 cartas consecutivas del mismo tipo.
3. *Poker*: (x, x, x, x, y), con $x \neq y$.
4. *Full house*: de la forma (x, x, x, y, y), con $x \neq y$.
5. *Flush*: 5 cartas del mismo tipo.
6. *Straight*: 5 cartas consecutivas independientemente del tipo.
7. *Tercia*: (x, x, x, y, z), con $x \neq y, z$; $y \neq z$.
8. *Dos pares*: (x, x, y, y, z), con $x \neq y, z$; $y \neq z$.
9. *Un par*: (x, x, y, z, w), $x \neq y, z, w$; $y \neq z, w$; $z \neq w$.

Ejemplo : Una urna con r bolas numeradas $1, 2 \ldots, r$. Se toma una muestra sin reemplazo de tamaño n. Se regresa y se toma una segunda muestra ahora de tamaño m. Calcule la probabilidad que ambas muestras tengan exactamente k bolas en común.

Solución: La primera muestra define una partición de n elementos contra $(n - r)$ no seleccionados, luego, la probabilidad pedida es

$$p_k = \frac{\binom{n}{k}\binom{r-n}{m-k}}{\binom{r}{m}}$$



Considere ahora que la segunda muestra define la partición, luego la probabilidad de tener k elementos en común en ambas muestras es

$$p_k^* = \frac{\binom{m}{k}\binom{r-m}{n-k}}{\binom{r}{n}}$$

Ejercicio : Muestre que $p_k = p_k^*$.

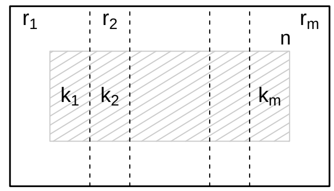
Ejercicio : Revise el problema de los alces de la lista 1.

Clases o categorías

Ahora se extiende la idea de partición a $m(> 2)$ subgrupos, clases o categorías. Suponga que se tienen r objetos donde cada uno de ellos es de uno y solo uno de los m tipos posibles. Si se toma una muestra de tamaño n, por determinar la probabilidad de que se elijan k_i objetos de la categoría C_j, $j = 1, \ldots, m$.

El total de distintas muestras $\binom{r}{n}$ y los k_j objetos de las clase C_j se pueden seleccionar de $\binom{r_j}{k_j}$ maneras para $j = 1, \ldots, m$. Así la probabilidad es

$$p = \frac{\binom{r_1}{k_1}\binom{r_2}{k_2}\cdots\binom{r_m}{k_m}}{\binom{r}{n}}$$



Ejemplo : En el ejemplo del comité de profesores determine la probabilidad de que 2 profesores sean de tiempo completo; 3 de medio tiempo; y 1 profesor de asignatura.

Solución:

$$p = \frac{\binom{25}{2}\binom{15}{3}\binom{35}{1}}{\binom{75}{6}} = 0.0210$$

Ejemplo : En un estacionamiento en este momento se tienen 25 autos de color gris; 18 de color negro; 16 blancos; 11 rojos; y 30 de otro color. Se toma una muestra aleatoria sin reemplazo de 10 automóviles. Determine la probabilidad de elegir 3 automóviles de color gris, 1 negro, 2 blancos, 0 rojos y 4 más de otro color.

Solución:

$$p = \frac{\binom{25}{3}\binom{18}{1}\binom{16}{2}\binom{11}{0}\binom{30}{4}}{\binom{100}{10}} = 0.007865$$

Ejercicio : Un cargamento tiene 20 paquetes de los cuales 7 están dañados. Los paquetes son inspeccionados uno a uno sin reemplazo hasta que se encuentra el cuarto paquete dañado. Calcule la probabilidad de que se requiera inspeccionar exactamente 12 paquetes.

Ejercicio : En un centro de cómputo, un servidor tiene 3 procesadores para recibir n tareas. Las tareas se asignan a los procesadores de manera aleatoria de tal forma que hay 3^n asignaciones posibles. Determina la probabilidad de que *exactamente* a uno de los procesadores no le sea asignada ninguna tarea.

2.5. Problemas de empates

El material que se presenta en esta sección se extrajo de Hoel et al. (1971)

Considere la permutación de n objetos. Se dice que *ocurrió un empate en la i-ésima posición* si el objeto i quedó en la posición i $(i = 1, \ldots, n)$. Sean $A_i = \{\text{Empate en posición } i\}$ y $A = \cup_{i=1}^{n} A_i = \{\text{Al menos un empate}\}$. Entonces, la probabilidad de al menos un empate está dada por la siguiente proposición.

Proposición : *Principio de inclusión–exclusión en probabilidad.*

$$\mathbb{P}\left(\cup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1\le i<j\le n} \mathbb{P}(A_i \cdot A_j) + \sum_{1\le i<j<k\le n} \mathbb{P}(A_i \cdot A_j \cdot A_k) + \cdots + (-1)^n \mathbb{P}(A_1 \cdots A_n) \quad (1)$$

donde los puntos $\cdot$ representan intersecciones de conjuntos.

$k = 2$: Se sigue del corolario C_8 de los axiomas de probabilidad.

$k = 3$:

$$\begin{aligned}\mathbb{P}(A_1 \cup A_2 \cup A_3) =& \mathbb{P}\big((A_1 \cup A_2) \cup A_3)\big) \\ =& \mathbb{P}(A_1 \cup A_2) + \mathbb{P}(A_3) - \mathbb{P}\big((A_1 \cup A_2) \cap A_3\big) \\ =& \big[\mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3)\big] - \big[\mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_1 \cap A_3) + \mathbb{P}(A_2 \cap A_3)\big] \\ &+ \mathbb{P}(A_1 \cap A_2 \cap A_3)\end{aligned}$$

$k = n$: Por inducción, siguiendo el mismo procedimiento se concluye que

$$\begin{aligned}\mathbb{P}\left(\cup_{i=1}^{n} A_i\right) =& \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1\le i<j\le n} \mathbb{P}(A_i \cdot A_j) + \sum_{1\le i<j<k\le n} \mathbb{P}(A_i \cdot A_j \cdot A_k) + \cdots \\ & \cdots + (-1)^n \mathbb{P}(A_1 \cdots A_n) \\ =& S_1 - S_2 + \cdots + (-1)^{n-1} S_n\end{aligned}$$

donde S_k representa la suma de probabilidades de la intersección de k eventos $A_{i_1} \cap \cdots \cap A_{i_k}$, donde $i_1 < \cdots < i_k$. Luego, S_k tiene $\binom{n}{k}$ sumandos.

Sea p_n la probabilidad de que no haya empates en las n posiciones. Luego, $1 - p_n = \mathbb{P}\left(\cup_{i=1}^{n} A_i\right)$, por lo que es necesario calcular $\mathbb{P}\left(A_{i_1} \cap \cdots \cap A_{i_k}\right)$ con $k = 1, \ldots, n$ y $i_1 < \cdots < i_k$, que como se vio en sección anterior esta probabilidad es $(n-k)!/n!$. Además, como la k-ésima suma S_k tiene $\binom{n}{k}$ términos,

$$\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^{n} (-1)^{k-1} \frac{1}{k!} = 1 - p_n$$

Por lo tanto, la probabilidad de no tener empates es

$$p_n = \sum_{k=0}^{n} (-1)^k \frac{1}{k!} \approx e^{-1} = 0.3679$$

y la probabilidad de al menos un empate $1 - p_n = 0.6321$. La siguiente tabla muestra las probabilidades de no empates y al menos un empate para varios números de casillas.

n	2	3	4	5	6	7	8	9
p_n	0.5000	0.3333	0.3750	0.3667	0.3681	0.3679	0.3679	0.3679
$1 - p_n$	0.5000	0.6667	0.6250	0.6333	0.6319	0.6321	0.6321	0.6321

Note que la probabilidad de al menos un empate en la permutación de n objetos es prácticamente independiente del número de objetos.

2.6. Cajas vacías

El material que se presenta en esta sección se extrajo de Hoel et al. (1971)

Para determinar la probabilidad $p_k(r,n)$ de exactamente k cajones vacíos al tirar n bolas en r casillas. Sea $A_i = \{\text{Cajón } i \text{ vacío}\}$ Luego, las n bolas deben acomodarse en las $(r-1)$ casillas restantes y esto sucede de $(r-1)^n$ maneras. Entonces,

$$\mathbb{P}(A_i) = \frac{(r-1)^n}{r^n} = \left(1 - \frac{1}{r}\right)^n$$

De manera simular, si $1 \leq i_1 < \cdots < i_k \leq r$, el evento $(A_{i_1} \cap \cdots \cap A_{i_k})$. ocurre si y solo si $(r-k)$ celdas reciben las n bolas. Así,

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \frac{(r-k)^n}{r^n} = \left(1 - \frac{k}{r}\right)^n$$

Luego,

$$S_k = \sum_{\{1 \leq i_1 < \cdots < i_k \leq r\}} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \binom{r}{k}\left(1 - \frac{k}{r}\right)^n$$

Por lo tanto,

$$\mathbb{P}\left(\cup_{k=1}^{r} A_k\right) = \sum_{k=1}^{r} (-1)^{k-1} \binom{r}{k} \left(1 - \frac{k}{r}\right)^n$$

Entonces, la probabilidad de tener todas las casillas ocupadas es

$$p_0(r,n) = 1 - \mathbb{P}\left(\cup_{k=1}^{r} A_k\right) = \sum_{k=0}^{r} (-1)^{k} \binom{r}{k} \left(1 - \frac{k}{r}\right)^n$$

Denote ahora $q_k(r,n) = \mathbb{P}\big(\{\text{Exactamente } k \text{ cajones } \textit{específicos} \text{ quedan vacíos}\}\big)$

Considere los primeros k cajones. El evento ocurre solo si las n pelotas van a las $(r-k)$ casillas restantes y ninguna de éstas queda vacía. Así el número de maneras para ocupar $(r-k)$ cajones con n bolas sin dejar uno vacío es $(r-k)^n p_0(r-k,n)$. Luego,

$$q_k(r,n) = \frac{(r-k)^n p_0(r-k,n)}{r^n} = \left(1 - \frac{k}{r}\right)^n p_0(r-k,n)$$

Además hay $\binom{r}{k}$ maneras de elegir los cajones vacíos, por lo que

$$p_k(r,n) = \binom{r}{k} q_k(r,n) = \binom{r}{k}\left(1 - \frac{k}{r}\right)^n p_0(r-k,n)$$

y sustituyendo la expresión para p_0

$$p_k(r,n) = \binom{r}{k} \sum_{j=0}^{r-k} (-1)^j \binom{r-k}{j} \left(1 - \frac{j+k}{r}\right)^n$$

para , $\quad k = 1, \ldots, r-1$.

Ejercicio : Problema de los cupones. Suponga que se inserta un cupón en cada caja de cereal. Suponga también que hay r distintos cupones y que todos son igualmente posibles de aparecer en cualquiera de las cajas. Si se compran n cajas de cereal. Calcule las siguientes probabilidades:

a) De haber obtenido al menos uno de cada tipo de cupón.

b) De no obtener exactamente k de los r tipos.

2.7. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 2, Barrios and Heiras (2024).

3. Probabilidad Condicional

3.1. Probabilidad condicional

1. Considere una caja con diez pelotas numeradas 1,2,...,10. Se extrae una pelota al azar.
   - ¿Cuál es la probabilidad de sacar el número 4?
   - ¿Cuál si se sabe que la pelota extraída fue un número par?
   - ¿Cuál si se sabe que fue un número primo?
2. Suponga una urna con r bolas rojas numeradas $1, 2, \ldots, r$ y b bolas blancas $1, 2, \ldots, b$.
   - ¿Cuál es la probabilidad del evento $A = \{\text{bola con número 1}\}$?
   - ¿Cuál es la probabilidad de que ocurra el evento A si se sabe que el evento $B = \{\text{bola blanca}\}$ ha ocurrido?
   - ¿Y cuál si se sabe que B no ha ocurrido?

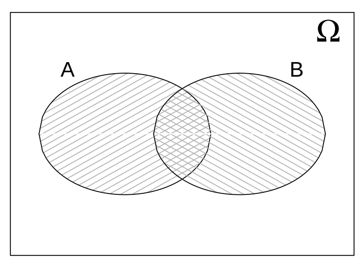
Definición : Sea $(\Omega, \mathcal{S}, \mathbb{P})$ un espacio de probabilidad, $B \in \mathcal{S}$ con $\mathbb{P}(B) > 0$. Se define la probabilidad del evento A dado el evento B, denotada por $\mathbb{P}_B(A)$, por

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \equiv \mathbb{P}(A|B)$$

para todo $A \in \mathcal{S}$. Si $\mathbb{P}(B) = 0$, $\mathbb{P}_B(\cdot)$ no está definido.

Nota: la notación $\mathbb{P}(A|B)$ es más común que $\mathbb{P}_B(A)$ para denotar la **probabilidad condicional de A dado el evento B**.

La *probabilidad condicional dado el evento B* se puede pensar como la medida de probabilidad, heredada de $\mathbb{P}$, definida sobre el espacio medible $(B, \mathcal{S}_B)$. Luego, $\mathbb{P}_B(A)$ es la probabilidad del evento más oscuro $(A \cap B)$ sobre la probabilidad del evento B.



Por mostrar que $\mathbb{P}_B$ es una medida de probabilidad sobre el espacio medible $(B, \mathcal{S}_B)$ donde $\mathcal{S}_B = \{A \cap B | A \in \mathcal{S}\}$.

Interpretación frecuentista.[1]

Suponga que el experimento aleatorio $\mathcal{E}$ se repite muchas veces, por decir n. Sean $N_n(A), N_n(B)$ y $N_n(A \cap B)$ las veces que ocurrió el evento A, B y $A \cap B$, respectivamente. Entonces, la interpretación frecuentista de la probabilidad quedaría como

$$\mathbb{P}(A) \approx \frac{N_n(A)}{n}; \quad \mathbb{P}(B) \approx \frac{N_n(B)}{n}; \quad \mathbb{P}(A \cap B) \approx \frac{N_n(A \cap B)}{n}$$

[1] Hoel, Port, and Stone (1971)

Si ahora uno se fija solamente en aquellos ensayos donde el evento B ocurrió, ésto se dio un número $N_n(B)$ veces, de los cuales $N_n(A\cap B)$ veces ocurrió el evento A. Así, la proporción de veces que ocurrió el evento A de aquellos que ocurrió B es

$$\frac{N_n(A\cap B)}{N_n(B)} = \frac{N_n(A\cap B)/n}{N_n(B)/n}$$

que para n grande debe aproximar $\dfrac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$

Ejemplo : En el caso de la urna, $\mathbb{P}(A) = \mathbb{P}(\{1\}) = \frac{2}{r+b}$.

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{1/(r+b)}{b/(r+b)} = \frac{1}{b} = \mathbb{P}_B(A)$$
$$\mathbb{P}(A|B^C) = \frac{\mathbb{P}(A\cap B^C)}{\mathbb{P}(B^C)} = \frac{1/(r+b)}{r/(r+b)} = \frac{1}{r} = \mathbb{P}_{B^C}(A)$$

Compare estas probabilidades con la **probabilidad incondicional**.

Ejemplo : Se lanzan dos monedas honestas.

a). Determine la probabilidad de que ambas hayan resultado *águila* si se sabe que la primera fue *águila*.

b). ¿Cuál es la probabilidad de que ambas sean *águila* si se sabe que al menos una lo fue?

Solución: Sean

$$\begin{aligned}\Omega &= \text{Espacio muestral} = \{\text{aa,as,sa,ss}\}\\ A &= \{\text{Ambas águilas}\} = \{\text{aa}\}\\ B &= \{\text{La primer moneda águila}\} = \{\text{aa,as}\}\\ C &= \{\text{Al menos un águila}\} = \{\text{aa,as,sa}\}\end{aligned}$$

Luego,

$$\text{a)}\quad \mathbb{P}(A|B) = \frac{\mathbb{P}(\{\text{aa}\})}{\mathbb{P}(\{\text{aa,as}\})} = \frac{1/4}{2/4} = \frac{1}{2}$$
$$\text{b)}\quad \mathbb{P}(A|C) = \frac{\mathbb{P}(\{\text{aa}\})}{\mathbb{P}(\{\text{aa,as,sa}\})} = \frac{1/4}{3/4} = \frac{1}{3}$$

Proposición : Sea $(\Omega, \mathcal{S}, \mathbb{P})$ un espacio de probabilidad, $B \in \mathcal{S}$ tal que $\mathbb{P}(B) > 0$. La medida de probabilidad condicional dado el evento B, $\mathbb{P}_B$ definida para toda $A \in \mathcal{S}$,

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$$

es una medida de probabilidad *legítima* en el sentido de que satisface los axiomas de probabilidad K_1, K_2 y K_3.

Demostración: En efecto,

K_1: Sea $A \in \mathcal{S}$, luego

$$\mathbb{P}_B(A) \stackrel{\text{def}}{=} \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} \geq 0$$

pues $\mathbb{P}$ es una medida de probabilidad (no negativa) y $\mathbb{P}(B) > 0$.

K_2: $\Omega \in \mathcal{S}$,

$$\mathbb{P}_B(\Omega) \stackrel{\text{def}}{=} \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$$

K_3: Sean $A_1, A_2, \dots \in \mathcal{S}$ mutuamente disjuntos, es decir, $A_i \cap A_j = \emptyset$, para todo $i \neq j$. Luego,

$$\begin{aligned}
\mathbb{P}_B(\cup_{i=1}^{\infty} A_i) &\stackrel{\text{def}}{=} \frac{\mathbb{P}(\cup_{i=1}^{\infty} A_i \cap B)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}(\cup_{i=1}^{\infty}(A_i \cap B))}{\mathbb{P}(B)}, \quad \text{los } (A_i \cap B) \text{ eventos ajenos} \\
&\stackrel{K_3}{=} \frac{\sum_{i=1}^{\infty} \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} \\
&= \sum_{i=1}^{\infty} \mathbb{P}_B(A_i)
\end{aligned}$$

Así pues, $\mathbb{P}_B$ cumple los axiomas de probabilidad y por lo tanto es una medida de probabilidad, la **probabilidad condicional dado el evento** B y $(B, \mathcal{S}_B, \mathbb{P}_B)$ un espacio de probabilidad.

Definición : Sean $A, B \in \mathcal{S}$, $\mathbb{P}(B) > 0$ y $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$. Entonces,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$$

La igualdad anterior se conoce como **regla de la multiplicación**. Así también si $\mathbb{P}(A) > 0$, la regla queda

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$$

Ejemplo : Suponga que la población de cierta localidad hay 40 % de hombres y 60 % de mujeres. Suponga que el 30 % de las mujeres fuma y 50 % de los hombres fuma. Si se selecciona un fumador al azar, ¿cuál es la probabilidad de que éste sea mujer? ¿Qué sea hombre?

Solución: Sean los eventos $M = \{\text{Mujer}\}$, $H = \{\text{Hombre}\}$, $F = \big\{\text{Fumador/a}\big\}$. Luego, $\mathbb{P}(H) = 0.40$, $\mathbb{P}(M) = 0.60$, $\mathbb{P}(F|M) = 0,30$, $\mathbb{P}(F|H) = 0.50$.

Note que $F = (F \cap H) \cup (F \cap M)$, eventos ajenos. Además, se sigue de la regla de la multiplicación $\mathbb{P}(F \cap H) = \mathbb{P}(F|H)\mathbb{P}(H) = .5(.4) = 0.20$, $\mathbb{P}(F \cap M) = \mathbb{P}(F|M)\mathbb{P}(M) = .3(.6) = 0.18$ y por lo tanto, $\mathbb{P}(F) = 0.20 + 0.18 = 0.38$. Entonces, la probabilidad de que un fumador elegido al azar sea mujer es

$$\mathbb{P}(M|F) = \frac{\mathbb{P}(M \cap F)}{\mathbb{P}(F)} = \frac{0.18}{0.38} = 0.4737$$

No siempre la ocurrencia o no ocurrencia de un evento afecta la probabilidad de ocurrencia de otro. En tal caso se dice que son eventos independientes.

Ejercicio :

a). Encuentre $\mathbb{P}(A|B)$ si:

$$i)\ A \cap B = \emptyset; \qquad ii)\ A \subseteq B; \qquad iii)\ B \subseteq A$$

b). Muestre que si $\mathbb{P}(A|B) > \mathbb{P}(A)$ entonces $\mathbb{P}(B|A) > \mathbb{P}(B)$. Interprete el resultado.

3.2. Independencia

Definición : Sea $(\Omega, \mathcal{S}, \mathbb{P})$ un espacio de probabilidad. $A, B \in \mathcal{S}$ se dicen **eventos independientes** si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Proposición : Sean $A, B \in \mathcal{S}$, eventos independientes. Entonces, si $\mathbb{P}(B) > 0$, $\mathbb{P}(A|B) = \mathbb{P}(A)$. Y por lo mismo, $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Demostración: Sean A, B dos eventos independientes y $\mathbb{P}(B) > 0$,

$$\mathbb{P}(A|B) \stackrel{\text{def}}{=} \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \stackrel{\text{ind}}{=} \frac{\mathbb{P}(A)\cancel{\mathbb{P}(B)}}{\cancel{\mathbb{P}(B)}} = \mathbb{P}(A)$$

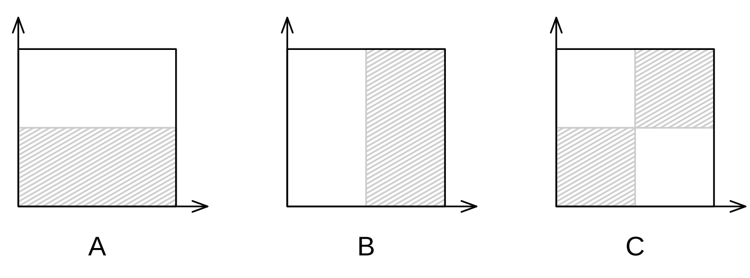
Ejemplo :

- Los eventos {Fumar} y {Cáncer pulmonar} no son independientes pues definitivamente fumar aumenta la probabilidad del cáncer pulmonar.
- En un mazo de cartas, la probabilidad de un *as* no se ve afectado por el evento *tréboles*. Luego, los eventos son independientes.
- Note que si $A, B \in \mathcal{S}$, eventos ajenos. Si $\mathbb{P}(A), \mathbb{P}(B) > 0$, entonces los eventos no pueden ser independientes.

Definición : Sean $A, B, C \in \mathcal{S}$. Se dicen **eventos mutuamente independientes** si son independientes a pares y $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.

Proposición : Eventos independientes a pares no son necesariamente mutuamente independientes.

Demostración: Suponga un espacio de probabilidad uniforme y considere los siguientes eventos



Luego, $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = 1/2$ y

$$\begin{aligned}
\mathbb{P}(A \cap B) &= 1/4 = 1/2 \cdot 1/2 = \mathbb{P}(A)\mathbb{P}(B) \\
\mathbb{P}(A \cap C) &= 1/4 = 1/2 \cdot 1/2 = \mathbb{P}(A)\mathbb{P}(C) \\
\mathbb{P}(B \cap C) &= 1/4 = 1/2 \cdot 1/2 = \mathbb{P}(B)\mathbb{P}(C)
\end{aligned}$$

Entonces, los eventos son independientes a pares pero

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\emptyset) = 0 \neq \frac{1}{8} = \frac{1}{2}\,\frac{1}{2}\,\frac{1}{2} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$$

por lo que no son mutuamente independientes.

Definición : Sean $A_1, A_2, \dots \in \mathcal{S}$. Se dicen **eventos mutuamente independientes** si para todo $k = 2, 3, \dots$ e índices $i_1, \dots, i_k \in \mathbb{N}$ distintos, se tiene que

$$\mathbb{P}\left(A_{i_1} \cap \cdots \cap A_{i_k}\right) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})$$

Nota: En principio, habría que verificar que los $\sum_{j=2}^{n} \binom{n}{j} = 2^n - n - 1$ distintas subcolecciones satisfacen la igualdad anterior. Sin embargo, en muchas ocasiones la independencia mutua se sigue de la independencia de los experimentos.

Definición : Sea $(\Omega, \mathcal{S}, \mathbb{P})$ un espacio de probabilidad. Considere $A, B, C \in \mathcal{S}$. Los eventos A y B se dicen **condicionalmente independientes dado el evento** C si $\mathbb{P}_C(A \cap B) = \mathbb{P}_C(A)\mathbb{P}_C(B)$, o bien,

$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B|C)$$

Proposición : Sean $A, B, C \in \mathcal{S}$, A y B eventos condicionalmente independientes dado C. Entonces, $\mathbb{P}(A|B \cap C) = \mathbb{P}(A|C)$.

Demostración:

$$\begin{aligned}
\mathbb{P}(A|B \cap C) &= \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(B \cap C)}, && \text{por definición} \\
&= \frac{\mathbb{P}(A \cap B|C)\mathbb{P}(C)}{\mathbb{P}(B|C)\mathbb{P}(C)}, && \text{regla de la multiplicacion} \\
&= \frac{\mathbb{P}(A|C)\mathbb{P}(B|C)}{\mathbb{P}(B|C)}, && \text{independencia condicional} \\
&= \mathbb{P}(A|C)
\end{aligned}$$

Proposición : Sea $(\Omega, \mathcal{S}, \mathbb{P})$ un espacio de probabilidad, $A, B \in \mathcal{S}$ eventos independientes, entonces también lo son A y B^C; A^C y B; A^C y B^C.

Demostración: $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^C) \stackrel{\text{ind}}{=} \mathbb{P}(A)\mathbb{P}(B) + \mathbb{P}(A \cap B^C)$. Luego,

$$\mathbb{P}(A \cap B^C) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\left[1 - \mathbb{P}(B)\right] = \mathbb{P}(A)\mathbb{P}(B^C)$$

Por lo que A y B^C son eventos independientes. Y así los otros casos.

Proposición : Sea $(\Omega, \mathcal{S}, \mathbb{P})$ un espacio de probabilidad, $A, B, C \in \mathcal{S}$ eventos mutuamente independientes. Entonces, A es independiente de $B \cup C$ y $B \cap C$.

Demostración:

i)

$$\begin{aligned}
\mathbb{P}\left(A \cap (B \cup C)\right) &= \mathbb{P}\left((A \cap B) \cup (A \cap C)\right) \\
&= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap C) - \mathbb{P}(A \cap B \cap C) \\
&= \mathbb{P}(A)\mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(C) - \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) \\
&= \mathbb{P}(A)\left[\mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(B)\mathbb{P}(C)\right] \\
&= \mathbb{P}(A)\left[\mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(B \cap C)\right] \\
&= \mathbb{P}(A)\mathbb{P}(B \cup C)
\end{aligned}$$

Por lo que A y $B \cup C$ son eventos independientes.

ii)

$$\mathbb{P}\left(A \cap (B \cap C)\right) = \mathbb{P}(A)\left[\mathbb{P}(B)\mathbb{P}(C)\right] = \mathbb{P}(A)\mathbb{P}(B \cap C)$$

Se sigue que A y $(B \cap C)$ son eventos independientes.

Ejemplo : Se tiene una sucesión de *ensayos independientes*. Cada ensayo tiene dos posibles salidas: *éxito* y *fracaso*. La probabilidad de éxito de cualquiera de los ensayos es p y la de fracaso $q = 1 - p$. Determine las siguientes probabilidades.

a). Al menos un éxito en los primeros n ensayos.

b). Exactamente k éxitos en los primeros n ensayos

c). Todos los ensayos exitosos.

Ensayos como los descritos anteriormente se conocen como **ensayos Bernoulli**.

Solución: Para $i = 1, 2, \ldots$, sean $E_i = \left\{\text{Éxito en el } i\text{-ésimo ensayo}\right\}$, con $\mathbb{P}(E_i) = p$; $F_i = E_i^C = \{\text{Fracaso en el } i\text{-ésimo ensayo}\}$, con $\mathbb{P}(F_i) = q = 1 - p$.

a). $\{\text{Ningún exito en } n \text{ ensayos}\} = F_1 \cap \cdots \cap F_n = \mathbb{E}_1^C \cap \cdots \cap E_n^C$. Luego, $\mathbb{P}(\cap_{i=1}^n F_i) \overset{\text{ind}}{=} \prod_{i=1}^n \mathbb{P}(F_i) = q^n$. Por lo que $\mathbb{P}(\cup_{i=1}^n E_i) = 1 - q^n$. Así,

$$\mathbb{P}\left(\{\text{Al menos un éxito en } n \text{ ensayos}\}\right) = 1 - (1-p)^n$$

b). Suponga k éxitos en los primeros ensayos y $n - k$ fracasos después. A saber, $E_1 \cap \cdots \cap E_k \cap F_1 \cap \cdots \cap F_{n-k} =: A$. Luego,

$$\mathbb{P}(A) \overset{\text{ind}}{=} \underbrace{p \cdots p}_{k}\, \underbrace{q \cdots q}_{n-k} = p^k q^{n-k}$$

Pero si no se restringe el orden donde ocurren los éxitos, hay $\binom{n}{k}$ posiciones donde acomodar los k éxitos en los n ensayos y cada una de estos arreglos se da con probabilidad de $p^k q^{n-k}$. Así,

$$\mathbb{P}\left(\{\text{Exactamente } k \text{ éxitos en } n \text{ ensayos}\}\right) = \binom{n}{k} p^k (1-p)^{n-k}$$

c). Considere el evento $\{\text{Todos éxitos}\} = \cap_{i=1}^\infty E_i$ y para $p < 1$,

$$\mathbb{P}\left(\cap_{i=1}^\infty E_i\right) = \mathbb{P}\left(\lim_{n\to\infty} \cap_{i=1}^n E_i\right) \overset{\text{cont}}{=} \lim_{n\to\infty} \mathbb{P}(\cap_{i=1}^n E_i) = \lim_{n\to\infty} p^n = 0$$

Por otro lado, si $p = 1$ entonces $\mathbb{P}\left(\{\text{Todos éxitos}\}\right) = 1$.

3.3. Teorema de probabilidad total (TPT)

Definición : Sea $(\Omega, \mathcal{S}, \mathbb{P})$ un espacio de probabilidad y $B_1, B_2, \cdots \in \mathcal{S}$ tales que: *i)* $B_i \cap B_j = \emptyset$, $i \neq j$; *ii)* $\Omega = \cup_{n=1}^\infty B_n$. Entonces, la colección $\{B_n\}$ se dice que es una **partición (infinita)** del espacio muestral Ω.

Proposición : Sea $(\Omega, \mathcal{S}, \mathbb{P})$ un espacio de probabilidad y $\{B_i\}$ una partición de Ω. Entonces, para todo $a \in \mathcal{S}$, $\mathbb{P}(A) = \sum_{i=1}^\infty \mathbb{P}(A \cap B_i)$.

Demostración: Note que para todo $A \in \mathcal{S}$, $A = A \cap \Omega = A \cap (\cup_{i=1}^{\infty} B_i) = \cup_{i=1}^{\infty}(A \cap B_i)$, con los $A \cap B_i$ ajenos, por ser $\{B_n\}$ una partición de Ω. Se sigue de K_3 que $\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap B_i)$.

Nota: En la definición y proposición anteriores la partición no tiene que se necesariamente infinita. Bien puede ser que $B_1, \ldots, B_n \in \mathcal{S}$, tales que $B_i \cap B_j = \emptyset$ y $\Omega = \cup_{i=1}^{n} B_i$. De hecho, defina $B_k = \emptyset$, $k = n+1, n+2, \ldots$ y tendría las condiciones anteriores.

Teorema de probabilidad total (TPT) Sea $(\Omega, \mathcal{S}, \mathbb{P})$ un espacio de probabilidad y $\{B_n\}$ una partición de Ω. Entonces, para todo evento $A \in \mathcal{S}$,

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

Demostración: Se sigue de la proposición anterior y regla de la multiplicación

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

Ejemplo : Suponga que se tienen 3 urnas U_i, $i = 1, 2, 3$, donde la i-ésima urna tiene b_i bolas blancas y n_i bolas negras. Se selecciona una urna y de ella se extrae un pelota. Si se sabe que la probabilidad de seleccionar la urna U_i es p_i, determine la probabilidad de que la pelota seleccionada sea negra.

Solución: Sea N el evento que denota que la bola seleccionada sea negra. Entonces, se sigue del teorema de probabilidad total (TPT) que

$$\begin{aligned}\mathbb{P}(N) &= \mathbb{P}(N|U_1)\mathbb{P}(U_1) + \mathbb{P}(N|U_2)\mathbb{P}(U_2) + \mathbb{P}(N|U_3)\mathbb{P}(U_3) \\ &= \frac{n_1}{b_1 + n1} p_1 + \frac{n_2}{b_2 + n_2} p_2 + \frac{n_3}{b_3 + n_3} p_3\end{aligned}$$

La manera de proceder en el siguiente ejemplo Blitzstein and Hwang lo llaman *condicionar en el primer paso.*

Ejemplo : (Tomado de Blitzstein and Hwang (2014)) Se tiene una amiba que después de un minuto se divide en dos, se queda como está o muere con la misma probabilidad. Después del minuto las amibas sobrevivientes se comportan igual y de manera independiente. Determine la probabilidad de que la población de amibas muera eventualmente.

Solución: Sean los eventos $M = \{\text{Población de amibas muere}\}$, $C = \{\text{Amiba se conserva}\}$ y $V = \{\text{Amiba se divide en dos}\}$. Entonces, por el teorema de probabilidad total, se tiene

$$\begin{aligned}\mathbb{P}(M) &= \mathbb{P}(M|M)\mathbb{P}(M) + \mathbb{P}(M|C)\mathbb{P}(C) + \mathbb{P}(M|V)\mathbb{P}(V) \\ &= 1\,\frac{1}{3} + \mathbb{P}(M)\frac{1}{3} + \mathbb{P}^2(M)\frac{1}{3}\end{aligned}$$

que da lugar a la ecuación cuadrática $p^2 - 2p + 1 = 0$, con $p = \mathbb{P}(M)$ y solución $p = 1$. Esto es, con certeza la población de amibas muere eventualmente.

Ejercicio : Recuerde el problema 2.119 de Wackerly et al. (2008) visto en la primera sección. Suponga que se lanzan repetidamente dos dados y se cuenta la suma de las caras hacia arriba en cada lanzamiento. Determine la probabilidad de que la suma 3 sale antes que la suma 7 utilizando el argumento de condicionar en el primer paso.

Regla de Bayes

Proposición : Bajo los supuestos del TPT, para todo $A \in \mathcal{S}$ y cualquier B_k elemento de la partición, se tiene que

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}$$

La igualdad anterior se conoce como la **regla de Bayes**.

Demostración: Se sigue de la definición de probabilidad condicional, aplicar la regla de la multiplicación de en el numerador y el TPT en el denominador. A saber,

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A \cap B_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}$$

Ejemplo : En una prueba con preguntas de *opción múltiple* el/la que responde sabe o no la respuesta. Sea p la probabilidad de que sepa la respuesta y $q = 1 - p$ que tenga que adivinarla. Suponga que un/a estudiante que adivina la respuesta la tendrá correcta con probabilidad de $1/m$, donde m es el número de opciones para la respuesta. Determine la probabilidad de que sepa la respuesta dado que la tuvo correcta.

Solución: Sean los eventos $C = \{\text{Respuesta correcta}\}$, $S = \{\text{Sabe la respuesta}\}$ Entonces, S y S^C constituyen una partición y de la regla de Bayes se sigue que

$$\begin{aligned}
\mathbb{P}(S|C) &= \frac{\mathbb{P}(C|S)\mathbb{P}(S)}{\mathbb{P}(C|S)\mathbb{P}(S) + \mathbb{P}(C|S^C)\mathbb{P}(S^C)} \\
&= \frac{1\,p}{1\,p + \frac{1}{m}(1-p)} \\
&= \frac{mp}{mp + 1 - p}
\end{aligned}$$

Ejemplo : Se tienen las siguientes estadísticas. En un grupo de estudio 60 % de las mujeres están embarazadas. Se practican pruebas de embarazo con un historial de 10 % de *falsos positivos* y 20 % de *falsos negativos*. Se selecciona una mujer al azar y se le practica una prueba y resulta positiva. Si se repite la prueba de manera independiente y resulta nuevamente positiva, ¿cuál es la probabilidad de que la mujer esté realmente embarazada?

Solución: Sean los eventos $E = \{\text{Mujer embarazada}\}$, $S_i = \{\text{Prueba } i \text{ resulta positivo}\}$, con $i = 1, 2$. Luego, $\mathbb{P}(S|E^C) = 0.10$, $\mathbb{P}(S^C|E) = 0.20$, $\mathbb{P}(E) = 0.60$. Entonces, aplicando la regla de Bayes y la independencia condicional de S_i dado E,

$$\begin{aligned}
\mathbb{P}(E|S_1 \cap S_2) &= \frac{\mathbb{P}(E \cap S_1 \cap S_2)}{\mathbb{P}(S_1 \cap S_2)} \\
&= \frac{\mathbb{P}(S_1 \cap S_2|E)\mathbb{P}(E)}{\mathbb{P}(S_1 \cap S_2|E)\mathbb{P}(E) + \mathbb{P}(S_1 \cap S_2|E^C)\mathbb{P}(E^C)} \\
&= \frac{\mathbb{P}(S_1|E)\mathbb{P}(S_2|E)\mathbb{P}(E)}{\mathbb{P}(S_1 \cap S_2|E)\mathbb{P}(E) + \mathbb{P}(S_1 \cap S_2|E^C)\mathbb{P}(E^C)} \\
&= \frac{(1-.2)(1-.2)(.6)}{(1-.2)^2(.6) + (.1)^2(.4)} \\
&= 0.990
\end{aligned}$$

Ejemplo : (Tomado de Wackerly et al. (2008)) Fusibles son construidos en 5 líneas de producción. Los fusibles son confiables y caros y se empacan en lotes de 100 fusibles. Como las pruebas son destructivas se inspeccionan solamente pocos de ellos para decidir si se acepta o no el lote.

Las 5 líneas producen bajo condiciones normales la misma tasa de defectuosos del 2 %. La línea 2 sufrió una avería y produjo por un mes con una tasa de defectuosos del 5 %.

Un consumidor recibe un lote de fusibles ese mes, prueba 3 artículos y uno de ellos es defectuoso. Determine la probabilidad de que el lote venga de la línea 2. ¿Cuál es la probabilidad de que venga de cualesquiera de las otras 4 líneas?

Solución: Sean los eventos $D = \{\text{Fusible defectuos}\}$, $L_i = \{\text{Línea de producción } i\}$, $i = 1, \ldots, 5$ y $A_1 = \{\text{Uno de los 3 fusibles defectuoso}\}$ Luego, $\mathbb{P}(L_i) = 0.2, i = 1, \ldots, 5; \mathbb{P}(D|L_i) = 0.02, i \neq 2, \mathbb{P}(D|L_2) = 0.05; \mathbb{P}(A_1|L_2) = \binom{3}{1}(.05)(.95)^2 = 0.1354; \mathbb{P}(A_1|L_2^C) = \binom{3}{1}(.02)(.98)^2 = 0.0578$. Y se sigue del teorema de probabilidad total (TPT)

$$\begin{aligned}
\mathbb{P}(A_1) &= \mathbb{P}(A_1|L_2)\mathbb{P}(L_2) + \mathbb{P}(A_1|L_2^C)\mathbb{P}(L_2^C) \\
&= .1354(.2) + .0576(.8) \\
&= 0.0732
\end{aligned}$$

Y de la regla de Bayes

$$\mathbb{P}(L_2|A_1) = \frac{\mathbb{P}(A_1|L_2)\mathbb{P}(L_2)}{\mathbb{P}(A_1)} = \frac{.1354(.2)}{.0732} = 0.370$$

Finalmente,

$$\mathbb{P}(L_2^C|A_1) = 1 - \mathbb{P}(L_1|A_1) = 1 - 0.37 = 0.630$$

Ejemplo : (Tomado de Ross (2010)) En una investigación criminal el inspector a cargo está 60 % convencido de que cierto sospechoso sea culpable. Un“nuevo” elemento de evidencia criminal tiene cierta característica (zurdo, calvo, flaco) está disponible. Si el 20 % de la población tiene esa característica, ¿qué tan convencido deberá estar el inspector sobre la culpabilidad del sospechoso?

Solución: Sean los eventos C={El sospechoso sea culpable} y R={El sospechoso tiene la característica del criminal}. Entonces,

$$\mathbb{P}(C|R) = \frac{\mathbb{P}(R|C)\mathbb{P}(C)}{\mathbb{P}(R|C)\mathbb{P}(C) + \mathbb{P}(R|C^C)\mathbb{P}(C^C)} = \frac{1(.6)}{1(.6) + .2(.4)} = 0.882$$

donde se supuso que la probabilidad de que el sospechoso tenga la característica sea 0.2, aún si es inocente.

Nota: El cambio en la probabilidad de la hipótesis (es culpable) a partir de la “mera” información se expresa de manera resumida por medio de los momios.

Definición : Se definen los **momios** (*odds*)) de un evento A por $\mathbb{P}(A)/\mathbb{P}(A^C)$. Los momios de cierto evento dice que tan probable es la ocurrencia del evento con respecto a la probabilidad de que no ocurra. Así por ejemplo, si $\mathbb{P}(A) = 2/3$, $\mathbb{P}(A^C) = 1/3$ y $\mathbb{P}(A) = 2\mathbb{P}(A^C)$. Luego, los momios son 2 (*“dos a uno”*). Si los momios fuesen α, se diría que *“ los momios son alfa a uno”* en favor de la hipótesis.

Considere que la hipótesis H es cierta con probabilidad $\mathbb{P}(H)$ y suponga se tiene nueva información I disponible. Entonces, la probabilidad de H (y de no H, H^C) es

$$\mathbb{P}(H|I) = \frac{\mathbb{P}(I|H)\mathbb{P}(H)}{\mathbb{P}(I)}; \qquad \mathbb{P}(H^C|I) = \frac{\mathbb{P}(I|H^C)\mathbb{P}(H^C)}{\mathbb{P}(I)};$$

Por lo que los nuevos momios a partir de la nueva información es

$$\frac{\mathbb{P}(H|I)}{\mathbb{P}(H^C|I)} = \frac{\mathbb{P}(I|H)\mathbb{P}(H)/\mathbb{P}(I)}{\mathbb{P}(I|H^C)\mathbb{P}(H^C)/\mathbb{P}(I)} = \frac{\mathbb{P}(H)}{\mathbb{P}(H^C)} \cdot \frac{\mathbb{P}(I|H)}{\mathbb{P}(I|H^C)}$$

que son los momios anteriores veces el cociente de probabilidades de la nueva información dada la certeza o no de la hipótesis.

Si se refiere al ejemplo anterior, los momios aumentan (disminuyen) si la probabilidad de la información es mayor (menor) si la hipótesis es verdadera (falsa).

Ejemplo : (Tomado de Ross (2010)) Se tienen dos urnas. La urna 1 tiene n bolas (moléculas) rojas y la urna 2 tiene n moléculas azules. A cada ensayo se extrae una bola de la urna 1 que es reemplazada por una bola de la urna 2. De esta manera la urna 1 queda vacía después de $2n$ extracciones.

a). Sea $R = \{\text{La última molécula extraída fue roja}\}$. Encuentre $\mathbb{P}(R)$.

b). Repita el experimento suponiendo r_i bolas rojas y a_i bolas azules en la urna U_i, $i = 1, 2$. Encuentre $\mathbb{P}(R)$.

Solución:

a). Sean R_k=$\{$La última molécula extraída es la k-ésima bola roja$\}$ y sea N_i=$\{$La k-ésima bola roja *no* fue extraída en el i-ésimo ensayo$\}$. Entonces, usando la notación $AB = A \cap B$, se sigue de la regla de la multiplicación

$$\begin{aligned}
\mathbb{P}(R_k) &= \mathbb{P}(R_k N_1 \cdots N_n) \\
&= \mathbb{P}(R_k|N_1 \cdots N_n)\mathbb{P}(N_1 \cdots N_n) \\
&= \mathbb{P}(R_k|N_1 \cdots N_n)\mathbb{P}(N_n|N_1 \cdots N_{n-1})\mathbb{P}(N_1 \cdots N_{n-1}) \\
&\vdots \\
&= \mathbb{P}(R_k|N_1 \cdots N_n)\mathbb{P}(N_n|N_1 \cdots N_{n-1}) \cdots \mathbb{P}(N_2|N_1)\mathbb{P}(N_1)
\end{aligned}$$

Ahora bien, note que $\mathbb{P}(N_1) = \frac{n-1}{n}$ y $\mathbb{P}(N_j|N_1 \cdots N_{j-1}) = \frac{n-1}{n}$, para $j = 2, \ldots, n$, y

$$\mathbb{P}(R_k|N_1 \cdots N_n) = \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdots \frac{1}{2} \cdot 1 = \frac{1}{n}$$

Y por lo tanto

$$\mathbb{P}(R_k) = \frac{1}{n} \underbrace{\frac{n-1}{n} \cdots \frac{n-1}{n}}_{n \text{ veces}} = \frac{1}{n}\left(1 - \frac{1}{n}\right)^n$$

Finalmente, los R_k son eventos ajenos, por lo que

$$\mathbb{P}(R) = \mathbb{P}\left(\cup_{k=1}^n R_k\right) = \sum_{k=1}^n \mathbb{P}(R_k) = n \cdot \frac{1}{n}\left(1 - \frac{1}{n}\right)^n \underset{n\to\infty}{\longrightarrow} e^{-1}$$

b). Suponga ahora que hay r_i bola srojas y a_i bolas azules en la urna U_i, $i = 1, 2$. Nuevamente sean R y R_k los eventos definidos en el inciso anterior. Entonces,

$$\mathbb{P}(R_k) = \frac{1}{r_1 + a_1}\left(1 - \frac{1}{r_1 + a_1}\right)^{r_2 + a_2} = p$$

donde el término entre paréntesis es la probabilidad de que la bola esté en la urna U_1 habiendo vaciado U_2.

Sea ahora W={La última molécula extraída es original de la urna U_1}. Entonces,

$$\mathbb{P}(W) = \mathbb{P}\left(\cup_{i=1}^{r_1} R_i \ \cup \ \cup_{j=1}^{a_1} A_j\right) = (r_1 + a_1)p = \left(1 - \frac{1}{r_1 + a_1}\right)^{r_2+a_2}$$

Finalmente, se sigue del teorema de probabilidad total (TPT)

$$\begin{aligned}
\mathbb{P}(R) &= \mathbb{P}(R|W)\mathbb{P}(W) + \mathbb{P}(R|W^C)\mathbb{P}(W^C) \\
&= \frac{r_1}{r_1 + a_1}\left(1 - \frac{1}{r_1 + a_1}\right)^{r_2+a_2} + \frac{r_2}{r_2 + a_2}\left[1 - \left(1 - \frac{1}{r_1 + a_1}\right)^{r_2+a_2}\right]
\end{aligned}$$

Si $a_1 + a_1 = n = R_2 + a_2$, grande, entonces

$$\mathbb{P}(R) \approx \frac{r_1}{n}e^{-1} + \frac{r_2}{n}(1 - e^{-1})$$

3.4. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 3, Barrios and Heiras (2024).

4. Variables Aleatorias

4.1. Variables aleatorias

Definición : Sea $(\Omega, \mathcal{S}, \mathbb{P})$ un espacio de probabilidad. La función real X con dominio en Ω se dice **variable aleatoria** (*v. a.*) si para todo $x \in \mathbb{R}$, la preimagen de x es un evento. Esto es, si

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto x \end{aligned} \qquad \text{y para todo } x \in \mathbb{R}, \big\{\omega \in \Omega : X(\omega) = x\big\} \in \mathcal{S} \tag{1}$$

La figura 2 muestra el experimento aleatorio $\mathcal{E}$ que arroja salidas ω. La variable aleatoria X mapea éstas en los reales, $X(\omega) = x \in \mathbb{R}$. En Teoría de la Medida la variable aleatoria

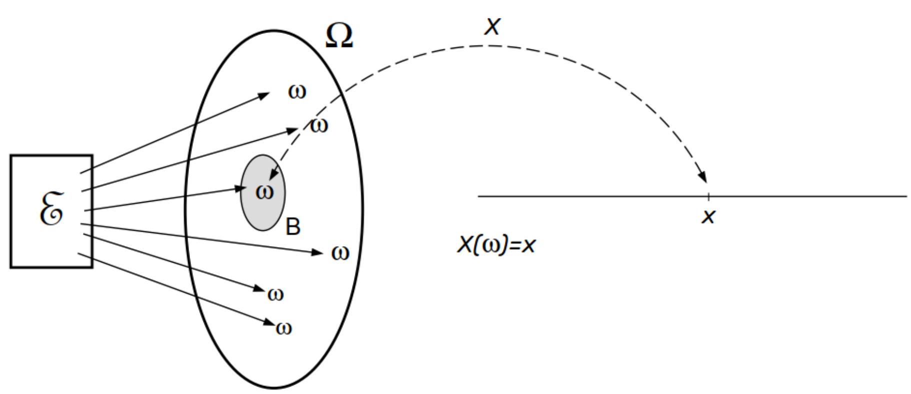


Figura 2: La variable aleatoria X mapea la salida ω en los reales. El evento B es la preimagen de x.

X que satisface la condición anterior (1) se diría que es una **función medible**.

Ejemplo : Considere el experimento aleatorio (EA) $\mathcal{E}$ que consiste en el lanzamiento de 3 monedas honestas. El espacio muestral (EM) sería $\Omega = \{aaa, aas, \ldots, sss\}$. Sea X la variable aleatoria que denota el número total de águilas en el lanzamiento de las monedas. La siguiente tabla describe la (función) variable aleatoria $X : \Omega \to \mathbb{R}$

ω	aaa	aas	asa	saa	ssa	sas	ass	sss
x	3	2	2	2	1	1	1	0

El rango de la función X es $R_X = \{0, 1, 2, 3\}$. Note,

$$\mathbb{P}(\{X = 1\}) \equiv \mathbb{P}(\big\{\omega \in \Omega : X(\omega) = 1\big\} = \mathbb{P}(\{ass, sas, ssa\})) = \frac{3}{8}$$

Los eventos $\{X = 1\}$ y $\{ass, sas, ssa\}$ se dicen eventos equivalentes. En realidad $\{X = 1\}$ es una manera abreviada de referirse al evento $\{\omega \in \Omega : X(\omega) = 1\} \in \mathcal{S}$, más práctica.

Definición : Sea $(\Omega, \mathcal{S}, \mathbb{P})$ un espacio de probabilidad (EP) y X una variable aleatoria (*v. a.*). Para todo $A \in \mathcal{B}(\mathbb{R})$, los eventos $\{X \in A\}$ y $\{\omega \in \Omega : X(\omega) \in A\}$ se dicen **eventos equivalentes**.

4.2. Variables aleatorias discretas

Definición : Sea $(\Omega, \mathcal{S}, \mathbb{P})$ un *EP*, y X una *v. a.*. Si R_X, los valores que puede tomar la *v. a.* X es un conjunto discreto (finito, o infinito numerable), $R_X = \{x_1, x_2, \ldots\}$, X se dice que es una **variable aleatoria discreta**.

Nota: Sea X una *v. a.* discreta con rango $R_X = \{x_1, x_2, \ldots\}$. Entonces,

$$\mathbb{P}(X = x_i) = \mathbb{P}(\{X = x_i\}) = \mathbb{P}\Big(\big\{\omega \in \Omega : X(w) = x_i\big\}\Big) = p_i \in [0, 1]$$

pues siendo $\mathbb{P}$ una medida de probabilidad se mostró ya que para todo evento A se tienen que $0 \leq \mathbb{P}(A) \leq 1$.

Definición : Sea X una *v. a.* discreta con rango $R_X = \{x_1, x_2, \ldots\}$. Se define f_X, la **función masa de probabilidad (*f. m. p.*) de** X tal que, para todo $x \in \mathbb{R}$,

$$\begin{aligned} f_X : \mathbb{R} &\longrightarrow [0,\ 1] \\ x &\longmapsto \mathbb{P}(X = x) \end{aligned}$$

Esto es,

$$f_X(x) = \mathbb{P}(X = x) = \mathbb{P}\Big(\big\{\omega \in \Omega : X(\omega) = x\big\}\Big), \quad \text{para todo } x \in \mathbb{R}$$

Definición : Sea X *v. a.* disreta con función masa de probabilidad f_X. El conjunto $S_X = \{x \in \mathbb{R} : f_X(x) > 0\}$ se dice el **soporte de (la distribución) de** X. En el caso de la *v. a.* discretas, $S_X = R_X$, el rango de X.

Proposición : Sea X una *v. a.* discreta y f_X su *f. m. p.*. Entonces,

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

ii) $\sum_{x_i \in S_X} f_X(x_i) = 1$.

Demostración:

i) Sea $x \in \mathbb{R}$, luego, $f_X(x) = \mathbb{P}\Big(\big\{\omega \in \Omega : X(\omega) = x\big\}\Big) \geq 0$, pues es una probabilidad.

ii) Sea $S_X = R_X = \{x_2, x_2, \ldots\}$. Sean $B_i = \big\{\omega \in \Omega : X(\omega) = x_i\big\} \in \mathcal{S}$. Los $\{B_i\}$ forman una partición de Ω.

Si $x \notin S_X$, $f_X(x) = \mathbb{P}\Big(\big\{\omega \in \Omega : X(\omega) = x\big\}\Big) = \mathbb{P}(\emptyset) = 0$.

Si $x \in S_X$, $x = x_i$, para algún $x_i \in S_X$ y $f_X(x_i) = \mathbb{P}\Big(\big\{\omega \in \Omega : X(\omega) = x_i\big\}\Big)$.

Se sigue de los axiomas de probabilidad,

$$\sum_{x_i \in S_X} f_X(x_i) = \sum_{x_i \in S_X} \mathbb{P}(X = x_i) = \mathbb{P}\left(\cup_{x_i \in S_X} \{X = x_i\}\right) = \mathbb{P}\left(\cup_i B_i\right) = \mathbb{P}(\Omega) = 1$$

Ejemplo : Considere nuevamente $\mathcal{E}$, el *EA* que consiste en el lanzamiento de 3 monedas honestas. El *EM*, $\Omega = \{aaa, \ldots, sss\}$, la familia de eventos $\mathcal{S} = \mathcal{P}(\Omega)$ y $\mathbb{P}$ la probabilidad uniforme. Para todo $\omega \in \Omega$, $\mathbb{P}(\{\omega\}) = 1/8$. Si X es la *v. a.* que representa el número de águilas, $X : \Omega \to \{0, 1, 2, 3\} = R_X$ y su *f. m. p.* está dada en la tabla

x	0	1	2	3	
p	1/8	3/8	3/8	1/8	1

Nota:

i) $f_X(x) \geq 0$. Pues por ejemplo, $f_X(\pi) = \mathbb{P}(X = \pi) = \mathbb{P}(\emptyset) = 0$, pero si $x_i \in \{0, 1, 2, 3\} = S_X$, $f_X(x_i) > 0$.

ii) $\sum_{x_i \in S_X} f_X(x_i) = f_X(0) + f_X(1) + f_X(2) + f_X(3) = 1/8 + 3/8 + 3/8 + 1/8 = 1.$

Definición : Sea $A \subseteq \mathcal{D}$. Se define la **función indicadora del conjunto** A por

$$\begin{aligned} \mathbb{1}_A : \mathcal{D} &\longrightarrow \mathbb{R} \\ x &\longmapsto \{0,1\} \end{aligned} \qquad \text{tal que} \qquad \mathbb{1}_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

Luego, la *f. m. p.* del ejemplo anterior se puede escribir como

$$f_X(x) = \binom{3}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{3-x} \mathbb{1}_{\{0,1,2,3\}}(x)$$

Nota: Sea X una *v. a.* discreta, $A \in \mathcal{B}(\mathbb{R})$, entonces

$$\begin{aligned} \mathbb{P}(X \in A) &= \mathbb{P}\Big(\big\{w \in \Omega : X(\omega) \in A\big\}\Big) \\ &= \mathbb{P}\left(\cup_{x_i \in A} \{X = x_i\}\right) \\ &= \sum_{x_i \in A} \mathbb{P}(X = x_i) \\ &= \sum_{x_i \in A} f_X(x_i) \end{aligned}$$

pues los $\{X = x_i\}$ son eventos ajenos para distintos x_i's y aplicando K_3.

Ensayos Bernoulli

Considere el experimento aleatorio $\mathcal{E}$ con solamente dos salidas posibles: éxito/fracaso, sí/no, correcto/defectuoso, presente/ausente, 1/0, etcétera.

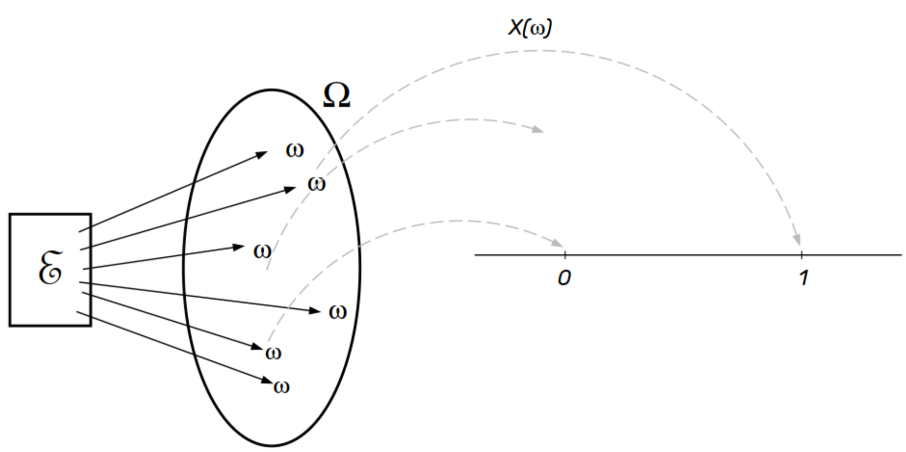


Figura 3: Ensayos Bernoulli.

Esta situación queda representado en la figura 3. Se pueden representar por variables aleatorias X binarias que toman valores 0 ó 1.

$$X(\omega) = \begin{cases} 1 & \text{éxito, sí, correcto, presente} \\ 0 & \text{fracaso, no, defectuoso, ausente} \end{cases}$$

Con probabilidad de éxito p $(0 \leq p \leq 1)$

$$\mathbb{P}(X=1)=p, \quad \mathbb{P}(X=0)=1-p=q$$

Luego, la correspondiente *f. m. p.* f_X es

$$f_X(x)=p\mathbb{1}_{\{1\}}(x)+(1-p)\mathbb{1}_{\{0\}}(x)$$

Experimentos aleatorios *independientes* como el recién descrito se dicen **ensayos Bernoulli** con probabilidad de éxito p.

Ejemplo : Recuerde el ejemplo del lanzamiento de 3 monedas. Defina Y_i la variable aleatoria indicadora con $Y_i=\begin{cases} 1 & \text{águila} \\ 0 & \text{sol} \end{cases}$, para $i=1,2,3$. Las variables aleatorias Y_i's se dicen **v. a.'s Bernoulli**.

Note que si X es la variable aleatoria que denota el número de águilas, entonces, $X=Y_1+Y_2+Y_3$.

Ejemplo : Considere ahora una sucesión de ensayos Bernoulli con probabilidad de éxito p. Sea Y la *v. a.* que denota el número de ensayos hasta el primer éxito. Entonces, el rango de Y es el soporte $S_Y=\{1,2,3,\dots\}$. Por encontrar f_Y, la *f. m. p.* de la *v. a.* Y.

$$\begin{aligned}
f_Y(1) &= \mathbb{P}(Y=1)=p \\
f_Y(2) &= \mathbb{P}(Y=2)=(1-p)p=qp \\
f_Y(3) &= \mathbb{P}(Y=3)=qqp=q^2p \\
&\vdots \\
f_Y(k) &= \mathbb{P}(Y=k)=q\cdots qp=q^{k-1}p
\end{aligned}$$

Así,

$$f_Y(y)=p(1-p)^{y-1}\mathbb{1}_{\{1,2,3,\dots\}}(y)$$

Alternativamente, sea Z es la variable aleatoria que denota el número de fracasos hasta el primer éxito, $Z=Y-1$ y $S_Z=\{0,1,2,\ldots\}$. Entonces, si $z\in S_Z$, se tiene que $\mathbb{P}(Z=z)=\mathbb{P}(Y-1=z)=\mathbb{P}(Y=z+1)=pq^z$.

Así, la *f. m. p.* de Z está dada por

$$f_Z(z)=p(1-p)^z\mathbb{1}_{\{0,1,2,\dots\}}(z)$$

Definición : En estas notas se definen los números **naturales extendidos** por $\mathbb{N}_0=\{0,1,2,3,\ldots\}=\{0\}\cup\mathbb{Z}^+$.

Nota: Se verifica que

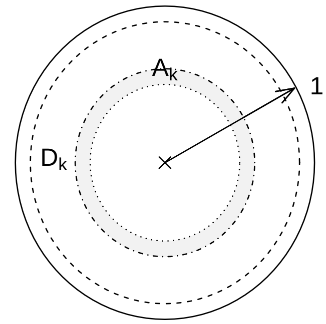
i) $f_Z(z)\geq 0$, para todo $z\in\mathbb{R}$.

ii) $\sum_{z_i\in S_Z} f_Z(z_i)=\sum_{k\in\mathbb{N}_0} p(1-p)^k=p\sum_{k=0}^{\infty} q^k=p\frac{1}{1-q}=1$.

Pues la sume es la serie geométrica convergente pues $q=1-p<1$.

Ejemplo : (Tomado de (Hoel et al. 1971)).

Considere una diana circular de radio 1, dividida en n discos concéntricos D_k, de radio k/n, $k = 1, \ldots, n$, como se muestra en la figura de la derecha.
Un dardo se lanza a la diana al azar y si cae en el aro $A_k = D_k \backslash D_{k-1}$ usted gana $n+1-k$ pesos. Sea D_0 el centro de la diana y X la *v. a.* que denota lo ganado. Determine f_X, la *f. m. p.* de X.



Solución: Note que el rango de X es $R_X = \{n, n-1, \ldots, 1\} = S_X$.

$$\mathbb{P}(X = n+1-k) = \mathbb{P}(A_k) = \mathbb{P}(D_k \backslash D_{k-1}) = \frac{\pi\left(\frac{k}{n}\right)^2 - \pi\left(\frac{k-1}{n}\right)^2}{\pi 1^2} = \frac{2k-1}{n^2}$$

para $k = 1, \ldots, n$.

Sea ahora $x = n+1-k$, luego $k = n+1-x$, y

$$\mathbb{P}(X = x) = \mathbb{P}(X = n+1-k) = \frac{2(n+1-x)-1}{n^2} = \frac{2(n-x)+1}{n^2}$$

Por lo tanto,

$$f_X(x) = \frac{2(n-x)+1}{n^2}\mathbb{1}_{\{1,2,\ldots,n\}\}}(x)$$

Note nuevamente que

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

ii) $\sum_{x_i \in S_x} f_X(x_i) = \sum_{x=1}^{n} \frac{2(n-x)-1}{n^2} = \frac{2n}{n^2}\sum_1^n 1 - \frac{2}{n^2}\sum_1^n x + \frac{1}{n^2}\sum_1^n 1 = 1.$

Ejemplo : Sea $\lambda > 0$ fija y N una *v. a.* discreta con soporte (rango) $S_N = \mathbb{N}_0 = \{0, 1, 2, \ldots\}$ y *f. m. p.* f_X dada por

$$f_N(n) = \frac{\lambda^n e^{-\lambda}}{n!}\mathbb{1}_{\mathbb{N}_0}(n)$$

Entonces N tiene el mismo soporte que Z del ejemplo anterior pero las probabilidades asignadas son distintas. Así,

$$f_Z(3) = \mathbb{P}(Z = 3) = p(1-p)^3, \qquad f_N(3) = \mathbb{P}(N = 3) = \frac{\lambda^3 e^{-\lambda}}{3!}$$

Para $p = 1/3$ y $\lambda = 2$, las primeras probabilidades serían

x	0	1	2	3
$f_Z(x)$	0.3333	0.2222	0.1481	0.0988
$f_N(x)$	0.1353	0.2707	0.2707	0.1804

Sin embargo, f_N cumple también que

i) $f_N(n) \geq 0$, para todo $n \in \mathbb{R}$.

ii) $\sum_{n \in S_N} f_N(n) = \sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} = e^{-\lambda}\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = 1$, pues la última suma es la expansión en serie de e^{λ}.

Ejemplo : Considere una rifa con n boletos (números) *igualmente probables*. Sea X la variable aleatoria que denota el número de boleto. Luego, el soporte (rango) de X es $S_X = \{1, 2, \ldots, n\}$, y $\mathbb{P}(X = x) = p$, para todo $x \in S_X$ y algún $0 < p < 1$, fijo. Entonces, $f_X(x) = p\mathbb{1}_{S_X}(x)$. Además, siendo f_X una *f. m. p.* se tiene que

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

ii) $\sum_{x_i \in S_X} f_X(x_i) = \sum_{i=1}^{n} p = np = 1$, lo que implica que $p = 1/n$. Por lo tanto,

$$f_X(x) = \frac{1}{n}\mathbb{1}_{\{1,2,\ldots,n\}}(x)$$

4.3. Funciones de distribución

Definición : Sea F una función real finita que satisface las siguiente propiedades:

a). F es una función no decreciente. Esto es, si $t_1 < t_2$, entonces $F(t_1) \leq F(t_2)$.

b). F es una función continua por la derecha. Esto es, para todo $t \in \mathbb{R}$

$$F(t^+) \equiv \lim_{\delta \to 0^+} F(t + \delta) = F(t)$$

se dice que F es una **función de distribución** sobre $\mathbb{R}$. [2]

Definición : Sea $(\Omega, \mathcal{S}, \mathbb{P})$ un espacio de probabilidad y X una variable aleatoria. Se define la **función de probabilidad acumulada** (*f. p. a.*) F_X por

$$F_X(x) = \mathbb{P}(X \leq x), \quad \text{para todo } x \in \mathbb{R}$$

Ejemplo : Considere nuevamente el ejemplo de la rifa con n boletos igualmente posibles. La *v. a.* X denota el número del boleto. $\mathbb{P}(X = x) = 1/n$, para todo $x \in S_X = \{1, \ldots, n\}$. Se vio que la correspondiente función masa de probabilidad (*f. m. p.*) está dada por $f_X(x) = \frac{1}{n}\mathbb{1}_{S_X}(x)$. Luego,

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} \mathbb{P}(X = x_i) = \frac{1}{n}\sum_{x_i \leq x} 1 = \frac{\lfloor x \rfloor}{n}$$

donde $\lfloor x \rfloor = \max\{z \in \mathbb{Z} : z \leq x\}$ representa la *función piso* de x. Así, la función de probabilidad acumulada de la *v. a.* X es

$$F_X(x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{\lfloor x \rfloor}{n} & \text{si } 1 \leq x < n \\ 1 & \text{si } x \geq 1 \end{cases}$$

La figura 4 muestra la *f. m. p.* y la *f. p. a.* del ejemplo que al asignar la misma probabilidad a cada uno de los puntos es conocida como **distribución uniforme**.

Proposición : Sea X una variable aleatoria con función de probabilidad acumulada F_X. Entonces,

a). $0 \leq F_X(t) \leq 1$, para todo $t \in \mathbb{R}$.

[2] Las funciones de distribución es un concepto del Análisis Real. Se utilizan, entre otros propósitos, para definir medidas de Lebesgue–Stieltjes y su relación con las integrales Riemann–Stieltjes, que se estudian en Teoría de la Medida. Vea por ejemplo, Ash (1972).

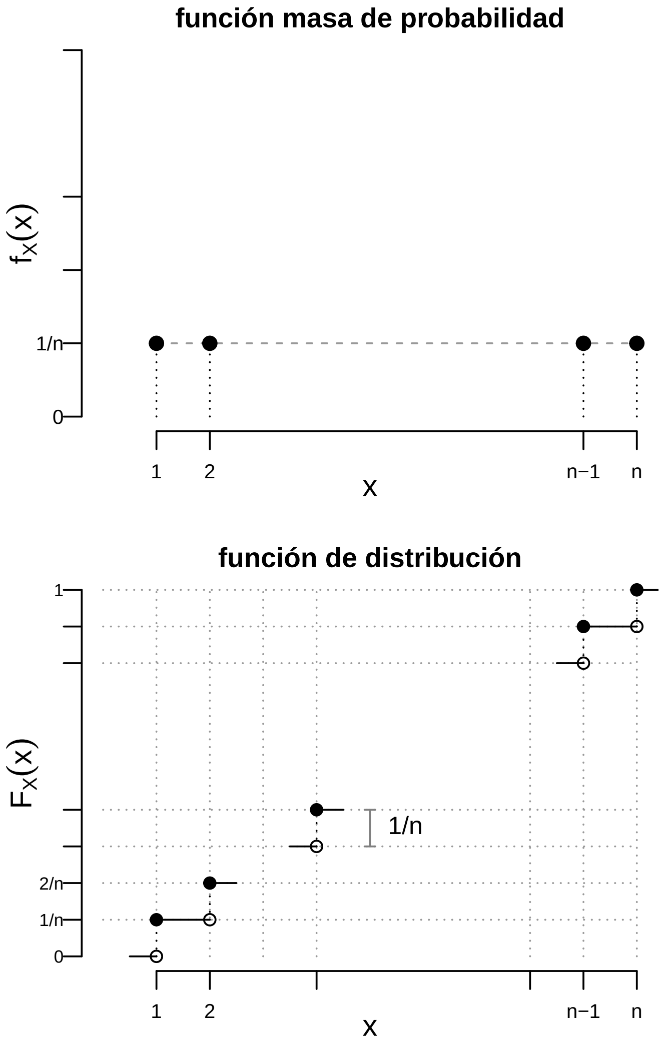


Figura 4: Distribución uniforme. Función masa de probabilidad (*f. m. p.*) y función de probabilidad acumulada (*f. p. a.*) o función de distribución.

b). F_X es una función no decreciente. Esto es, si $t_1 < t_2$, entonces $F_X(t_1) \leq F_X(t_2)$.

c). $F_X(-\infty) = 0$ y $F_X(+\infty) = 1$.

d). F_X es una función continua por la derecha. Esto es, $F_X(t^+) = F(t)$, para toda $t \in \mathbb{R}$.

Demostración:

a). Se sigue de la definición de F_X pues para todo $t \in \mathbb{R}$, $0 \leq F_X(t) = \mathbb{P}(X \leq t) \leq 1$.

b). Sean $t_1 < t_2$. Luego, $\{X \leq t_1\} \subseteq \{X \leq t_2\}$ y en consecuencia $F_X(t_1) = \mathbb{P}(X \leq t_1) \leq \mathbb{P}(X \leq t_2) = F_X(t_2)$.

c). La función F_X es monótona (b) y acotada (a), lo que garantiza la existencia tanto de los límites por la derecha de la función, $F_X(t^+)$, como por la izquierda $F_X(t^-)$. Por lo que también existen $F_X(+\infty) = \lim_{t\to\infty} F_X(t)$ y $F_X(-\infty) = \lim_{t\to-\infty} F_X(t)$.

Sean $B_n = \big\{\omega \in \Omega : X(\omega) \leq n\big\}$. Entonces, $\cdots \subseteq B_{-1} \subseteq B_0 \subseteq B_1 \subseteq \cdots$ con $\cap_{n=0}^{-\infty} B_n = \emptyset$ y $\cup_{n=0}^{\infty} B_n = \Omega$. Se sigue de la continuidad de $\mathbb{P}$,

$$F_X(-\infty) = \lim_{t\to-\infty} F_X(t) = \lim_{n\to-\infty} \mathbb{P}(B_n) = \mathbb{P}\left(\lim_{n\to-\infty} \cap_{i=0}^{n} B_i\right) = \mathbb{P}(\emptyset) = 0$$

$$F_X(+\infty) = \lim_{t\to\infty} F_X(t) = \lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}\left(\lim_{n\to\infty} \cup_{i=0}^{n} B_i\right) = \mathbb{P}(\Omega) = 1$$

d). Para mostrar que la función F_X es continua por la derecha, sea $t \in \mathbb{R}$ y sean $B_n = \{\omega \in \Omega : X(\omega) \leq t + 1/n\}$. Luego, $B_n \subseteq B_{n-1}$, $\cap_{n=1}^{\infty} B_n = \{\omega \in \Omega : X(\omega) \leq t\}$ y $\mathbb{P}\left(\cap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(X \leq t+1/n) = \mathbb{P}(X \leq t)$. Esto es, la función F_X es continua por la derecha. Similarmente, se tiene que $F_X(t^-) = \lim_{n\to\infty} F_X(t-1/n) = \mathbb{P}(X < t)$.

Corolario : Sea X una variable aleatoria con función de probabilidad acumulada F_X. Entones, F_X es una función de distribución.

Demostración: Se sigue de la proposición anterior y de la definición de función de distribución.

Proposición : Sea F una función de distribución. Entonces existe (se puede construir) un espacio de probabilidad $(\Omega, \mathcal{S}, \mathbb{P})$ y una variable aleatoria X en él tal que, X tiene a F como función de probabilidad acumulada. Esto es, para todo $t \in \mathbb{R}$, $\mathbb{P}(X \leq t) = F_X(t) = F(t)$.

Nota: Por el corolario y proposición anteriores muchos textos se refieren a F_X como la **función de distribución (de probabilidad) de** X.

En el resto de las notas se evitará el subíndice X en la función de distribución F salvo que se quiera hacer explícito que es la asociada a la *v. a.* X.

Corolario : X *v. a.* con *f. p. a.* F_X. Entonces $\mathbb{P}(X = t) = F_X(t^+) - F_X(^-)$.

Demostración: Se sigue de corolario a los axiomas pues $\{X = t\} = \{X \leq t\} \backslash \{X < t\}$.

Definición : X se dice una **variable aleatoria continua** si su función de probabilidad acumulada F_X es continua en todo $t \in \mathbb{R}$.

Corolario : Sea X una *v. a.* continua, entonces $\mathbb{P}(X = t) = 0$ para todo $t \in \mathbb{R}$.

Demostración: Sea F_X la *f. p. a.*. Se sigue de la continuidad de F_X que $\mathbb{P}(X = t) = F_X(t^+) - F_X(t^-) = 0$.

Proposición : Sea X *v. a.* con *f. p. a.* F. Entonces, si $a < b$,

$$\mathbb{P}(a < X \leq b) = F(b) - F(a)$$

Demostración: Sea $a < b$. Luego, $\{a < X \leq b\} = \{X \leq b\} \backslash \{X \leq a\}$ y de corolario a los axiomas se sigue que

$$\mathbb{P}(a < x \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F(b) - F(a)$$

Corolario ; Sea X es una *v. a.* continua y $a < b$. Entonces,

$$\mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b)$$

Ejemplo : Considere una sucesión de ensayos Bernoulli con probabilidad de éxito $p = 0.3$. Sea Y el número de fracasos hasta el primer éxito.

a). Determine $\mathbb{P}(Y = 3)$.

b). Calcule $\mathbb{P}(2 \leq Y \leq 4)$.

c). Determine F_Y, la *f. p. a.* de Y.

d). Calcule $\mathbb{P}(3 \leq Y \leq 10)$.

e). Determine $\mathbb{P}(Y \geq 5)$.

Solución: Sea Y el número de fracasos hasta el primer éxito. Luego, el soporte (rango) de Y es $S_Y = \{0, 1, 2, \dots\}$ y con $q = 1 - p$,

$$f_Y(y) = \mathbb{P}(Y = y) = pq^y \mathbb{1}_{S_Y}(y)$$

a). $\mathbb{P}(Y = 3) = pq^3 = .3(1 - .3)^3 = 0.1039$.

b). Calcule $\mathbb{P}(2 \leq Y \leq 4) = p[q^2 + q^3 + q^4] = .3(.7)^2[1 + .7 + .7^2] = 0.3219$.

c). Sea $y \in \{0, 1, 2, \dots\}$,

$$F_Y(y) = \mathbb{P}(Y \leq y) = \sum_{k=0}^{y} f_Y(k) = p \sum_{k=0}^{y} q^k = p\frac{1 - q^{y+1}}{1 - q} = 1 - q^{y+1}$$

Entonces,

$$F_Y(y) = \begin{cases} 0 & \text{si } y < 0 \\ 1 - q^{\lfloor y \rfloor + 1} & \text{si } 0 \geq y \end{cases}$$

Entonces, $F_Y(y) = 1 - (.7)^{\lfloor y \rfloor + 1}$, para $y \geq 0$. Note que $F_Y(0) = 1 - .7 = 0.3 = p$.

d). $\mathbb{P}(3 \leq Y \leq 10) = \mathbb{P}(2 < Y \leq 10) = F_Y(10) - F_Y(2) = .9718 - .657 = 0.3148$.

e). $P(Y \geq 5) = 1 - \mathbb{P}(Y < 5) = 1 - F_Y(4) = (.7)^5 = 0.1681$.

Ejercicio : Sea T una *v. a.* con *f. p. a.* F dada por

$$F(t) = \begin{cases} 0 & \text{si } t < \theta \\ 1 - \left(\frac{\theta}{t}\right)^{\alpha} & \text{si } t \geq \theta \end{cases}$$

con $\theta > 0$ y $\alpha > 1$, fijos

a). Muestre que la función F satisface las propiedades de una *f. p. a.*.

b). Con $\theta = 2$ y $\alpha = 3$, verifique:

a) $\mathbb{P}(T \geq 5) = 0.064$.

b) $\mathbb{P}(4 \leq T \leq 6) = 0.0880$.

c) $\mathbb{P}(4 \leq T \leq 6 | T \geq 5) = 0.4213$.

Definición : De manera alternativa se define la **variable aleatoria** X en el espacio de probabilidad $(\Omega, \mathcal{S}, \mathbb{P})$ como la función real con dominio en Ω, tal que, para todo $x \in \mathbb{R}$, $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{S}$.

Ejemplo : Se hace girar una flecha sujeta a un tablero circular. Sea θ el ángulo final de la flecha. Luego, $0 < \theta < 2\pi$. La probabilidad de que la flecha apunte en un subintervalo de $(0, 2\pi]$ es proporcional a la longitud del subintervalo. Sea X la *v. a.* definida por $X(\theta) = \theta/2\pi$. Encuentre F, la función de distribución de X.

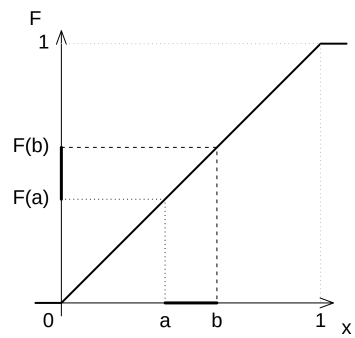
Solución: Note que ningún valor de θ lleva a $x < 0$, por lo que $F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\emptyset) = 0$, si $x < 0$. Por otro lado, si $\leq x \leq 1$, el evento $\{X \leq x\}$ ocurre si y solo si $\{\theta \leq 2\pi x\}$. Por lo que $F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\Theta \leq 2\pi x) = \frac{2\pi x}{2\pi} = x$. Finalmente, note que si $x > 1$, todos los valores de θ llevan a que $X(\theta) \leq x$. Así,

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

En este caso, se dice que la *v. a.* X *se distribuye uniformemente* en el intervalo $[0, 1]$.

Si $0 < a < b < 1$,

$$\mathbb{P}(a \leq X \leq b) = F(b) - F(a) = b - a$$



Note que a diferencia de este ejemplo de los boletos para la rifa y el número de ensayos hasta el primer éxito, las correspondientes *f. p. a.* son funciones escalonadas no decrecientes con los brincos localizados en los puntos del soporte. Por otro lado, los dos últimos ejemplos la *f. p. a.* son funciones continuas no decrecientes.

4.4. Variables aleatorias continuas

Definición : Sea X una *v. a.* con *f. p. a.* F_X. Si F_X es continua en todo $\mathbb{R}$, X se dice un **variable aleatoria continua**. Si además F_X es una función suave tal que para todo $x \in \mathbb{R}$,

$$F_X(x) = \int_{-\infty}^{x} h(u)du$$

para alguna función no negativa h, X se dice **absolutamente continua (con respecto a la integral)** y h su correspondiente **función de densidad de probabilidad (*f. d. p.*)**.

Proposición : Sea X una *v. a.* continua con *f. p. a.* F_X y *f. d. p.* f_X. Entonces,

a). $F_X(x) = \displaystyle\int_{-\infty}^{x} f_X(u)du, \quad$ para todo $x \in \mathbb{R}$.

b). $f_X(x) = \dfrac{dF_X(x)}{dx}$.

Nota: Es posible que X *v. a.* continua con *f. p. a.* F_X sea no derivable. En este caso, X no tiene función de densidad.

Proposición : Sea X una *v. a.* continua con *f. d. p.* f_X y *f. p. a.* F_X. Entonces,

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

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

Demostración:

a). Se sigue de la defición de *f. d. p.*, f_X debe ser necesariamente no negativa.

b). $F_X(x) = \int_{-\infty}^{x} f_X(x)dx$. Luego,

$$1 = \lim_{x\to\infty} F_X(x) = \int_{-\infty}^{\infty} f_X(u)du$$

Definición : Una función f que satisface las propiedades a) y b) anteriores se dice **función de densidad propia** o legítima.

Ejemplo : Sea X la *v. a.* distribuida uniformemente en $[0,1]$ tiene como *f. p. a.* F_X donde

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

Entonces, su *f. d. p.* está dada por

$$f_X(x) = \frac{d}{dx}F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } 0 \leq x < 1 \\ 0 & \text{si } x \geq 1 \end{cases}$$

Luego, $f_X(x) = \mathbb{1}_{(0,1)}(x)$.

Nota:

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

ii) $\displaystyle\int_{-\infty}^{\infty} f_X(u)du = \int_{-\infty}^{\infty} \mathbb{1}_{(0,1)}(u)du = \int_{0}^{1} 1\,du = 1$

Ejemplo : Considere el lanzamiento al azar un dardo a una diana circular de radio R. Sea Y la *v. a.* que denota la distancia del centro al dardo. Determine la *f. p. a.* y la *f. d. p.* de Y.

Solución:

a). *f. p. a.*, F_Y:

i) $F_Y(y) = 0$, si $y < 0$.

ii) $F_Y(y) = \mathbb{P}(Y \leq y) = \frac{\text{Área disco radio } y}{\text{Área disco radio } R} = \frac{\pi y^2}{\pi R^2} = \frac{y^2}{R^2}$, si $0 \leq y < R$.

iii) $F_Y(y) = 1$, si $y \geq R$.

b). *f. d. p.*, f_X:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{2y}{R^2}\mathbb{1}_{(0,R)}(y)$$

c). Si $R = 5$, $\mathbb{P}(Y \leq 2) = \left(\frac{2}{5}\right)^2 = \frac{4}{25} = 0.16$.

d). Note que $\displaystyle\int_{-\infty}^{\infty} f_Y(y)dy = \int_{0}^{R} \frac{2y}{R^2}dy = \left.\frac{y^2}{R^2}\right|_{0}^{R} = 1$.

Ejemplo : En una estación de servicio hay 2 bombas para despachar gasolina cada una con su tanque de almacenamiento y capacidad de 1 m^3. Sea W la *v. a.* que denota la cantidad despachada con *f. d. p.* dada por

$$f_W(w) = w\mathbb{1}_{(0,1]}(w) + (2-w)\mathbb{1}_{(1,2]}(w)$$

a). Grafique la *f. d. p.* f_W.

b). Detemine la *f. p. a.* F_W y grafíquela.

c). Calcule $\mathbb{P}(.8 \leq W \leq 1.2)$.

d). Determine $\mathbb{P}(W \geq 1.5|X > 0.9)$.

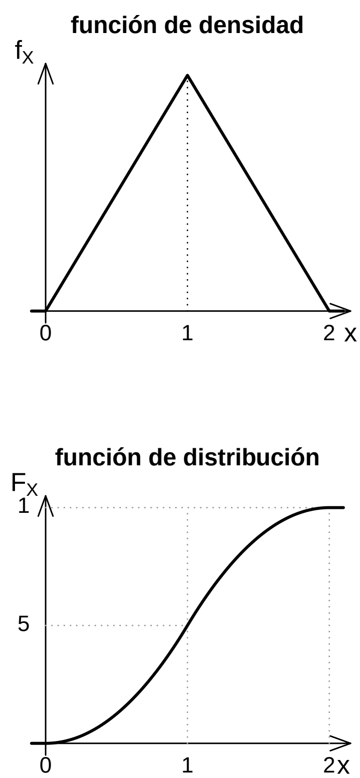
Solución:

a).

$$f_W(w) = w\mathbb{1}_{(0,1]}(w) + (2-w)\mathbb{1}_{(1,2]}(w)$$

El panel superior de la gráfica muestra la función de densidad f_W.

b).



i) $w < 0$.

$$F_W(w) = \int_{-\infty}^{w} f_W(u)du = \int_{-\infty}^{w} 0du = 0$$

ii) $0 \leq w < 1$.

$$F_W(w) = \int_{-\infty}^{w} f_W(u)du = \int_{0}^{w} udu = \frac{w^2}{2}$$

ii) $1 \leq w < 2$.

$$\begin{aligned} F_W(w) &= \int_{-\infty}^{w} f_W(u)du \\ &= \int_{0}^{1} f_W(u)du + \int_{1}^{w} f_W(u)du \\ &= \frac{1}{2} + \int_{1}^{w} (2-u)du \\ &= -1 + 2w - \frac{w^2}{2} \end{aligned}$$

El panel inferior de la gráfica muestra la correspondiente *f. p. a.* F_W, de la *v. a.* W.

$$F_W(w) = \begin{cases} 0 & \text{si } \ w < 0 \\ w^2/2 & \text{si } \ 0 \leq w < 1 \\ -1 + 2w - w^2/2 & \text{si } \ 1 \leq w < 2 \\ 1 & \text{si } \ w \geq 2 \end{cases}$$

Note que $F_W(0) = 0$, $F(1) = 1/2$, y $F_W(2) = 1$.

c). $\mathbb{P}(.8 \leq W \leq 1.2) = F_W(1.2) - F_W(.8) = 0.36$.

d).

$$\mathbb{P}(W \geq 1.5|W > .9) = \frac{\mathbb{P}(W \geq 1.5)}{1 - \mathbb{P}(W \leq .9)} = \frac{.125}{1 - .405} = 0.2100$$

Ejemplo : Recuerde X *v. a.* con *f. p. a.* dada por

$$F(x) = \begin{cases} 0 & \text{si } x < \theta \\ 1 - \big(\theta/x\big)^{\alpha} & \text{si } x \geq \theta \end{cases}$$

con $\theta > 0$ y $\alpha > 1$, fijos. Encuentre ka correspondiente *f. d. p.* de X y verifique que es una función de densidad propia.

Solución:

$$f_x(x) = \frac{d}{dx} F_X(x) = \alpha\theta^{\alpha} x^{-\alpha-1} \mathbb{1}_{(\theta,\infty)}(x)$$

Además,

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

ii) $\displaystyle\int_{\mathbb{R}} f_X(u)du = \int_{\theta}^{\infty} \alpha\theta^{\alpha} x^{-\alpha-1} dx = \alpha\theta^{\alpha} \frac{x^{-\alpha}}{-\alpha}\Big|_{\theta}^{\infty} = 1$

por lo que f_x es efectivamente una función de densidad propia.

Ejemplo : (Tomado de León-García (2008).)

El tiempo de espera T de un taxi es cero si el cliente encuentra un auto disponible en el sitio y se distribuye *uniformemente* a lo largo de una hora $(0, 1]$ sin no hay autos en el sitio. Si la probabilidad de encontrar un taxi en el sitio es p,

a). Determine la *f. p. a.* F_T y grafíquela.

b). Determine la función de probabilidad f_T y grafique.

Solución:

a). *i)* Si $t < 0$, $F_T(t) = \mathbb{P}(T < t) = 0$.

ii) Si $0 \leq t < 1$, se sigue del teorema de probabilidad total (TPT),

$$\begin{aligned} F_T(t) &= \mathbb{P}(T \leq t) \\ &\stackrel{\text{TPT}}{=} \mathbb{P}\big(T \leq t|\text{taxi en el sitio}\big)\, p + \mathbb{P}\big(T \leq t|\text{no taxi}\big)\,(1-p) \\ &= 1\,p + t(1-p) \end{aligned}$$

pues si no hay taxi en el sitio el tiempo de espera sigue se distribuye uniformemente en $(0, 1]$ y la consecuente *f. p. a.* es $F(t) = t$,

iii) Si $t > 1$, $F_T(t) = \mathbb{P}(T \leq 1) = 1$,

b). La correspondiente función de probabilidad es mixta pues tiene un punto de probabilidad positiva ($t = 0$) y ek resto se comporta como *v. a.* continua uniforme. Entonces,

$$f_T(t) = p\mathbb{1}_{\{0\}}(t) + (1-p)\mathbb{1}_{(0,1]}(t)$$

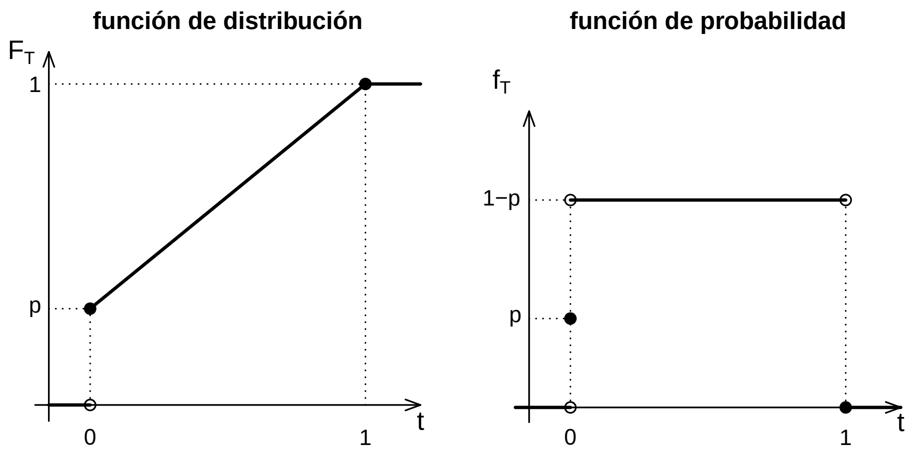


Figura 5: Funciones de distribución y de probabilidad de la variable aleatoria *mixta* T, el tiempo de espera del taxi. Los puntos llenos ($\bullet$) y vacíos ($\circ$) representan si la función alcanza el valor en el punto o no, respectivamente.

La figura 5 muestra las funciones de distribución (*f. p. a.*) y de probabilidad (*f. d. p.*) del tiempo de espera T,

Definición : Sea X una *v. a.* con *f. p. a.* F_X, se dice de **tipo mixto** (discreta–continua) si F_X tiene saltos en $\{x_1, x_2, \dots\}$ conjunto numerable pero también es creciente en el intervalos. En el ejemplo anterior, F_X tiene la forma

$$F_X(x) = pF_1(x) + (1-p)F_2(x), \quad x \in \mathbb{R}$$

para algún $0 < p < 1$ y con F_1 la *f. p. a.* de una *v. a.* del tipo discreta y F_2 la *f. p. a.* de una *v. a.* del tipo continuo.

4.5. Función de supervivencia

Definición : Sea T una *v. a.* con *f. p. a.* F_T. Se define la **función de supervivencia de** T por

$$G_T(t) = \mathbb{P}(T > t) = 1 - F_T(t), \quad t \in \mathbb{R}$$

Ejemplo : Sea T una *v. a.* con función de distribución F_T dada por $F_T(t) = 1 - e^{-\lambda t}$, para $t > 0$ y algún $\lambda > 0$, fijo.

a). Muestre que F_T satisface las propiedades de una *f. p. a.*.

b). Grafique F_T y G_T, la función de supervivencia de T.

c). Determine f_T, la *f. d. p.* de T y grafíquela.

d). Para $\lambda = 3$, calcule $\mathbb{P}(4 \le T \le 6 | T \ge 5)$.

Solución:

a). F_T satisface las propiedades de una función de distribución de probabilidad. A saber,

i) $F_T(0) = 0$ y $F_T(+\infty) = 1$. Por lo tanto $0 \le F_T(t) \le 1$, para todo $t \in \mathbb{R}$.

ii) $e^{-\lambda t} \searrow 0$ monótonamente, luego $F_T = 1 - e^{-\lambda t}$ es creciente.

iii) $F_T(0) = 0$ y $F_T(+\infty) = 1$.

iv) $e^{-\lambda t}$ es una función continua, luego F_T también lo es.

b). La figura 6 muestra las funciones de distribución F_T, de supervivencia G_T y de densidad f_T de la variable aleatoria T distribuida *exponencialmente*.

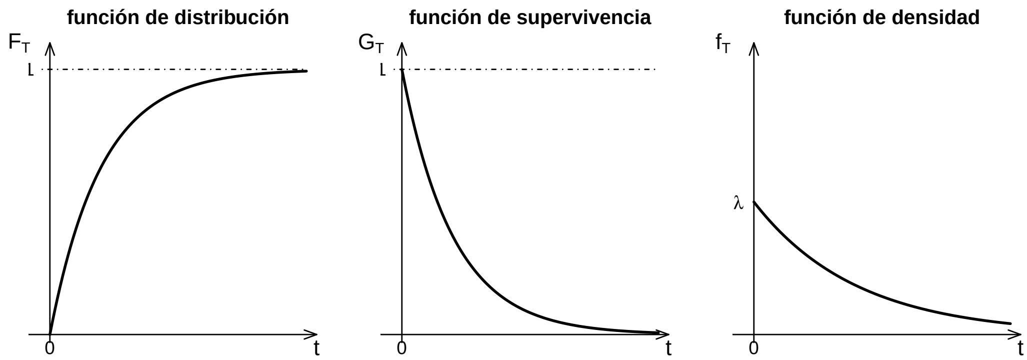


Figura 6: Funciones de distribución F_T, de supervivencia G_T y de densidad f_T de la variable aleatoria T distribuida *exponencialmente*.

c). La función de densidad queda como la derivada de la función de distribución. Así,

$$f_T(t) = \frac{d}{dt} F_T(t) = \lambda e^{-\lambda t}, \quad t > 0$$

d). Si $\lambda = 3$, se tiene

$$\begin{aligned}
\mathbb{P}(4 \leq T \leq 6 | T \geq 5) &= \frac{\mathbb{P}\left(\{4 \leq T \leq 5\} \cap \{T \geq 5\}\right)}{\mathbb{P}(T \geq 5)} \\
&= \frac{\mathbb{P}(5 \leq T \leq 6)}{\mathbb{P}(T \geq 5)} \\
&= \frac{F_T(6) - F_T(5)}{1 - F_T(5))} \\
&= 1 - e^{-\lambda \cdot 1} \\
&= \mathbb{P}(T \leq 1) \\
&= 0.950
\end{aligned}$$

Nota: En este caso particular (*distribución exponencial*) se tiene que

$$\mathbb{P}(4 \leq T \leq 6 | T \geq 5) = \mathbb{P}(T \leq 1)$$

que se interpreta como que T *"olvida"* que ya ocurrió $\{T > 5\}$. Se verá más adelante que esta propiedad de *"pérdida de memoria"* es exclusiva de la distribución exponencial si T es una *v. a.* continua. En el caso discreto, la pérdida de memoria se da exclusivamente con la *distribución geométrica*, empleada para modelar el número de experimentos hasta el primer éxito en una sucesión de ensayos Bernoulli.

4.6. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 4, Barrios and Heiras (2024).

5. Características de una Variable Aleatoria

5.1. Descripción de la distribución de probabilidades

Sea $(\Omega, \mathcal{S}, \mathbb{P})$ un espacio de probabilidad y X una variable aleatoria definida para representar o modelar cierta característica de las salidas del experimento aleatorio $\mathcal{E}$. La **ley de probabilidades** es una manera informal de referirse o describir a el comportamiento aleatorio de X. Rigurosamente este comportamiento queda completamente determinado por su función de distribución (*f. p. a.*) y/o su función masa de probabilidad (*f. m. p.*) o función de densidad de probabilidad (*f. d. p.*), según la naturaleza discreta o continua de X, pues recuerde:

a). Discreto:

i) $F_X(x) = \sum_{x_i \le x} f_X(x_i)$

ii) $f_X(x) = F_X(x^+) - F_X(x^-)$

b). Continuo:

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

ii) $f_X(x) = \frac{d}{dx} F_X(x)$

Uno puede responder prácticamente cualquier pregunta contando con una de las funciones anteriores pues en principio teniendo una se obtiene la otra y podría uno determinar la probabilidad de cualquier evento.

Por otro lado, en ocasiones puede uno darse cierta idea de cómo es la distribución de probabilidades si el detalle de las funciones anteriores con algunas características e indicadores que se definen ahora. A saber, *cuantiles*, *moda(s)* y *momentos*.

Definición : Sea X una variable aleatoria (va) con función de probabilidad acumulada (*f. p. a.*) F_X y $0 < p < 1$. Se define el **p-ésimo cuantil** de la *v. a.* X (o de la distribución de X) y se denota por x_p a

$$x_p = \min\{x \in \mathbb{R} : p \le F_X(x)\}$$

Luego, siempre se tiene que $p \le F_X(x_p)$. Se sigue que si F_X es continua, entonces $p = F_X(x_p)$. La figura 7 ilustra los casos para distribuciones discretas y continuas.

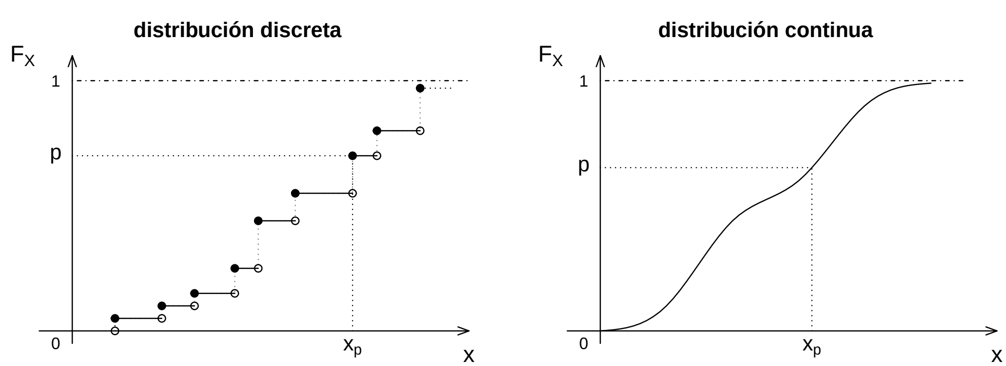


Figura 7: Localización del p-ésimo cuantil en funciones de distribución (*f. p. a.*), casos discreto y continuo.

Note que si X es una *v. a.* discreta con soporte S_X y *f. p. a.* F y si $0 < p < 1$, tal que $p < F(x_p)$, en ocasiones se redefine el p-ésimo cuantil x_p por

$$x_p = x_{k-1} + \left[\frac{x_k - x_{k-1}}{p_k - p_{k-1}}\right](p - p_{k-1})$$

resultado de la **interpolación lineal** en la *f. p. a.* F_X. Más formalmente, si $x_{k-1}, x_k \in S_X$ tales que $x_{k-1} < x_k$ y $F_X(x_{k-1}) = p_{k-1} < p$ y $F_X(x_k) = p_k > p$, entonces sería razonable esperar que el p-ésimo cuantil se localice entre x_{k-1} y x_k como se muestra en la figura 8.

$$x_p = x_{k-1} + \left[\frac{x_k - x_{k-1}}{p_k - p_{k-1}}\right](p - p_{k-1})$$

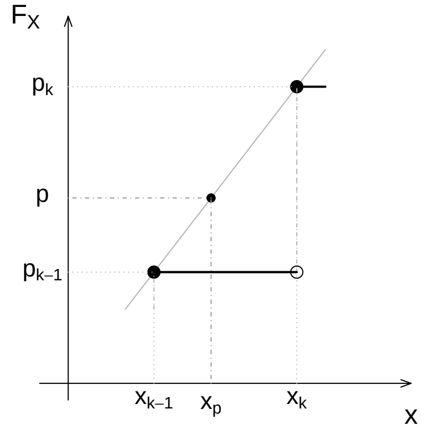


Figura 8: Interpolación lineal para la estimación del p-ésimo cuantil de una distribución discreta.

Definición : La siguiente tabla muestra algunos cuantiles representativos que ayudan a describir la distribución de X.

concepto	notación	probabilidad acumulada
primer cuartil	q_1	0.25
mediana	x_{med}	0.50
tercer cuartil	q_3	0.75
primer decil	d_1	0.10
⋮		
noveno decil	d_9	0.90

La figura 9 presenta la mediana, primer y tercer cuartil de la distribución mostrada en las gráficas de la función de densidad y función de distribución.

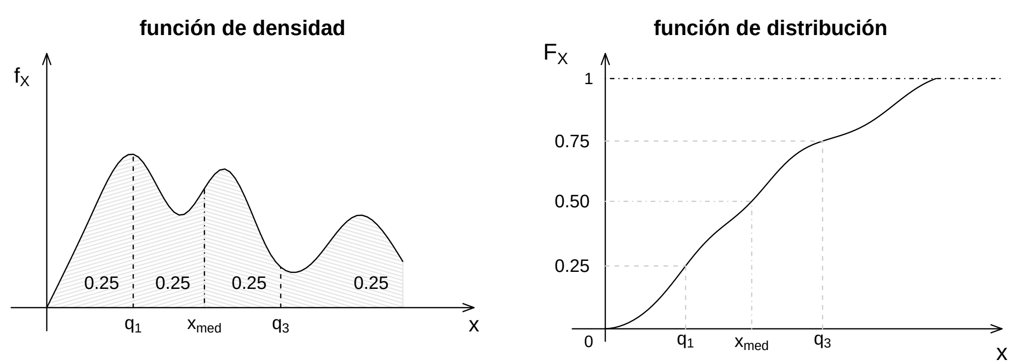


Figura 9: Localización de la mediana x_{med}, el primer q_1 y tercer q_3 cuartil de la distribución de X.

Note que la mediana x_{med} *"divide en dos partes iguales"* la distribución de probabilidades. En este sentido la mediana es una *medida de la localización* de las distribución de la variable aleatoria.

Definición : Se define el **rango intercuartílico** como la distancia entre cuartiles,

$$R_q = q_3 - q_1$$

Note que el rango intercuartílico abarca el 50 % de la probabilidad pues necesariamente $0.5 = F(q_3) - F(q_1)$. Luego, es una *medida de la dispersión* de la variable aleatoria en el sentido que, si $R_q(X) < R_q(Y)$, se diría que la distribución de la *v. a.* Y es más dispersa que la de X, o bien, que la variación de Y es mayor que la de X.

En ocasiones, por ejemplo en las ciencias sociales, es más común expresar los cuantiles en términos porcentuales, conocidos como **percentiles**. Así, el .05-cuantil se dice el *5-percentil.*

Definición : Se define la **moda** de la variable aleatoria X (o de su distribución) al valor de x en el soporte de la distribución donde se localiza un máximo de la función de densidad (*f. d. p.*) o masa de probabilidad (*f. m. p.*). Si la función de probabilidad tiene máximos locales la distribución se dice **multimodal**. La figura 10 muestra el caso de distribuciones con una, dos y más modas.

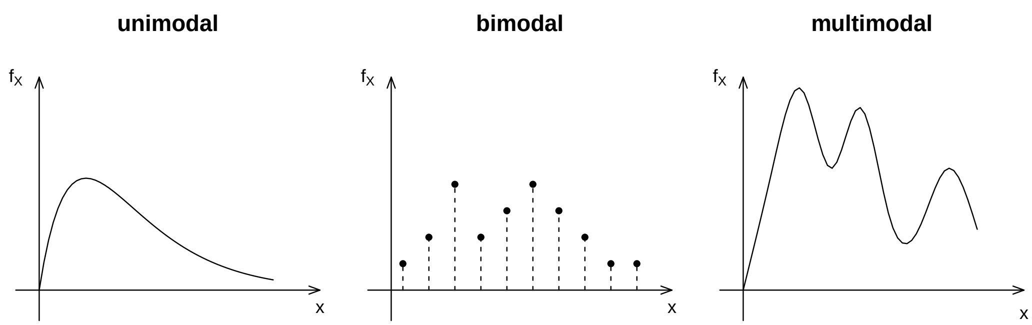


Figura 10: Funciones de densidad de probabilidad unimodal, bimodal y multimodal. El panel central corresponde a una distribución discreta.

5.2. Valor esperado

Definición : Sea X una variable aleatoria discreta con función masa de probabilidad f_X y soporte $S_X = \{x_1, x_2, \dots\}$. Se define el **valor esperado de** X por

$$\mathbb{E}[X] = \sum_{x_i \in S_X} x_i f_X(x_i)$$

siempre que $\sum_{x_i \in S_X} |x_i| f_X(x_i) < \infty$. $\mathbb{E}[X]$ se conoce también como la **media** o **esperanza matemática** de la *v. a.* X y en ocasiones se denota por μ_X o simplemente μ.

Nota: Si $\sum_{x_i \in S_X} |x_i| f_X(x_i) = \infty$, se dice que X no tiene valor esperado.

Ejemplo : Sea X una *v. a.* discreta distribuida uniformemente en $S_X = \{x_1, \dots, x_n\}$. Entonces, necesariamente $\mathbb{P}(X = x_i) = p$, para toda $x_i \in S_X$ y por lo tanto $p = 1/n$. Luego,

$$\mathbb{E}[X] = \sum_{x_i \in S_X} x_i f_X(x_i) = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$$

el *promedio* de las x_i's.

Ejemplo : Sea Y una *v. a.* discreta con soporte $S_Y = \{y_1, \ldots, y_n\}$ y función masa de probabilidad f_Y.

$$\mathbb{E}[Y] = \sum_{y_i \in S_Y} y_i f_Y(y_i) = \sum_{i=1}^{n} y_i \mathbb{P}(Y = y_i)$$

Luego, el valor esperado es un *promedio ponderado* de las y_i's, con pesos $p_i = \mathbb{P}(Y = y_i)$.

Ejemplo : Sea X una *v. a.* que representa un ensayo Bernoulli con probabilidad de éxito p. Luego, $S_X = \{0, 1\}$ y $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 0) = 1 - p = q$ y

$$\mathbb{E}[X] = 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) = p$$

La figura 11 muestra el caso binario. Suponga que el volumen de las pelotas refleja el peso de la probabilidad y la distancia al apoyo el valor que toma la variable. El punto de apoyo ▲ que permite que la barra quede horizontalmente balanceada se localiza en el valor esperado $\mathbb{E}[X]$.

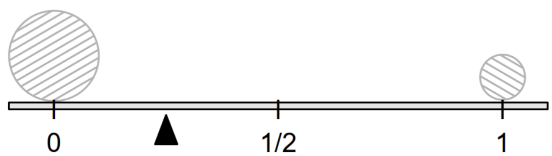


Figura 11: Variable binaria X. El volumen de las pelotas refleja la probabilidad (peso) del punto. El apoyo ▲ se localiza en el valor esperado $\mathbb{E}[X]$.

Esto sucede en general. Si X es una *v. a.* discreta con *f. m. p.* f_X y soporte $S_X = \{x_1, x_2, \ldots\}$ y valor esperado finito

$$\mathbb{E}[X] = \sum_{x_i \in S_X} x_i f_X(x_i)$$

localiza el *"centro de masa (gravedad o centroide) de los datos"*. Los x_i representan la localización de la masa (peso) $p_i = f_X(x_i)$.

Ejemplo : Sea X una *v. a.* que denota el número de éxitos en n ensayos Bernoulli. Se dice que X sigue una **distribución binomial**, siendo su soporte $S_X = \{0, 1, \ldots, n\}$ y *f. m. p.*

$$f_X(x) = \binom{n}{x} p^x q^{n-x}$$

con $q = 1 - p$. Luego se tiene,

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=0}^{n} x \, \frac{n!}{x!(n-x)!} \, p^x q^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)(n-x)!} p^{x-1} q^{(n-1)-(x-1)}] \\
&= np \sum_{y=0}^{n-1} \frac{(n-1)!}{y! \, (n-1-y)!} p^y q^{(n-1)-y} \\
&= np
\end{aligned}$$

Ejemplo : Sea Y la *v. a.* que denota el número de ensayos Bernoulli (con probabilidad de éxito p) hasta el primer éxito. Luego, $S_Y = \{1, 2, 3, \dots\}$ y $f_Y(y) = \mathbb{P}(Y = k) = pq^{k-1}$. Y se dice que sigue una **distribución geométrica**. Entonces,

$$\begin{aligned}
\mu &= \sum_{k=1}^{\infty} kpq^{k-1} \\
\mu &= p\left[1 + 2q + 3q^2 + 4q^3 + \cdots\right] \\
q\mu &= p\left[q + 2q^2 + 3q^3 + 4q^4 + \cdots\right] \\
\mu(1-q) &= p\left[1 + q + q^2 + q^3 + \cdots\right] \\
&= p\frac{1}{1-q} \\
&= 1
\end{aligned}$$

Por lo tanto, $\mu = \mathbb{E}[Y] = \dfrac{1}{1-q} = \dfrac{1}{p}$.

Ejemplo : Sea X una *v. a.* con *f. m. p.* $f_X(x) = \frac{1}{x(x+1)}\mathbb{1}_{\{1,2,3,\dots\}}(x)$. Muestre que la *f. m. p.* es propia pero X no tiene valor esperado.

Solución:

i) $f_X(x) = \frac{1}{x(x+1)}\mathbb{1}_{\{1,2,3,\dots\}}(x) \geq 0$, para todo $x \in \mathbb{R}$.

ii) $\sum_{x=1}^{\infty} \frac{1}{x(x+1)} = \text{lím}_{n\to\infty} \sum_{x=1}^{n} \frac{1}{x(x+1)} = \text{lím}_{n\to\infty} \frac{n}{n+1} = 1$.

iii) $\sum_{x=1}^{\infty} xf_X(x) = \sum_{x=1}^{\infty} \frac{x}{x(x+1)} = \sum_{x=1}^{\infty} \frac{1}{x+1} \to \infty$.

Los primeros dos incisos verifican que f_X es propia. El tercero muestra que X no tiene valor esperado.

Ejercicio . Sea $\lambda > 0$ fijo y N la *v. a.* con *f. m. p.* dada por

$$f_N(n) = \frac{\lambda^n e^{-\lambda}}{n!}\mathbb{1}_{\{0,1,2\dots\}}(n)$$

Muestre que $\mathbb{E}[N] = \lambda$.

Teorema del estadístico inconsciente (TEI, LOTUS) Caso discreto. (Rincón (2014); Hoel et al. (1971)) Sea X una variable aleatoria discreta con función masa de probabilidad f_X y soporte $S_X = \{x_1, x_)2, \dots\}$. Sea $g : \mathbb{R} \to \mathbb{R}$ una función medible tal que $g(X)$ es una variable aleatoria con valor esperado finito. Entonces,

$$\mathbb{E}[g(X)] = \sum_{x_i \in S_X} g(x_i) f_X(x_i)$$

Demostración: Sea $Y = g(X)$. Y es una *v. a.* discreta con soporte $S_Y = \{y_1, y_2, \dots\}$. Luego, $y_j \in S_Y$ si existe al menos un $x_i \in S_X$ tal que $y_j = g(x_i)$. Sean $A_j = \left\{x_i \in S_X : g(x_i) = y_j\right\}$. Entonces,

$$\mathbb{P}(Y = y_j) = \mathbb{P}(X \in A_j) = \sum_{x_i \in A_j} f_X(x_i)$$

Se sigue que los A_j son disjuntos y $\cup j A_j = S_X$. Así,

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_{y_j \in S_Y} y_j \mathbb{P}(Y = y_j) \\
&= \sum_{y_j} y_j \sum_{x_i \in A_j} \mathbb{P}(X = x_i) \\
&= \sum_{y_j} \sum_{x_i \in A_j} g(x_i) f_X(x_i) \\
&= \sum_{x_i \in S_X} g(x_i) f_X(x_i)
\end{aligned}$$

Nota: Sea $A \in \mathcal{B}(\mathbb{R})$. Defina $X = \mathbb{1}_A$, con $\mathbb{1}_A$ la función indicadora del evento A. Luego, X es una *v. a.* Bernoulli con *"éxito"* $(X = 1)$ si ocurre A y *"fracaso"* $(X = 0)$, si no ocurre A y con probabilidades $p = \mathbb{P}(X = 1) = \mathbb{P}(A)$ y $q = 1 - p = \mathbb{P}(X = 0) = \mathbb{P}(A^C)$, respectivamente. Entonces,

$$\mathbb{E}[X] = 0\,\mathbb{P}(X = 0) + 1\,\mathbb{P}(X = 1) = p = \mathbb{P}(A) = \mathbb{E}[\mathbb{1}_A]$$

Definición : Sea X una variable aleatoria continua con función de densidad de probabilidad f_X. Se dice quie X tiene valor esperado finito si $\int_{\mathbb{R}} |x| f_X(x)dx < \infty$, y en tal caso, se define su **valor esperado o media** por

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x)dx \equiv \mu_X$$

Ejemplo : Sea U la *v. a.* que sigue la **distribución uniforme** en el intervalo $(0, 1)$. Se ha visto que las correspondientes *f. p. a.* y *f. d. p.* están dadas por

$$F_U(u) = \begin{cases} 0 & \text{si } u < 0 \\ u & \text{si } 0 \leq u < 1 \\ 1 & \text{si } u \geq 1 \end{cases} \qquad \text{y} \qquad f_U(u) = \mathbb{1}_{(0,1)}(u)$$

respectivamente. Entonces,

$$\mu_U = \mathbb{E}[U] = \int_{\mathbb{R}} u f_U(u)du = \int_{\mathbb{R}} u \mathbb{1}_{(0,1)} du = \int_0^1 u du = \frac{1}{2}$$

Note que en este caso la media $\mu_U = 1/2$ coincide con la mediana $u_{0.5}$, pues $0.5 = F_U(1/2) = 1/2$. Esto no siempre es el caso, considere por ejemplo $\alpha > 1$ y $\theta > 0$ fijos y Y una *v. a.* con *f. p. a.* $F_Y(y) = 1 - (\theta/y)^\alpha \mathbb{1}_{(\theta,\infty)}(y)$ y *f. d. p.* $f_Y(y) = \alpha\theta^\alpha y^{-\alpha-1} \mathbb{1}_{(\theta,\infty)}(y)$. Y se dice que sigue una **distribución de Pareto**. Entonces,

I). Media:

$$\begin{aligned}
\mathbb{E}[Y] &= \int_{\mathbb{R}} y f_Y(y)dy \\
&= \int_{\mathbb{R}} y\alpha\theta^\alpha y^{-\alpha-1} \mathbb{1}_{(\theta,\infty)}(y) \\
&= \alpha\theta \int_\theta^\infty y^{-\alpha} dy \\
&= \frac{\alpha}{\alpha - 1}\theta
\end{aligned}$$

II). Mediana:
$$\frac{1}{2} = F_Y(y) = 1 - \left(\frac{\theta}{y}\right)^{\alpha}$$
que resolviendo la ecuación para y, $y_{\text{med}} = y_{0.5} = \theta 2^{1/\alpha}$.

III). Si por ejemplo, $\alpha = 2$ y $\theta = 3$,
$$\mu_Y = \frac{2}{1}3 = 6; \qquad y_{0.5} = 3\sqrt{2} = 4.2426$$

IV). Note que si $\alpha = 1$, Y no tiene valor esperado.

Ejemplo : Sea T una *v. a.* con *f. d. p.* $f_T(t) = \lambda e^{-\lambda t}\mathbb{1}_{(0,\infty)}(t)$, para algún $\lambda > 0$ fijo. Se dice que T sigue la **distribución exponencial**. Se ha visto ya que la correspondiente *f. p. a.* es
$$FT(t) = \mathbb{P}(T \leq t) = 1 - e^{\lambda t}, \quad t \geq 0$$

I). Media:
$$\mathbb{E}[T] = \int_{\mathbb{R}} t f_T(t)dt = \int_{\mathbb{R}} t\lambda e^{-\lambda t}\mathbb{1}_{(0,\infty)}(t)dt = \lambda\int_0^{\infty} te^{\lambda t}dt = \frac{1}{\lambda} = \mu_T$$
donde se aplicó un paso de integración por partes.

II). Mediana: $0.5 = F_T(t) = 1 - e^{-\lambda t}$, que resolviendo para t,
$$t_{\text{med}} = t_{0.5} = \frac{1}{\lambda}\ln 2$$

III). Si $\lambda = 2$,
$$\mu_T = \mathbb{E}[T] = \frac{1}{2} = 0.5; \qquad t_{0.50} = \frac{1}{2}\ln 2 = 0.3466$$

Ejemplo : Considere la *v. a.* X con *f. d. p.* dada por $f_X(x) = \dfrac{1}{\pi(1+x^2)}\mathbb{1}_{\mathbb{R}}(x)$. X se dice que sigue la **distribución Cauchy**. Luego,

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

II).
$$\int_{\mathbb{R}} f_X(x)dx = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1}{1+x^2}dx = \frac{1}{\pi}\arctan x\Big|_{-\infty}^{\infty} = \frac{1}{\pi}\left[\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right] = 1$$

III).
$$\lim_{u\to\infty}\int_{-u}^{u}\frac{x}{\pi(1+x^2)}dx = \lim_{u\to\infty}\log(1+x^2)\Big|_{-u}^{u} = ???$$
y se tiene que
$$\int_{\mathbb{R}}|x|f_X(x)dx = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{|x|}{1+x^2}dx = +\infty$$
por lo que la distribución Cauchy no tiene valor esperado.

Teorema del estadístico inconsciente (TEI, LOTUS) Caso continuo. (Rincón (2014)) Sea X una variable aleatoria continua con función de densidad de probabilidad f_X y g, una función real (medible) tal que $g(X)$ es una *v. a.* con valor esperado finito. Entonces,
$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)f_X(x)dx$$

Demostración:

I). Suponga que g es una función no negativa. Entonces,

$$\begin{aligned}
\mathbb{E}[g(X)] &= \int_0^\infty \mathbb{P}(g(X) > y)dy \qquad \text{(Prob. 6:16, Rincón (2014))} \\
&= \int_0^\infty \mathbb{P}(X \in g^{-1}((y,\infty)))dy \\
&= \int_0^\infty \int_{\{u:y<g(u)<\infty\}} f_X(x)dx\ dy \\
&= \int \int_{\{u:y<g(u)<\infty\}} \int_0^{g(x)} f_X(x)dy\ dx \qquad \text{(Fubini)} \\
&= \int_0^\infty g(x) f_X(x)dx
\end{aligned}$$

II). Para el caso general de la función g, se definen la *parte positiva* y *parte negativa* de la función g por

$$g^+(x) = \begin{cases} g(x) & \text{si } g(x) \geq 0 \\ 0 & \text{si } g(x) < 0 \end{cases} \qquad \text{y} \qquad g^-(x) = \begin{cases} 0 & \text{si } g(x) \geq 0 \\ g(x) & \text{si } g(x) < 0 \end{cases}$$

respectivamente. Note que ambas funciones son no negativas. Luego,

$$g(x) = g^+(x) - \left(g^-(x)\right), \quad x \in \mathbb{R}$$

Así,

$$\begin{aligned}
\mathbb{E}[g(X)] &= \mathbb{E}[g^+(X) - \left(g^-(X)\right)] \\
&= \mathbb{E}[g^+(X)] - \mathbb{E}[g^-(X)] \\
&= \int_{\mathbb{R}} g^+(x) f_X(x)dx - \int_{\mathbb{R}} -g^-(x) f_X(x)dx \\
&= \int_{\mathbb{R}} \left[g^+(x) - (-g^-(x))\right] f_X(x)dx \\
&= \int_{\mathbb{R}} g(x) f_X(x)dx
\end{aligned}$$

Ejemplo : Sea U *v. a.* distribuida uniformemente en $(0,1)$. Entonces, $f_U(u) = \mathbb{1}_{(0,1)}(x)$ y $F_U(u) = u$, si $0 \leq u \leq 1$.

Sea $Y = U^2$. Luego, si $0 \leq y \leq 1$, la *f. p. a.* de Y es

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(U^2 \leq y) = \mathbb{P}(U \leq \sqrt{y}) = \sqrt{y}$$

y la correspondiente *f. d. p.*

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{2} y^{-1/2} \mathbb{1}_{(0,1)}(y)$$

Por lo tanto,

$$\mathbb{E}[Y] = \int_{\mathbb{R}} y f_Y(y)dy = \frac{1}{2} \int_0^1 y^{1/2} dy = \frac{1}{3}$$

Alternativamente, se sigue del TEI

$$\mathbb{E}[U^2] = \int_{\mathbb{R}} u^2 f_U(u)du = \int_0^1 u^2 du = \frac{1}{3}$$

Propiedades de valor esperado

Proposición : Sea X una variable aleatoria con valor esperado finito.

a). Si $a \in \mathbb{R}$, y $\mathbb{P}(X = a) = 1$, entonces $\mathbb{E}[X] = a$.

b). Sea $b \in \mathbb{R}$, entonces $\mathbb{E}[bX] = b\mathbb{E}[X]$.

c). $a, b \in \mathbb{R}$, entonces $\mathbb{E}[a + bX] = a + b\mathbb{E}[X]$.

d). $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.

Demostración: Sin pérdida de generalidad suponga que X es una *v. a.* continua con f, *f. d. p.*.

a). $E[X] = a\mathbb{P}(X = a) + \xi\mathbb{P}(X \neq a) = a$.

b). $\mathbb{E}[bX] \overset{\text{TEI}}{=} \int_{\mathbb{R}}(bx)f(x)dx = b\int_{\mathbb{R}} xf(x) = b\mathbb{E}[X]$.

c). $\mathbb{E}[a + bX] \overset{\text{TEI}}{=} \int_{\mathbb{R}}(a + bx)f(x)dx = a\int_{\mathbb{R}} f(x)dx + b\int_{\mathbb{R}} xf(x)dx = a + b\mathbb{E}[X]$.

d). Para todo $X \in \mathbb{R}$, $-|x| \leq x \leq |x|$. Entonces,

$$\int_{\mathbb{R}} -|x|f(x)dx \leq \int_{\mathbb{R}} xf(x)dx \leq \int_{\mathbb{R}} |x|f(x)dx$$

Luego, $\mathbb{E}[-|X|] \leq \mathbb{E}[X] \leq \mathbb{E}[|X|]$. Por lo tanto $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.

Proposición : Sea X una variable aleatoria con $\mathbb{P}(|X| \leq M) = 1$. Entonces se tiene $\mathbb{E}[|X|] \leq M$.

Proposición : Sea X una variable aleatoria y g_1, g_2 funciones medibles tal que $g_1(x) \leq g_2(x)$, para todo $x \in \mathbb{R}$. Entonces, si existen,

$$\mathbb{E}[g_1(X)] \leq \mathbb{E}[g_2(X)]$$

Proposición : Sean X y Y dos variables aleatorias con valor esperado finito. Entonces,

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

Demostración: Esta proposición enunció antes y se demuestra en el curso siguiente, Cálculo de Probabilidades II.

Nota: El valor esperado $\mathbb{E}$ puede verse como un operador actuando sobre el conjunto de variables aleatorias $\mathcal{F}$ (funciones medibles) con valor esperado, $\begin{array}{rl} \mathbb{E}: & \mathcal{F} \longrightarrow \mathbb{R} \\ & X \longrightarrow \mu_X \end{array}$.

En este sentido, $\mathbb{E}$ es un operador lineal. A saber,

i) $\mathbb{E}[a + bX] = a + b\mathbb{E}[X]$

ii) $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

5.3. Momentos de una variable aleatoria

Definición : Sea X una variable aleatoria. Se define el r**-ésimo momento de** X por

$$\mu_r = \mathbb{E}\left[X^r\right], \quad r \in \mathbb{N}$$

siempre que $\mathbb{E}\left[|X|^r|\right] < \infty$.

$$\mathbb{E}\left[X^r\right] = \begin{cases} \sum_{x_i} x_i^r f_X(x_i) & \text{caso discreto} \\ \int_{\mathbb{R}} x^r f_X(x)dx & \text{caso continuo} \end{cases}$$

Definición : Sea X una variable aleatoria con media (valor esperado) μ_X. Se define el r**-ésimo momento central de** X por

$$\nu_r = \mathbb{E}\left[(X - \mu_X)^r\right], \quad r \in \mathbb{N}$$

siempre que $\mathbb{E}\left[|X|^r\right] < \infty$.

Proposición : Sea X una variable aleatoria con su r-ésimo momento finito. Entonces X tiene su k-ésimo momento finito, $k = 1, 2 \ldots, r$, pero no necesariamente existe su $(r+1)$ momento.

Demostración:

i) Suponga X v. a. continua con r-ésimo momento finito. y $k \leq r$.

$$\begin{aligned}
\mathbb{E}\left[|X|^k\right] &= \int_{\mathbb{R}} |x|^k f(x)dx \\
&= \int_{-1}^{1} |x|^k f(x)dx + \int_{\mathbb{R}\setminus[-1,1]} |x|^k fx(d)dx \\
&\leq \int_{-1}^{1} 1 \cdot f(x)dx + \int_{\mathbb{R}\setminus[-1,1]} |x|^k f(x)dx \\
&\leq 1 + \int_{\mathbb{R}\setminus[-1,1]} |x|^r f(x)dx \\
&\leq 1 + \mathbb{E}\left[|X|^r\right] \\
&< \infty
\end{aligned}$$

ii) Considere ahora la v. a. X con f. d. p. $f(x) = \frac{3}{x^4}\mathbb{1}_{(1,\infty)}(x)$.

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

b) $\int_{\mathbb{R}} f(x)dx = \int_1^\infty \frac{3}{x^4}dx = 1$

c) $\mathbb{E}\left[X^2\right] = \int_{\mathbb{R}} x^2 f(x)dx = \int_1^\infty x^2 \frac{2}{x^4}dx = 3$.

d) $\mathbb{E}\left[X^3\right] = \int_{\mathbb{R}} x^3 f(x)dx = \int_1^\infty 3\frac{3}{x^4}dx = 3\lim_{b\to\infty} \log b = \infty$.

Luego, X tiene segundo momento finito pero no tercero.

Proposición : Sea X una variable aleatoria con r-ésimo momento finito, entonces X tiene r-ésimo momento central finito y viceversa.

Demostración: Sea μ la media de la *v. a.* X. Entonces,

$$\mathbb{E}\left((X-\mu)^r\right) = \mathbb{E}\left[\sum_{k=0}^{r}\binom{r}{k}(-1)^k\mu^{r-k}X^k\right] = \sum_{k=0}^{r}\binom{r}{k}(-1)^k\mu^{r-k}\mathbb{E}[X^k] < \infty$$

por la proposición anterior. Por otro lado, si $\mathbb{E}\left[(X-\mu)^r\right] < \infty$, entonces todos sus sumandos son también finitos.

Nota: Se ha visto que el primer momento (media) de la *v. a.* X es una medida de la localización de la distribución. El segundo, tercer y cuarto momento centrales son medidas de la variabilidad, sesgo o asimetría y lo "plana o picuda" de la distribución respectivamente.

Definición : Sea X una variable aleatoria con segundo momento finito. Se define la **varianza** (de la distribución) de X por

$$\mathrm{var}(X) = \mathbb{E}\left[(X-\mu_X)^2\right] \equiv \sigma_X^2$$

donde $\mu_X = \mathbb{E}[X]$ representa la media de X.

Nota: La varianza σ_X^2 de X es una medida de la variabilidad o dispersión de X, expresada en unidades de X al cuadrado, lo que hace difícil su interpretación. Por ejemplo, si X representa el consumo mensual de agua con media $\mu_X = 9\mathrm{m}^3$ y varianza $\sigma_X^2 = 5\mathrm{m}^6$.

Definición : Sea X una variable aleatoria con varianza finito. Se define la **desviación estándar** (de la distribución) de X por

$$\mathrm{de}(X) = \sqrt{\mathrm{var}(X)} \equiv \sigma_X$$

La desviación estándar σ_X es también una medida de la variabilidad de X esperada en las mismas unidades lo que facilita su interpretación. Por ejemplo, el consumo mensual de agua con valor medio $\mu_X = 9\mathrm{m}^3$ y desviación estándar $\sigma_X = 2.24\mathrm{m}^3$.

Proposición : Sea X una variable aleatoria finita. Entonces,

$$\mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X]$$

Demostración: Sea X la *v. a.* con media μ y varianza σ^2. Luego,

$$\begin{aligned}\sigma^2 &= \mathbb{E}\left[(X-\mu)^2\right]\\ &= \mathbb{E}\left[X^2 - 2\mu X + \mu^2\right]\\ &= \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2\\ &= \mathbb{E}[X^2] - \mu^2\end{aligned}$$

Proposición : Sea X una variable aleatoria con varianza finita σ^2. Sean $a, b \in \mathbb{R}$. Entonces,

a). $\mathrm{var}(a) = 0$.

b). $\mathrm{var}(bX) = b^2\mathrm{var}(X)$

c). $\mathrm{var}(a+bX) = b^2\mathrm{var}(X)$.

Demostración:

a). $\mathbb{E}[a] = a$ y $\mathbb{E}[(a - \mathbb{E}[a])^2] = 0$.

b).

$$\operatorname{var}(bX) \overset{\text{TEI}}{=} \int_{\mathbb{R}} (bx - \mathbb{E}[bX])^2 f(x)dx = \int_{\mathbb{R}} b^2(x-\mu)^2 f(x)dx = b^2\operatorname{var}(X)$$

c).

$$\operatorname{var}(a + bX) = \int_{\mathbb{R}} \Big[(\cancel{a} + bX) - \underbrace{\mathbb{E}[a + bX]}_{\cancel{a}+b\mu}\Big]^2 f(x)dx = b^2\operatorname{var}(X)$$

Definición : Sea X una variable aleatoria con media μ_X y varianza finitas σ_X^2. Se define el **coeficiente de variación** de X por

$$\operatorname{cv}(X) = \frac{\sigma_X}{\mu_X}$$

Algunos textos lo presentan dividido por $|\mu_X|$, por lo que $\operatorname{cv}(X) > 0$.

El coeficiente de variación es una medida (relativa) adimensional (sin unidades) de la dispersión, por lo que permite comparar variabilidades entre de distintos conceptos. Por ejemplo, si X denota el consumo mensual de agua con $\mu_X = 12\text{m}^3$ y $\sigma_X = 2.1\text{m}^3$; y sea Y que denota el consumo mensual de electricidad con $\mu_Y = 8\text{kWh}$ y $\sigma_Y = 1.9\text{kWh}$. Entonces,

$$\operatorname{cv}(X) = \frac{\sigma_X}{\mu_X} = 0.175, \quad \text{y} \quad \operatorname{cv}(Y) = \frac{\sigma_Y}{\mu_Y} = 0.2375$$

por lo que el consumo mensual de electricidad es "más variable" que el consumo de agua.

Definición : Sea X una variable aleatoria con varianza finita. La **estandarización** de X consiste en *"centrarla"* y *"escalarla"* de manera que el resultado sea una *v. a.* con media cero y varianza 1.

A saber, si X tiene media μ_X y varianza σ_X^2 y si Y es la estandarización de X, entonces $Y = \dfrac{X - \mu_X}{\sigma_X}$. Luego,

i) $\mathbb{E}[Y] = \mathbb{E}\left[\frac{X-\mu_X}{\sigma_X}\right] = \frac{1}{\sigma_X^2}\mathbb{E}]X - \mu_X] = 0$.

1. $\operatorname{var}(Y) = \operatorname{var}\left(\frac{X-\mu_X}{\sigma_X^2}\right) = \frac{1}{\sigma_X}\operatorname{var}(X - \mu_X) = \frac{1}{\sigma_X^2}\operatorname{var}(X) = 1$.

iii) Note que Y, la *"normalización"* de X, no tiene unidades.

Definición : Sea X una variable aleatoria con varianza finita. Se define la **relación señal–ruido** (SNR en inglés) de X por

$$\text{SNR}(X) = \frac{|\mu_X|}{\sigma_X} = \frac{1}{\operatorname{cv}(X)}$$

Definición : Sea X una variable aleatoria con tercer momento finito. Se define el **sesgo o asimetría de** X por su tercer momento central

$$\nu_3 = \mathbb{E}\left[(X - \mu_X)^3\right]$$

Nota: Si la distribución es unimodal:

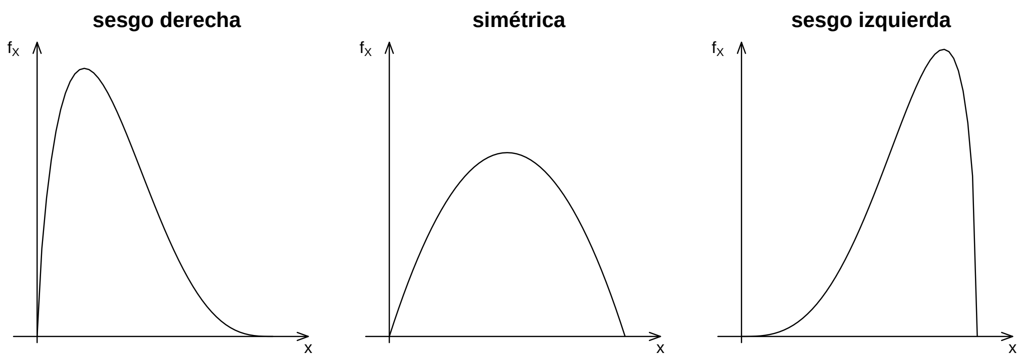


Figura 12: Densidades simétrica y sesgadas.

i) $\nu_e \approx 0$, la distribución es **simétrica**.

ii) $\nu_e > 0$, la distribución con **sesgo** (cola larga) **a la derecha**.

iii) $\nu_e < 0$, la distribución con **sesgo** (cola larga) **a la izquierda**.

La figura 12 ilustra los tres casos anteriores.

Definición : Sea X una variable aleatoria con tercer momento finito. Se define el **coeficiente de asimetría o sesgo (de Fisher-Pearson)** por

$$\eta_3 = \mathbb{E}\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^3\right]$$

Ejercicio : Verifique que $\eta_3 = \dfrac{1}{\sigma_X^3}\left[\mathbb{E}[X^3] - 3\mu_X\sigma_X^2 - \mu_X^3\right]$.

Definición : Sea X una variable aleatoria con cuarto momento finito. Se define la **curtosis de (la distribución)** X por

$$\nu_4 = \mathbb{E}\left[(X-\mu_X)^4\right]$$

La curtosis es una medida de lo *" ligero o pesado de las colas"* de la función de densidad de X. Se define el **coeficiente de curtosis de** X por

$$\eta_4 = \mathbb{E}\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^4\right]$$

Luego, el **exceso de curtosis** es $\varepsilon = \eta_4 - 3$.

Nota: En estadística se acostumbraba, en ocasiones, caracterizar las distribuciones de acuerdo a la curtosis.

i) Si $\varepsilon \approx 0$ la densidad se dice **mesocúrtica**.

ii) Si $\varepsilon > 0$ la densidad se dice **leptocúrtica** (colas ligeras).

iii) Si $\varepsilon < 0$ la densidad se dice **platicúrtica** (colas pesadas).

La figura 13 ilustra los 3 casos de densidades recién descritas.

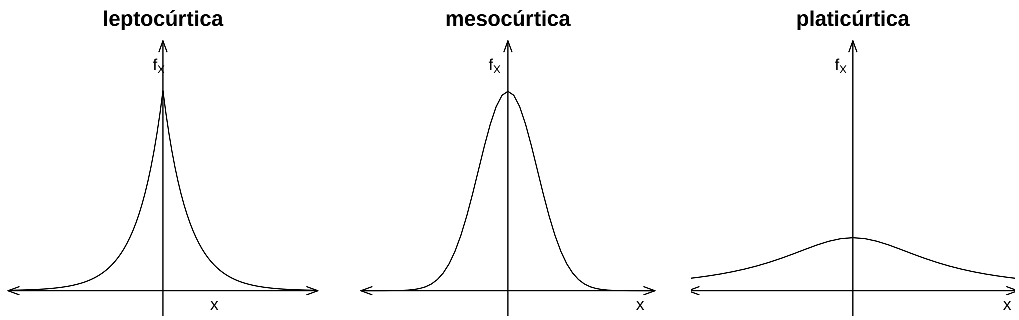


Figura 13: Densidades para los distintos casos de curtosis.

5.4. Esperanza Condicional

Ejemplo : Sea X un variable aleatoria distribuida uniformemente en $S_X = \{1, 2, \ldots, 19, 20\}$. Se extrae un número al azar. El número esperado es

$$\mathbb{E}[X] = \frac{1}{20}\sum_{x=1}^{20} 1 = \frac{1}{20}\frac{20(21)}{20} = 10.5$$

Suponga ahora que sabe que el número extraído es primo. ¿Modificaría el valor esperado?

Ejemplo : Suponga que un teléfono celular tiene una vida media útil de 4 años y una garantía de 6 meses. ¿Cual diría que es es el tiempo de vida media de un teléfono celular que falló durante el periodo de garantía?

Sea X una variable aleatoria discreta con soporte $S_X = \{x_1, x_2, \ldots\}$ y valor esperado finito, $B \in \mathcal{B}(\mathbb{R})$, tal que

$$\mathbb{P}(B) = \sum_{x_i \in B} f_X(x_i) = \sum_{x_i \in B} \mathbb{P}(X = x_i) > 0$$

Se vió ya que para $A \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}(X \in A | X \in B) = \frac{\mathbb{P}(X \in A \cap B)}{\mathbb{P}(X \in B)} = \frac{1}{\mathbb{P}(X \in B)} \sum_{x_i \in A} f_X(x_i)\mathbb{1}_B(x_i)$$

Así, en general,

$$\mathbb{P}(X \in A | B) = \sum_{x_i \in A} f(x_i | B)$$

donde $f(x|B) = \dfrac{1}{\mathbb{P}(X \in B)} f_X(x)\mathbb{1}_B(x)$.

Definición : Sea X una variable aleatoria con función masa de probabilidad f_X y $B \in \mathcal{B}(\mathbb{R}))$, tal que $\mathbb{P}(X \in B) > 0$. Se define la **función masa de probabilidad condicional de X dado $X \in B$**, por

$$f(x|B) = \frac{1}{\mathbb{P}(X \in B)} f_X(x)\mathbb{1}_B(x), \qquad \text{para todo } x \in \mathbb{R}$$

Nota: $f(x|B)$ es una *f. m. p.* propia. A saber, *i)* $f(x|B) = \frac{1}{\mathbb{P}(X\in B)} f_X(x)\mathbb{1}_B(x) \geq 0, \quad$ para todo $x \in \mathbb{R}$.

ii) $\sum_{x_i\in S_X} f(x_i|B) = \sum_{x_i\in S_X} \frac{1}{\mathbb{P}(X\in B)} f_X(x)\mathbb{1}_B(x) = \frac{1}{\mathbb{P}(X\in B)} \sum_{x_i\in B} f_X(x) = 1$

A partir de la *f. m. p.* condicional se define la **esperanza condicional de X dado** $X \in B$.

$$\mathbb{E}[X|B] = \sum_{x_i\in S_X} x_i f(x_i|B) = \frac{1}{\mathbb{P}(X\in B)} \sum_{x_i\in B} x_i f_X(x_1)$$

Ejemplo : El valor de X que se distribuye uniformemente $S_X = \{1, 2, \ldots, 19, 20\}$ si se sabe que salió un número primo.

Solución: $B = \{$Número primo menor que $20\} = \{2, 3, 5, 7, 11, 13, 17, 19\}$, $n(B) = 8$, y como X se distribuye uniformemente en S_X. $\mathbb{P}(B) = n(B)/n(S_X) = 8/20$.

$$\begin{aligned}
\mathbb{E}[X|B] &= \sum_{x_i\in S_X} x_i f(x_i|B) \\
&= \frac{1}{\mathbb{P}(X\in B)} \sum_{x_i\in B} x_i f_X(x_i) \\
&= \frac{1}{8/20} \sum_{x_i\in B} x_i \frac{1}{20} \\
&= \frac{20}{8}\frac{1}{20}(2 + 3 + \cdots + 19) \\
&= \frac{77}{8} \\
&= 9.625
\end{aligned}$$

Mientras que

$$\mathbb{E}[X] = \sum_{x_i\in S_X} x_i f_X(x_i) = \sum_{x=1}^{20} x\frac{1}{20} = \frac{1}{20}\frac{20(21)}{2} = \frac{21}{2} = 10.5$$

Suponga ahora que Y es una variable aleatoria continua con *f. d. p.* f_Y, $C \in \mathcal{B}(\mathbb{R})$, tal que $\mathbb{P}(Y \in C) > 0$. Si $A \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}
\mathbb{P}(Y\in A|Y\in C) &= \frac{\mathbb{P}(Y\in A\cap C)}{\mathbb{P}(Y\in C)} \\
&= \frac{1}{\mathbb{P}(Y\in C)} \int_{A\cap C} f_Y(y)dy \\
&= \frac{1}{Y\in C} \int_A f_Y(y)\mathbb{1}_C(y)dy \\
&= \int_A f(y|C)dy
\end{aligned}$$

donde $f(y|C) = \frac{1}{\mathbb{P}(Y\in C)} f_Y(y)\mathbb{1}_C(y)$, para todo $y \in \mathbb{R}$.

Nota: $f(y|C)$ es una *f. d. p.* propia. En efecto,

i) $f(y|C) = \frac{1}{\mathbb{P}(Y\in C)} f_Y(y)\mathbb{1}_C(y) \geq 0, \quad$ para todo $y \in \mathbb{R}$.

ii) $\int_{\mathbb{R}} f(y|C)dy = \frac{1}{\mathbb{P}(Y\in C)} \int_{\mathbb{R}} f_Y(y)\mathbb{1}_C(y)dy = \frac{1}{\mathbb{P}(Y\in C)} \int_C f_Y(y)dy = 1$

Luego,

$$\begin{aligned} E[Y|C] &= \int_{\mathbb{R}} y f(y|C) dy \\ &= \int_{\mathbb{R}} y \frac{1}{\mathbb{P}(Y \in C)} f_Y(y) \mathbb{1}_C(y) dy \\ &= \frac{1}{\mathbb{P}(Y \in C)} \int_C y f_Y(y) dy \end{aligned}$$

define la **esperanza condicional de Y dado $Y \in C$.**

Ejemplo : Sea T la vida útil de un teléfono celular. Suponga que $T \sim \text{Exp}(\lambda)$, con $\mathbb{E}[T] = 4$ años. Luego, $f_T(t) = \frac{1}{4} e^{-t/4} \mathbb{1}_{(0,\infty)}(t)$ y $E[T] = 1/\lambda = 4$.

Si la garantía es de 6 meses, $\mathbb{P}(T \leq 0.5) = 1 - e^{-\frac{1}{4}0.5} = 0.1175$. Esto es, la probabilidad de que el celular falle durante el periodo de garantía es 0.1175.

$$\begin{aligned} \mathbb{E}[T|T \leq .5] &= \int_{\mathbb{R}} t f(t|C) dt \\ &= \int_{\mathbb{R}} t \frac{1}{\mathbb{P}(C)} f_T(t) \mathbb{1}_C(t) dt \\ &= \frac{1}{\mathbb{P}(T \leq 0.5)} \int_0^{0.5} \frac{1}{4} t e^{-t/4} dt \\ &= \frac{0.0288}{0.1175} \\ &= 0.2448 \approx 2.93 \text{ meses} \end{aligned}$$

a diferencia de los 4 años del caso incondicional.

Proposición : El correspondiente *teorema de estadístico inconsciente*: sea g una función real tal que $g(Y)$ es una variable aleatoria con valor esperado finito,

$$\mathbb{E}[g(Y)|C] = \frac{1}{\mathbb{P}(Y \in C)} \int_C g(y) f_Y(y) dy$$

Luego, se puede definir la **varianza condicional de Y dado $Y \in C$**

$$\text{var}(Y|C) = \mathbb{E}[Y^2|C] - \mathbb{E}^2[Y|C]$$

5.5. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 5, Barrios and Heiras (2024).

6. Función Generadora de Momentos

Definición : Se dice que g es la **función generadora** de la sucesión $\{a_k\}$ si

$$g(t) = \sum_{k=1}^{\infty} a_k t^k, \quad \text{para todo } |t| < T$$

Ejemplo :

a). $g(t) = \dfrac{1}{1-t} = \displaystyle\sum_{k=0}^{\infty} t^k$ es la función generadora de la sucesión $1, 1, 1, \ldots$.

b). La sucesión $1, b, b^2, b^3, \ldots$ tiene como función generadora
$g(t) = \dfrac{1}{1-bt} = \displaystyle\sum_{k=0}^{\infty} (bt)^2$.

c). La función generadora de la sucesión $0, 1, 0, 1, \ldots$, es $g(t) = \dfrac{1}{1-t^2}$ pues $\frac{1}{1-t^2} = \sum_{k=1}^{\infty} t^{2k}$

d). La función $g(t) = e^t = \sum_{k=1}^{0} t^k/k!$ es la generadora de la sucesión $\left\{\dfrac{1}{k!}\right\}$.

6.1. Función generadora de momentos

Definición : Sea X una variable aleatoria. Se define su **función generadora de momentos (*f. g. m.*)** por

$$m_X(t) = \mathbb{E}\left[e^{tX}\right], \quad \text{para todo } \ t \in \mathbb{R}$$

siempre que exista el valor esperado.

Ejemplo : Parea X que sigue una **distribución Bernoulli** con probabilidad de exito p, que en estas notas se denota por $X \sim \mathrm{Ber}(p)$. Entonces, su *f. m. p.*

$$f_X(x) = p\mathbb{1}_{\{1\}}(x) + q\mathbb{1}_{\{0\}}(x)$$

con $q = 1 - p$ y su *f. g. m.*,

$$m_X(t) = \mathbb{E}\left[e^{tX}\right] = e^{t\,0} f_X(0) + e^{t\,1} f_X(1) = q + pe^t$$

para $t \in \mathbb{R}$,

Ejemplo : Sea T una variable aleatoria continua con *f. d. p.* $f_T(t) = \lambda e^{-\lambda t}\mathbb{1}_{(0,\infty)}(t)$, para alguna $\lambda > 0$ fija. En este caso se dice que la *v. a.* T sigue una **distribución exponencial**, $T \sim \mathrm{Exp}(\lambda)$. La correspondiente *f. g. m.* es

$$m_T(u) = \mathbb{E}\left[e^{uT}\right] = \frac{\lambda}{\lambda - u} = (1 - u/\lambda)^{-1}, \quad \text{para } u < \lambda$$

En efecto,

$$\begin{aligned} m_T(u) &= \mathbb{E}\big[e^{uT}\big] \\ &\overset{\text{TEI}}{=} \int_{\mathbb{R}} e^{ut} f_T(t)dt \\ &= \int_0^\infty e^{ut}\lambda e^{-\lambda t}dt \\ &= \frac{\lambda}{\lambda - u}\int_0^\infty (\lambda-u)e^{-(\lambda-u)t}dt \\ &= \frac{\lambda}{\lambda-u} \end{aligned}$$

La última integral es 1 pues es la integral sobre todo el soporte $(0,\infty)$, de una función de densidad propia.

Nota: Hay variables aleatorias que no tienen función generadora de momentos. Tal es el caso de la *distribución Cauchy* que tiene ninguno de sus momentos o la *distribución lognormal* (se verá más adelante), que sí tiene todos su momentos finitos pero no tiene *f. g. m.*.

Teorema : Sea X una variable aleatoria con función generadora de momentos m_X. Entonces m_X determina de manera única la distribución de X.

Proposición : Saa X una variable aleatoria con función generadora de momentos m_X, infinitamente diferenciable. Entonces,

$$\mathbb{E}\big[X^r\big] = \frac{d^r}{dt^r}m_X(t)\Big|_{t=0}, \quad r = 1, 2, \ldots$$

Demostración: Si existe $m_X(t)$ en una vecindad del cero $|t| < T$,

$$\begin{aligned} m_X(t) &= \mathbb{E}\big[e^{tX}\big] \\ &= \mathbb{E}\left[\sum_{k=o}^{\infty}\frac{1}{k!}(tX)^k\right] \\ &= \sum_{k=o}^{\infty}\frac{t^k}{k!}\mathbb{E}[X^k] \\ &= 1 + t\mathbb{E}[X] + \frac{t^2}{2!}\mathbb{E}[X^2] + \frac{t^3}{3!}\mathbb{E}[X^3] + \cdots \end{aligned}$$

Luego,

$$\begin{aligned} m'_X(t) &= \mathbb{E}[X] + \frac{2t}{2!}\mathbb{E}[X^2] + \frac{3t^2}{3!}\mathbb{E}[X^3] + \cdots \\ m''_X(t) &= \mathbb{E}[X^2] + \frac{2\cdot 3t}{3!}\mathbb{E}[X^3] + \frac{3\cdot 4t}{4!}\mathbb{E}[X^4] + \cdots \\ &\vdots \\ m_X^{(k)}(t) &= \frac{k!}{k!}\mathbb{E}[X^k] + \frac{(k+1)_k}{(k+1)!}t\mathbb{E}[X^{k+1}] + \cdots \end{aligned}$$

Y evaluando en $t = 0$, se tiene la proposición

$$\begin{aligned} m'_X(0) &= \mathbb{E}[X] \\ m'_X(0) &= \mathbb{E}[X'^2] \\ &\vdots \\ m_X^{(r)}(0) &= \mathbb{E}[X^r] \end{aligned}$$

Ejemplo : Sea $X \sim \text{Ber}(p)$. Se ha visto ya que $\mathbb{E}[X] = p$ y $m_X(t) = q + pe^t$. Por lo que $m'_X(t) = pe^t \cdot 1$ y $m_X(0) = pe0 = p$. Así, $\mathbb{E}[X] = p = m'_X(0)$.

Ejemplo : Sea $T \sim \text{Exp}(\lambda)$, esto es, $f_T(t) = \lambda e^{-\lambda T} \mathbb{1}_{(0,\infty)}(t)$. Se ha visto que $\mathbb{E}[T] = 1/\lambda$ y que $m_T(u) = \lambda/(\lambda - u) = (1 - u/\lambda)^{-1}$. Por lo que $m'_T(u) = (1 - u/\lambda)^{-2}(1/\lambda)$ y en consecuencia $m'_T(0) = 1/\lambda = \mathbb{E}[T]$.

Ejercicio : La *v. a.* U sigue la **distribución uniforme** en el intervalo (a, b). Esto es, $U \sim \text{unif}(a, b)$, con *f. d. p.* $f_U(u) = \frac{1}{b-a} \mathbb{1}_{(a,b)}(u)$.

a). Determine el valor esperado de U, $\mathbb{E}[U] = (a + b)/2$.
b). Calcule la correspondiente *f. g. m.*. A saber, $m_U(t) = \dfrac{e^{bt} - e^{at}}{t(b - a)}$.
c). Verifique $\mathbb{E}[U]$, empleando la *f. g. m.* m_U.

Proposición : Sea X una variable aleatoria con función generadora de momentos m_X. Sean $a, b \in \mathbb{R}$ y $Y = a + bX$. Entonces, para todo t,

$$m_Y(t) = e^{at} m_X(bt)$$

Esto es, $m_{a+bX}(t) = e^{at} m_X(bt)$.

Demostración:

$$m_{a+bX}(t) = \mathbb{E}[e^{t(a+bX)}] = \mathbb{E}[e^{at}e^{btX}] = e^{at}\mathbb{E}[e^{(bt)X}] = e^{at}\, m_X(bt)$$

Ejercicio : Sean X e Y como la proposición anterior. Utilizando la *f. g. m.* muestre que:

i) $\mathbb{E}[Y] = a + b\mathbb{E}[X]$
ii) $\text{var}(Y) = b^2 \text{var}(X)$.

Proposición : Problema de los momentos (T. J. Stiltjes) Considere la variable aleatoria X y sus momentos $\mu_r = \mathbb{E}[X^r]$, $r = 1, 2, 3, \ldots$. Dados los momentos $\mu_1, \mu_2, \mu_3, \ldots$, ¿es posible recuperar la distribución de X? Respuesta: NO.

Ejemplo :

a). Sea X *v. a.* con *f. d. p.* dada por

$$f_X(x) = \frac{1}{24} e^{x^{1/4}} (1 - \alpha \sin x^{1/4}) \mathbb{1}_{\mathbb{R}\setminus\{0\}}(x)$$

con $0 < \alpha < 1$, fija. Se tiene entonces que

i) f_X es una función de densidad propia.
ii) El r-ésimo momento $\mu_r = \mathbb{E}[X^r]$ no depende de α, por lo que no hay manera de recuperar completamente f_X a partir de sus momentos.

b). Sea la *v. a.* X con *f. d. p.*

$$f_X(x) = \frac{1}{x\sqrt{\sigma\pi}} \exp\left\{-\frac{1}{2}(\log x)^2\right\} \mathbb{1}_{(0,\infty)}(x)$$

y la *v. a.* Y, una "perturbación de X" con *f. d. p.*

$$f_Y(y) = f_X(y) \left[1 + \sin(2\pi \log y)\right] \mathbb{1}_{(0,\infty)}(y)$$

En este caso se puede mostrar que $\mathbb{E}[X^r] = \mu_r = \mathbb{E}[Y^r]$, $r = 1, 2 \ldots$. Por lo que no hay forma de saber con certeza a qué distribución se refiere a partir simplemente de los momentos μ_r.

Teorema de unicidad de *f. g. m.* Sean X y Y variables aleatorias con funciones de distribución F_X y F_Y, y funciones generadoras de momentos m_X y m_Y, respectivamente. Si $m_X(t) = m_Y(t)$, para todo $t \in \mathbb{R}$. Entonces, $F_X(w) = F_Y(w)$, para todo $w \in \rfloor(F)$, puntos de continuidad de F.

$$\begin{array}{ccc} X & & Y \\ \downarrow & & \downarrow \\ m_X(t) & \longleftrightarrow & m_Y(t) \\ \Updownarrow & & \Updownarrow \\ F_X(w) & \longleftrightarrow & F_Y(w) \end{array}$$

Ejercicio : Sea $X_1, X_2, \ldots,$ sucesión de ensayos Bernoulli con probabilidad de éxito p. $X_i \sim \text{Ber}(p)$. Verifique las siguientes *f. g. m.*.

a). Sea X la suma de éxitos en n ensayos. X se dice que sigue la **distribución binomial**, $X \sim \text{Bin}(n, p)$.

$$f_X(x) = \binom{n}{x} p^x q^{n-x} \mathbb{1}_{\{0,1,\ldots,n\}}(x)$$
$$m_X(t) = (q + pe^t)^n, \quad t \in \mathbb{R}$$

b). Y denota el número de fracasos hasta el primer éxito. Y sigue la **distribución geomética**, $Y \sim \text{Geom}(p)$.

$$f_Y(y) = pq^y \mathbb{1}_{\{0,1,2,\ldots\}}(y)$$
$$m_Y(t) = p(1 - qe^t)^n, \quad t < -\ln(1-p)$$

c). N *v. a.* entera no negativa que sigue la **distribución Poisson**, $N \sim \text{Po}(\lambda)$.

$$f_N(n) = \frac{\lambda^n e^{\lambda}}{n!} \mathbb{1}_{\{0,1,2,\ldots\}}(n)$$
$$m_N(t) = e^{\lambda(e^t - 1)}, \quad t \in \mathbb{R}$$

6.2. Función característica

Definición : Se define la **función característica** (*f. c.*) de la *v. a.* X por

$$\varphi_X(t) = \mathbb{E}[e^{itX}]$$

para todo t y con $i = \sqrt{-1}$.

Corolario : Sea X *v. a.* con función generadora de momentos m_X y función característica φ_X. Entonces,

$$\varphi_X(t) = m_X(it)$$

para todo t donde m_X exista.

Proposición : Sea X *v. a.* continua con función de densidad de probabilidad f_X. Entonces, su función característica φ_X siempre existe.

Demostración Recuerde que $e^{ix} = \cos x + i \sin x$, por lo que $|e^{ix}|^2 = \cos^2 x + \sin^2 x = 1$. Luego, se tiene para todo t,

$$|\varphi_X(t)| = \left| \int_{\mathbb{R}} e^{itx} f_X(x) dx \right| \leq \int_{\mathbb{R}} |e^{itx}| f_X(x) dx = \int_{\mathbb{R}} 1 \cdot f_X(x) dx = 1$$

Así, la función característica *siempre existe*.

La proposición anterior se presentó el caso de X *v. a.* continua, pero es el mismo resultado si X es una *v. a.* discreta. Sustituya la integral por suma.

Recuerde que toda variable aleatoria X tiene su función de probabilidad acumulada o de distribución F_X que tiene asociada una función (densidad, masa o mixta) de probabilidad f_X. La definición de función característica y la proposición anterior muestran que la distribución de X también tiene asociada φ_X.

Fórmula de Inversión [(Harris 1966), Teorema 4-5.4] Sea h cualquier número positivo y sean x y $x+h$ un par de puntos de continuidad de F_X. Entonces,

$$F_X(x+h) - F_X(x) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{1-e^{-ith}}{it} e^{-itx} \varphi_X(t)dt$$

Se sigue del resultado anterior que toda función característica tiene asociada una función de distribución por lo que φ_X es una manera de *caracterizar* la distribución de X y de ahí su nombre.

Ejemplo :

- $X \sim \text{Bin}(n,p)$. Entonces, $\varphi_X(t) = m_X(it) = (q + pe^{it})^n$.
- $N \sim \text{Po}(\lambda)$. Entonces, $\varphi_N(t) = \exp\{\lambda(e^{it} - 1)\}$.
- $Y \sim \text{Exp}(\lambda)$. Entonces, $\varphi_Y(t) = (1 - it/\lambda)^{-1}$.
- $W \sim$ Cauchy. Entonces, $\varphi_W(t) = e^{-|t|}$.

Se sabe que la distribución Cauchy no tiene valor esperado. Por otro lado, la proposición anterior indica que todas las funciones de probabilidad tiene su correspondiente función característica y el ejemplo muestra que la distribución Cauchy tiene como *f. c.* $\varphi(t) = e^{-|t|}$. En este caso no es posible usar el resultado $\varphi(t) = m_X(it)$ pues la *f. g. m.* no existe y si existiese note que la función φ no es diferenciable en 0.

Proposición : Sea X *v. a.* con *f. p. a.* $F_X(x)$ y *f. c.* $\varphi_X(t)$. Entonces, $\varphi_X(t)$ es real si y solo si F_X es simétrica, en el sentido que $F_X(-x) = 1 - F_X(x)$, para todo x.

Demostración: Vea Apéndice 3 de (Harris 1966).

Note que efectivamente la *f. d. p.* de la distribución Cauchy(0,1) es simétrica al rededor de 0 y su *f. c.* es una función real. Lo mismo sucede con la distribución normal centrada en 0, su correspondiente *f. c.* es $\varphi(t) = e^{-\sigma^2 t^2/2}$, función real.

6.3. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 6, Barrios and Heiras (2024).

7. Desigualdades

Desigualdad de Markov: Sea X una variable aleatoria positiva con media finita, $\mu_X < \infty$. Entonces, para todo $c > 0$, se tiene

$$\mathbb{P}\left(X \geq c\right) \leq \frac{\mathbb{E}[X]}{c} = \frac{\mu_X}{c}$$

Demostración: Sin pérdida de generalidad suponga que X es una *v. a.* continua positiva con *f. d. p.* f. Sea $\mathbb{E}[X] = \mu_X$, entonces,

$$\begin{aligned}\mu_X = \mathbb{E}[X] &= \int_0^\infty x f(x) dx \\ &\geq \int_c^\infty x f(x) dx \\ &\geq c \int_c^\infty f(x) dx \\ &= c\mathbb{P}\left(X \geq c\right)\end{aligned}$$

y se sigue la desigualdad.

7.1. Desigualdad de Chebyshev

Desigualdad de Chebyshev: Sea X una variable aleatoria con media μ_X y varianza σ_X^2 finita. Entonces

$$\mathbb{P}\left(|X - \mu_X| \geq \epsilon\right) \leq \frac{\sigma_X^2}{\epsilon^2}$$

Demostración: Sean $\mu_X = \mathbb{E}[X]$ y $\sigma_X^2 = \text{var}(X)$. Defina $Y = (X - \mu_X)^2$. Entonces, Y es una *v. a.* positiva y por la desigualdad anterior

$$\begin{aligned}\mathbb{P}\left(Y \geq \epsilon^2\right) &\leq \frac{\mathbb{E}[Y]}{\epsilon^2} \\ \mathbb{P}\left((X - \mu_X)^2 \geq \epsilon^2\right) &\leq \frac{\mathbb{E}[(X-\mu)^2]}{\epsilon^2} \\ \mathbb{P}\left(|X - \mu_X| \geq \epsilon\right) &\leq \frac{\sigma_X^2}{\epsilon^2}\end{aligned}$$

Algunos textos presentan alternativamente la desigualdad de Chebyshev en términos de la desviación estándar. A saber, para todo $k > 0$,

$$\mathbb{P}\left(|X - \mu_X| < k\sigma_X\right) \geq 1 - \frac{1}{k^2}$$

Ejercicio : Verifique el resultado anterior.

En ocasiones la desigualdad de Chebyshev puede ser muy conservadora pero en otros en que *no puede ser mejorada*, en el sentido que la cota se alcanza,

Ejemplo : Considere la *v. a.* X con la siguiente *f. m. p.*

x	-1	0	$+1$
p	$1/8$	$6/8$	$1/8$

En este caso se tiene $\mathbb{E}[X] = 0$, $\text{var}(X) = \mathbb{E}[X^2] = 2/8$ y

$$\begin{aligned}\mathbb{P}\left(|X - \mu_X| \geq 2 \cdot \frac{1}{2}\right) &= \mathbb{P}\left(|X| \geq 1\right) = \frac{2}{8} \\ &\leq \frac{\sigma_X^2}{\epsilon^2} = \frac{1/4}{1^2} = \frac{1}{4}\end{aligned}$$

En el caso anterior se alcanza la igualdad por lo que la Desigualdad de Chebyshev no puede ser mejorada. Sin embargo, en ocasiones la desigualdad puede resultar demasiado conservadora al indicar, por ejemplo, que $\mathbb{P}\big(|x-\mu_X|\geq\epsilon\big)\leq 1$.

Deigualdad de Chebyshev de un solo lado: Sea X una *v. a.* con media cero y varianza finita. Entonces, para todo $a>0$,

$$\mathbb{P}\left(X\geq a\right)\leq\frac{\sigma_X^2}{a^2+\sigma_X^2}$$

Demostración: Vea Ross (2006).

Ejemplo : El número de clientes por día es una caja tiene una media de 20 clientes y una desviación estándar de 2 clientes. ¿Qué diría de la probabilidad de tener mañana entre 17 y 23 clientes?

Solución: $\mu_X=\mathbb{E}[X]=20$ y $\sigma_X^2=\mathrm{var}(X)=2^2=4$. Se sigue de la desigualdad de Chebyshev que

$$\begin{aligned}\mathbb{P}\left(17\leq X\leq 23\right)&=\mathbb{P}\left(16<X<24\right)\\&=\mathbb{P}\left(-4<X-\mu<4\right)\\&=\mathbb{P}\big(|X-\mu|<4\big)\\&=\mathbb{P}\big(|X-\mu|<2(2)\big)\\&\geq 1-\frac{1}{2^2}=\frac{3}{4}\end{aligned}$$

Note que si $\sigma=1$, entonces $k=4$ para cubrir el intervalo $[17,23]$ y en tal caso

$$\mathbb{P}\left(17\leq X\leq 23\right)=1-\frac{1}{16}=\frac{15}{16}$$

Esto es, si se reduce la dispersión (varianza) aumenta la probabilidad del intervalo $[17,23]$.

Ejemplo : Se sabe por experiencia que el tiempo medio para reparar una máquina es de 6.2 h y una desviación estándar de 3.52 h. Suponga que un empleado nuevo se lleva 22.5 h en reparar una máquina. ¿Considera usted que necesita de capacitación o entrenamiento?

Solución: Sea Y el tiempo (aleatorio) de reparación de una máquina. Si se aplica la Desigualdad de Chebyshev de un lado se tiene

$$\mathbb{P}\left(Y\geq 225\right)=\mathbb{P}\big((Y-6.2)\geq 16.3\big)\leq\frac{3.52^2}{16.3^2+3.52^2}=0.044$$

La probabilidad es relativamente baja, por lo que se sugiere entrenamiento.

7.2. Desigualdad de Jensen

Definición : Una función h continua en $\mathbb{R}$ se dice **convexa** si para todo $x_0\in\mathbb{R}$, existe una línea recta $\ell(x)=a+bx$ que pasa por $(x_0,h(x_0))$ y queda por debajo de $h(x)$. Esto es, $\ell(x)\leq h(x)$, para todo $x\in\mathbb{R}$.

Desigualdad de Jensen. Sea X una *v. a.* continua con esperanza finita y h una función convexa tal que $h(\mathbb{E}[X])$ existe. Entonces, se tiene que

$$\mathbb{E}\left[h(X)\right]\geq h\left(\mathbb{E}[X]\right)$$

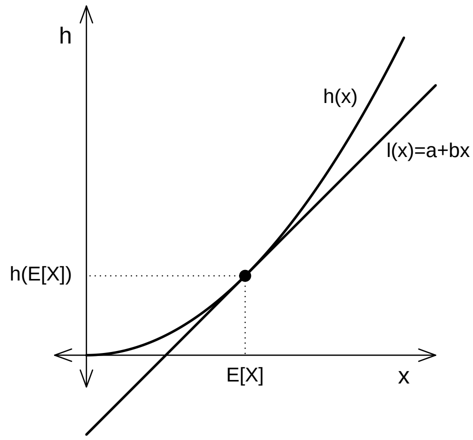


Figura 14: Función convexa h con $(\mathbb{E}[X], h(\mathbb{E}[X]))$, punto de tangencia de h con la recta $\ell(x) = a + bx$.

Demostración: Puesto que h es una función convexa, existe una recta $\ell(x) = a + bx$ que pasa por $\big(\mathbb{E}[X], h(\mathbb{E}[X])\big)$ y satisface que $\ell(x) = a + bx \leq h(x)$. (La recta ℓ es la de la definición de arriba de función convexa.) Se tiene que $h(X) \geq \ell(X)$ y tomando valor esperado de ambos lados

$$\mathbb{E}[h(X)] \geq \mathbb{E}[\ell(X)] = \mathbb{E}[a + bX] = a + b\mathbb{E}[X] = \ell(\mathbb{E}[X]) = h(\mathbb{E}[X])$$

y se tiene la desigualdad.

Ejemplo : Sea $X \sim \text{unif}(0, 1)$ y $h(x) = x^2$. Entonces, $E[X] = 1/2$ y $h(\mathbb{E}[X]) = 1/4$.

Por otro lado, $\mathbb{E}[h(X)] = \mathbb{E}[X^2] = 1/3$. Así se cumple que

$$\mathbb{E}[h(X)] = \frac{1}{3} \geq \frac{1}{4} = h(\mathbb{E}[X])$$

7.3. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 7, Barrios and Heiras (2024).

8. Distribuciones Discretas Paramétricas

8.1. Distribución Uniforme

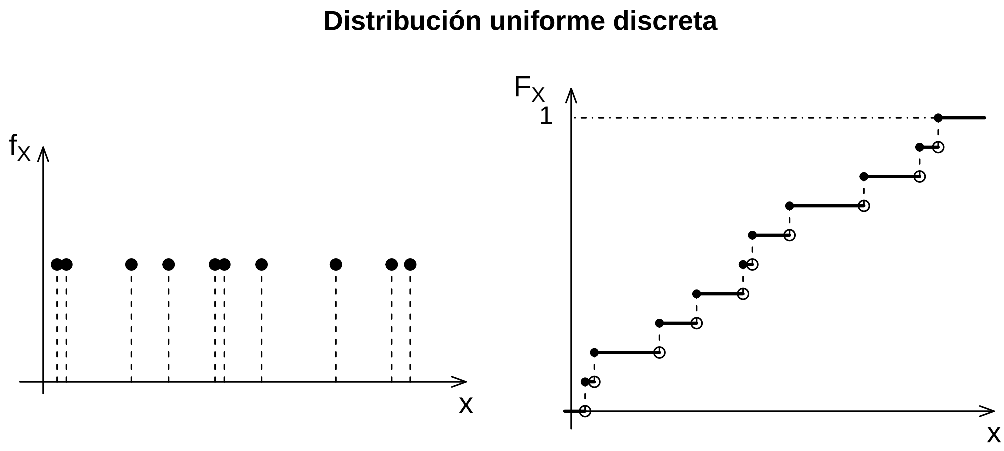


Figura 15: **Distribución uniforme**: Función masa de probabilidad f_X y correspondiente función de probabilidad acumulada F_X.

Sea U la variable aleatoria que sigue la **distribución uniforme (discreta)** en los puntos $u_1, u_2, \ldots, u_n \in \mathbb{R}$. Se denota $U \sim \text{unif}\{u_1, \ldots, u_n\}$. La figura 15 muestra el caso de una *f. m. p.* y la correspondiente función de distribución.

Parámetros: $u_1, u_2, \ldots, u_n \in \mathbb{R}$.

Soporte: $S_U = \{u_1, \ldots, u_n\}$.

Función masa de probabilidad:

$$f_U(u) = \frac{1}{n}\mathbb{1}_{S_U}(u)$$

Media:

$$\mathbb{E}[U] = \sum_{u_i \in S_U} u_i f_U(u_i) = \frac{1}{n}\sum_{i=1}^{n} u_i = \bar{u}$$

Varianza:

$$\text{var}(U) = \sum_{u_i \in S_U} (u_i - \mathbb{E}[U]))^2 f_U(u_i) = \frac{1}{n}\sum_{i=1}^{n}(u_i - \bar{u})^2$$

Función generadora de momentos:

$$m_U(t) = \sum_{u_i \in S_U} e^{tu_i} f_U(u_i) = \frac{1}{n}\sum_{i=1}^{n} e^{tu_i}$$

Uso: modela selección al azar; loterías.

Ejemplo : Sea U distribuida uniformemente en los primero 10 elementos de la serie de Fibonacci. $S_U = \{1, 1, 2, 3, 4, 8, 13, 21, 34, 55\}$.

$$\mathbb{E}[u] = \frac{1}{10}\sum_{i=1}^{10} u_i = \frac{143}{10} = 14.3 = \bar{u}$$

$$\text{var}(U) = \frac{1}{10}\sum_{i=1}^{10}(u_i - \bar{u})^2 = 285.01$$

$$\text{de}(U) = \sqrt{285.01} = 16.88$$

8.2. Distribución Bernoulli

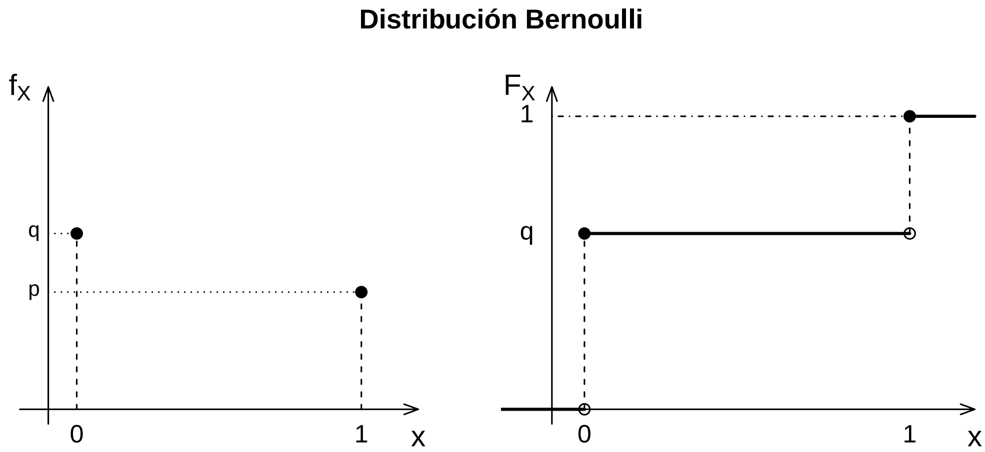


Figura 16: **Distribución Bernoulli**: Función masa de probabilidad f_X y correspondiente función de probabilidad acumulada F_X.

Sea X la variable aleatoria que sigue la **distribución Bernoulli** con probabilidad de éxito p. Se denota, $X \sim \text{Ber}(p)$. La figura 16 muestra el caso de una *f. m. p.* y la correspondiente función de distribución.

Parámetro: $0 \leq p \leq 1$.

Soporte: $S_X = \{0, 1\}$.

Función masa de probabilidad: sea $q = 1 - p$,

$$f_X(x) = q\mathbb{1}_{\{0\}}(x) + p\mathbb{1}_{\{1\}}(x)$$

Media:

$$\mathbb{E}[X] = 0 \cdot q + 1 \cdot p = p$$

Varianza:

$$\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X] = 0^2q + 1^2p - p^2 = p(1-p) = pq$$

Función generadora de momentos:

$$m_X(t) = qe^{0t} + pe^{1t} = q + pe^t$$

Uso: modela resultados aleatorios binarios: éxito/fracaso; enfermo/sano; encendido/apagado; abierto/cerrado; 1/0; etcétera.

Ejemplo : Considere una moneda cargada 6/10 al águila. Si $X = 1 =$ sol y $X = 0 =$ águila. Entonces, $p = 0.6$ y

$$\mathbb{E}[X] = p = 6/10 = 0.6$$
$$\text{var}(X) = pq = .6(.4) = 0.24$$
$$\text{de}(X) = \sqrt{.24} = 0.4899$$

8.3. Distribución Binomial

Sea X la variable aleatoria que sigue la **distribución Binomial** de tamaño n y probabilidad de éxito p. Se denota, $X \sim \text{Bin}(n, p)$. La figura 17 muestra el caso de una *f. m. p.* y la correspondiente función de distribución.

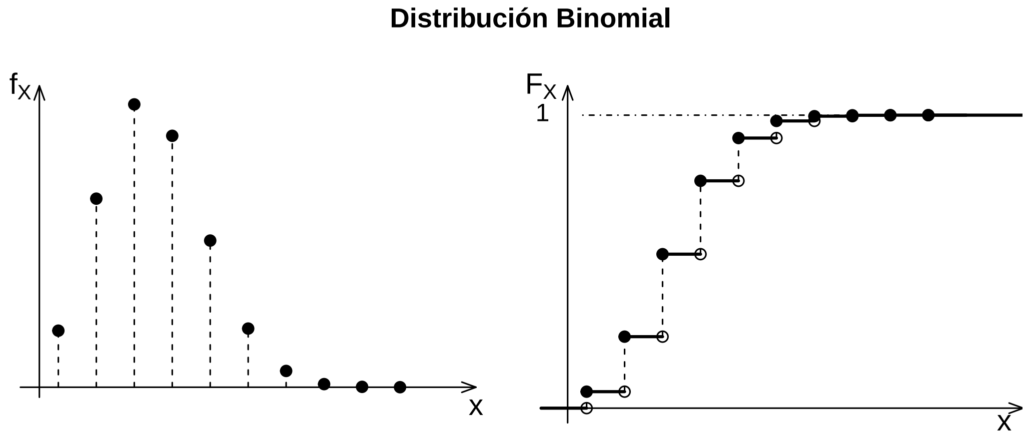


Figura 17: **Distribución binomial**: Función masa de probabilidad f_X y correspondiente función de probabilidad acumulada F_X.

Parámetros: $n \in \mathbb{N} = \{1, 2, 3, \ldots\}$ y $0 \leq p \leq 1$.

Soporte: $S_X = \{0, 1, \ldots, n\}$.

Función masa de probabilidad: sea $q = 1 - p$,

$$f_X(x) = \binom{n}{x} p^x q^{n-x} \mathbb{1}_{S_X}(x)$$

Media:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=0}^{n} \binom{n}{x} x p^x q^{n-x} \\
&= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-1-(x-1)} \\
&= np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y q^{n-1-y} \qquad ; y = x - 1 \\
&= np
\end{aligned}$$

Función generadora de momentos:

$$\begin{aligned}
m_X(t) &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x} \\
&= \left(pe^t + q\right)^n \qquad ; \text{teorema del binomio}
\end{aligned}$$

Así,

$$m_X(t) = \left(pe^t + q\right)^n, \quad t \in \mathbb{R}$$

Luego,

$$\begin{aligned}
m'_X(t) &= n(q+e^t)^{n-1}pe^t \\
m'_X(0) &= n(q+p)^{n-1}pe^t \\
&= np \\
m''_X(t) &= np\left[(q+pe^t)^{n-1}e^t + (n-1)\left(q+pe^t\right)^{n-2}pe^{2t}\right] \\
m''_X(0) &= np[1+(n-1)p] \\
&= np + n^2p^2 - np^2
\end{aligned}$$

Varianza:

$$\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X] = m''_X(0) - \left(m'_X(0)\right)^2 = np(1-p) = npq$$

Uso: modela el número de éxitos en n ensayos Bernoulli. Por ejemplo: número de soles en 30 volados; número de niños enfermos de 20 seleccionados al azar; etcétera.

Ejemplo : Suponga una línea de producción de partes automotrices. Las partes se embarcan en cajas de 100 y en el proceso de empacado 3 de cada 100 cajas son dañadas. Si usted inspecciona 20 cajas seleccionadas al azar, ¿cuántas cajas espera que estén dañadas? Calcule la probabilidad de que encontrar más de una caja dañada.

Solución: Sea X la *v. a.* que denota el número da cajas dañadas de las 20 inspeccionadas. $X \sim \text{Bin}(20, 3/100)$.

$$\begin{aligned}
\mathbb{E}[X] &= np = 20(3/100) = 0.60 \\
\text{de}(X) &= \sqrt{npq} = \sqrt{20(0.03)(0.97)} = \sqrt{0.582} = 0.7629 \\
\mathbb{P}(X \leq 1) &= \mathbb{P}(X=0) + \mathbb{P}(X=1) = \binom{20}{0}p^0q^{20} + \binom{20}{1}pq^{19} = 0.8801 \\
\mathbb{P}(X > 1) &= 1 - 0.8801 = 0.1199
\end{aligned}$$

En algunos textos se incluyen tablas de probabilidad para la distribución binomial entre otras. La figura 18 muestra un extracto tomado de **?** (**?**).

8.4. Distribución Geométrica o Pascal

Considere una sucesión de ensayos Bernoulli con probabilidad de éxito p. Sea X una variable aleatoria que denota el número de fracasos hasta el primer éxito. Luego, su rango o soporte es $R_X = \{0, 1, 2, \dots\}$ y si $q = 1-p$,

$$\begin{aligned}
\mathbb{P}(X=0) &= p \\
\mathbb{P}(X=1) &= qp \\
\mathbb{P}(X=2) &= qqp = q^2p \\
&\vdots \\
\mathbb{P}(X=k) &= \underbrace{q\cdots q}_{k \text{ veces}}\,p
\end{aligned}$$

La *v. a.* X se dice que sigue la **distribución geométrica** con probabilidad de éxito p y se denota $X \sim \text{Geom}(p)$. La figura 19 muestra el caso de una *f. m. p.* y la correspondiente función de distribución.

1. Distribución Binomial

$$X \sim \text{Binomial}(n, \pi)$$

$$p = P(X \leq x) = \sum_{k=0}^{x} \binom{n}{k} \pi^k (1-\pi)^{n-k} = 1 - \alpha$$

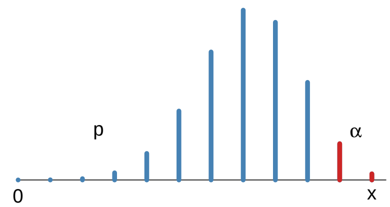


Tabla 1A. Probabilidades acumuladas p de la distribución binomial ($n = 5, 6, 7, 8, 9$).

		π														
	x	0.01	0.05	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9	0.95	0.99
$n=5$	0	0.951	0.774	0.590	0.328	0.237	0.168	0.078	0.031	0.010	0.002	0.001	0.000	0.000	0.000	0.000
	1	0.999	0.977	0.919	0.737	0.633	0.528	0.337	0.188	0.087	0.031	0.016	0.007	0.000	0.000	0.000
	2	1.000	0.999	0.991	0.942	0.896	0.837	0.683	0.500	0.317	0.163	0.104	0.058	0.009	0.001	0.000
	3	1.000	1.000	1.000	0.993	0.984	0.969	0.913	0.813	0.663	0.472	0.367	0.263	0.081	0.023	0.001
	4	1.000	1.000	1.000	1.000	0.999	0.998	0.990	0.969	0.922	0.832	0.763	0.672	0.410	0.226	0.049
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=6$	0	0.941	0.735	0.531	0.262	0.178	0.118	0.047	0.016	0.004	0.001	0.000	0.000	0.000	0.000	0.000
	1	0.999	0.967	0.886	0.655	0.534	0.420	0.233	0.109	0.041	0.011	0.005	0.002	0.000	0.000	0.000
	2	1.000	0.998	0.984	0.901	0.831	0.744	0.544	0.344	0.179	0.070	0.038	0.017	0.001	0.000	0.000
	3	1.000	1.000	0.999	0.983	0.962	0.930	0.821	0.656	0.456	0.256	0.169	0.099	0.016	0.002	0.000
	4	1.000	1.000	1.000	0.998	0.995	0.989	0.959	0.891	0.767	0.580	0.466	0.345	0.114	0.033	0.001
	5	1.000	1.000	1.000	1.000	1.000	0.999	0.996	0.984	0.953	0.882	0.822	0.738	0.469	0.265	0.059
	6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=7$	0	0.932	0.698	0.478	0.210	0.133	0.082	0.028	0.008	0.002	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.998	0.956	0.850	0.577	0.445	0.329	0.159	0.063	0.019	0.004	0.001	0.000	0.000	0.000	0.000
	2	1.000	0.996	0.974	0.852	0.756	0.647	0.420	0.227	0.096	0.029	0.013	0.005	0.000	0.000	0.000
	3	1.000	1.000	0.997	0.967	0.929	0.874	0.710	0.500	0.290	0.126	0.071	0.033	0.003	0.000	0.000
	4	1.000	1.000	1.000	0.995	0.987	0.971	0.904	0.773	0.580	0.353	0.244	0.148	0.026	0.004	0.000
	5	1.000	1.000	1.000	1.000	0.999	0.996	0.981	0.938	0.841	0.671	0.555	0.423	0.150	0.044	0.002
	6	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.992	0.972	0.918	0.867	0.790	0.522	0.302	0.068
	7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=8$	0	0.923	0.663	0.430	0.168	0.100	0.058	0.017	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.997	0.943	0.813	0.503	0.367	0.255	0.106	0.035	0.009	0.001	0.000	0.000	0.000	0.000	0.000
	2	1.000	0.994	0.962	0.797	0.679	0.552	0.315	0.145	0.050	0.011	0.004	0.001	0.000	0.000	0.000
	3	1.000	1.000	0.995	0.944	0.886	0.806	0.594	0.363	0.174	0.058	0.027	0.010	0.000	0.000	0.000
	4	1.000	1.000	1.000	0.990	0.973	0.942	0.826	0.637	0.406	0.194	0.114	0.056	0.005	0.000	0.000
	5	1.000	1.000	1.000	0.999	0.996	0.989	0.950	0.855	0.685	0.448	0.321	0.203	0.038	0.006	0.000
	6	1.000	1.000	1.000	1.000	1.000	0.999	0.991	0.965	0.894	0.745	0.633	0.497	0.187	0.057	0.003
	7	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.996	0.983	0.942	0.900	0.832	0.570	0.337	0.077
	8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=9$	0	0.914	0.630	0.387	0.134	0.075	0.040	0.010	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.997	0.929	0.775	0.436	0.300	0.196	0.071	0.020	0.004	0.000	0.000	0.000	0.000	0.000	0.000
	2	1.000	0.992	0.947	0.738	0.601	0.463	0.232	0.090	0.025	0.004	0.001	0.000	0.000	0.000	0.000
	3	1.000	0.999	0.992	0.914	0.834	0.730	0.483	0.254	0.099	0.025	0.010	0.003	0.000	0.000	0.000
	4	1.000	1.000	0.999	0.980	0.951	0.901	0.733	0.500	0.267	0.099	0.049	0.020	0.001	0.000	0.000
	5	1.000	1.000	1.000	0.997	0.990	0.975	0.901	0.746	0.517	0.270	0.166	0.086	0.008	0.001	0.000
	6	1.000	1.000	1.000	1.000	0.999	0.996	0.975	0.910	0.768	0.537	0.399	0.262	0.053	0.008	0.000
	7	1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.980	0.929	0.804	0.700	0.564	0.225	0.071	0.003
	8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.990	0.960	0.925	0.866	0.613	0.370	0.086
	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Figura 18: Extracto de las tablas de probabilidad de la distribución binomial.

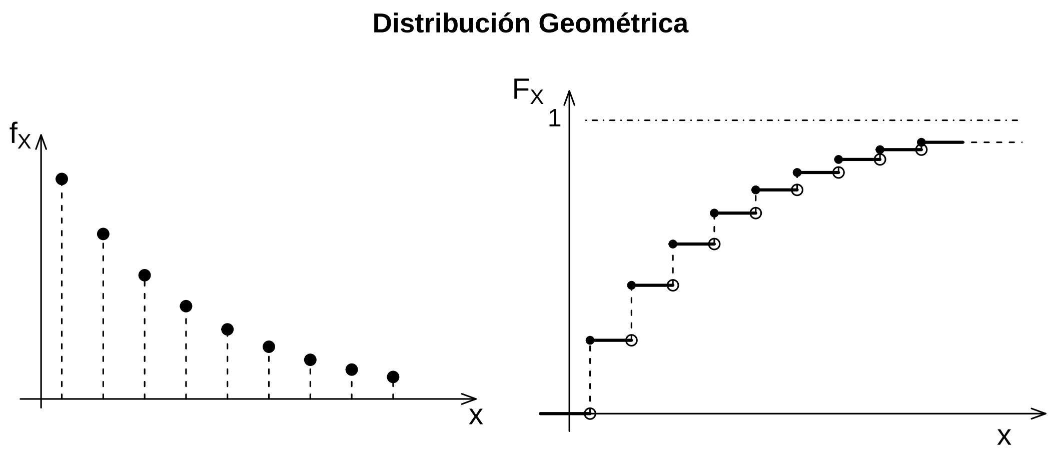


Figura 19: **Distribución geométrica**: Función masa de probabilidad f_X y correspondiente función de probabilidad acumulada F_X.

Parámetro: $0 \leq p \leq 1$.

Soporte: $S_X = \mathbb{N}_0 = \{0, 1, 2, \dots\}$.

Función masa de probabilidad: sea $q = 1 - p$,

$$f_X(x) = pq^x \mathbb{1}_{S_X}(x)$$

Se vio ya antes que f_X es una *f. m. p.* propia en el sentido que: *i)* $f_X(x) \geq 0$, para todo $x \in \mathbb{R}$; y *ii)* $\sum_{x=0}^{\infty} f_X(x) = 1$.

Se vio también que si Y sigue una distribución geométrica con soporte en $\mathbb{N}$, $f_Y(y) = pq^{y-1}\mathbb{1}_{\mathbb{N}}(y)$ y $\mathbb{E}[Y] = 1/p$, por lo que si $X \sim \text{Geom}(p)$ con soporte en $\mathbb{N}_0$, $X = Y - 1$, $f_X(x) = pq^x \mathbb{1}_{\mathbb{N}_0}(x)$ y $\mathbb{E}[X] = q/p$. Una manera alternativa de calcular el valor esperado de X es,

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=0}^{\infty} xpq^x \\
&= pq \sum_{x=1}^{\infty} xq^{x-1} \\
&= pq \sum_{x=1}^{\infty} \frac{d}{dq} q^x \\
&= pq \frac{d}{dq} \left(\sum_{x=0}^{\infty} q^x - 1 \right) \\
&= pq \frac{d}{dq} \left(\frac{q}{1-q} \right) \\
&= pq \frac{1}{(1-q)^2} \\
&= q/p
\end{aligned}$$

Función generadora de momentos: $m_X(t) = \mathbb{E}[e^{tX}]$

$$\begin{aligned} m_X(t) &= \sum_{x=0}^{\infty} e^{tx} pq^x \\ &= p \sum_{x=0}^{\infty} \left(qe^t\right)^x \\ &= p \frac{1}{1-qe^t} \end{aligned}$$

para $qe^t < 1$. Por lo tanto,

$$m_X(t) = p(1-qe^t)^{-1}, \qquad \text{para } t < \ln 1/q$$

Ahora,

$$\begin{aligned} m_X'(t) &= -p(1-qe^t)^{-2}(-qe^t) \\ m_X'(0) &= pq(1-q)^{-2} = qp^{-1} \\ m_X''(t) &= -pqe^t(1-qe^t)(1-qe^t)^{-3}(-qe^t) + pq(1-qe^t)^{-2}e^t \\ m_X''(0) &= pq(1-q)^{-2}[2q/p+1] = 2q^2/p^2 \end{aligned}$$

Media:

$$\mathbb{E}[X] = m_X(0) = \frac{q}{p}$$

Varianza:

$$\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X] = m_X''(0) - \left(m_X'(0)\right)^2 = 2\frac{q^2}{p^2} + \frac{q}{p} = \frac{q}{p^2}$$

Proposición : Sea X variable aleatoria distribuida geométricamente con probabilidad de éxito p y soporte $S_X = \mathbb{N}_0$. Entonces, para $q = 1-p$,

$$\mathbb{P}(X \geq x) = q^x, \quad \text{para } x \in \{0,1,2,\dots\}$$

Demostración:

$$\begin{aligned} \mathbb{P}(X \geq x) &= 1 - \mathbb{P}(X \leq x-1) \\ &= 1 - \sum_{k=0}^{x-1} pq^k \\ &= 1 - p\frac{1-q^x}{1-q} \\ &= q^x \end{aligned}$$

Corolario : Si $X \sim \text{Geom}(p)$ con soporte $\mathbb{N}_0$, entonces, para $m,n \in \mathbb{N}_0$,

$$\mathbb{P}(X \geq m+n | X \geq m) = \mathbb{P}(X \geq n)$$

Demostración:

$$\mathbb{P}(X \geq m+n | X \geq m) = \frac{\mathbb{P}(X \geq m+n)}{\mathbb{P}(X \geq m)} = \frac{q^{m+n}}{q^m} = \mathbb{P}(X \geq n)$$

Pérdida de memoria

La siguiente propiedad se incluye como la presnta Hoel et al. (1971). Considere un dispositivo que no mejore ni empeora con el tiempo. El dispositivo puede fallar por algún evento aleatorio que ocurre homogéneamente en el tiempo.

Sea Y la variable aleatoria que denota las unidades de tiempo medido de manera discreta (horas, días, etc.) hasta que el dispositivo falla. Luego, $S_Y = \{1, 2, 3, \dots\}$, El evento $\{Y = n\}$ ocurre si y solo si el dispositivo falla al tiempo n.

Ahora bien, intuitivamente, si el dispositivo no mejora ni empeora con el tiempo, se debe tener que

$$\mathbb{P}(Y > n + m | Y > n) = \mathbb{P}(Y > m)$$

propiedad que se conoce como **pérdida de memoria**, pues Y "olvida" que ya operó n unidades e ir por m unidades adicionales es igual que si apenas empezó a operar. Si se cumple la pérdida de memoria se tiene que

$$\mathbb{P}(Y > n + m) = \mathbb{P}(Y > n)\mathbb{P}(Y > m), \qquad n, m = 0, 1, 2, \dots \tag{2}$$

Si la variable aleatoria discreta Y satisface la propiedad anterior entonces sigue una distribución geométrica. A saber,

a). Suponga que $n = m = 0$. Luego, $\mathbb{P}(Y > 0) = \mathbb{P}(Y > 0)^2$, por lo que $\mathbb{P}(Y > 0) = 0$ ó $\mathbb{P}(Y > 0) = 1$. Pero si $\mathbb{P}(Y > 0) = 0$, entonces $\mathbb{P}(Y = 0) = 1$ y el dispositivo nunca funcionó.

b). Suponga que $\mathbb{P}(Y > 0) = 1$, sea $p = \mathbb{P}(Y = 1)$ y $q = 1 - p = \mathbb{P}(Y > 1)$.

$$\begin{aligned}
\mathbb{P}(Y > 2) &= \mathbb{P}(Y > 1)\mathbb{P}(Y > 1) = q^2 \\
\mathbb{P}(Y > 3) &= \mathbb{P}(Y > 1)\mathbb{P}(Y > 2) = q^3 \\
&\vdots \\
\mathbb{P}(Y > n) &= \mathbb{P}(Y > 1)\mathbb{P}(Y > n - 1) = q^n
\end{aligned}$$

Por lo que

$$\mathbb{P}(Y = n) = \mathbb{P}(Y > n - 1) - \mathbb{P}(Y > n) = q^{n-1}(1 - q) = pq^{n-1}, \quad n = 1, 2, \dots$$

ı) Si $p = 0$, $\mathbb{P}(Y > n) = 0 \Longrightarrow \mathbb{P}(Y = \infty) = 1$. ¡El dispositivo nunca falla!

ii) Si $p = 1$, $\mathbb{P}(Y = 1) = 1$. !El dispositivo nunca funcionó!

Sea ahora $X = Y - 1$. Entonces, el soporte de X, $S_X = \{0, 1, 2, \dots\}$

$$f_X(x) = \mathbb{P}(X = x) = \mathbb{P}(Y = x + 1) = pq^x$$

Por lo tanto, $X \sim \text{Geom}(p)$.

Note que la propiedad pérdida de memoria anterior podría escribirse como

$$\begin{aligned}
\mathbb{P}(Y > n + m | Y > n) &= \mathbb{P}(Y > m) \\
\mathbb{P}(X \geq n + m | X \geq n) &= \mathbb{P}(Y \geq m)
\end{aligned}$$

Proposición : Sea X una variable aleatoria discreta. X tiene la propiedad *pérdida de memoria* si y solo si X sigue la distribución geométrica.

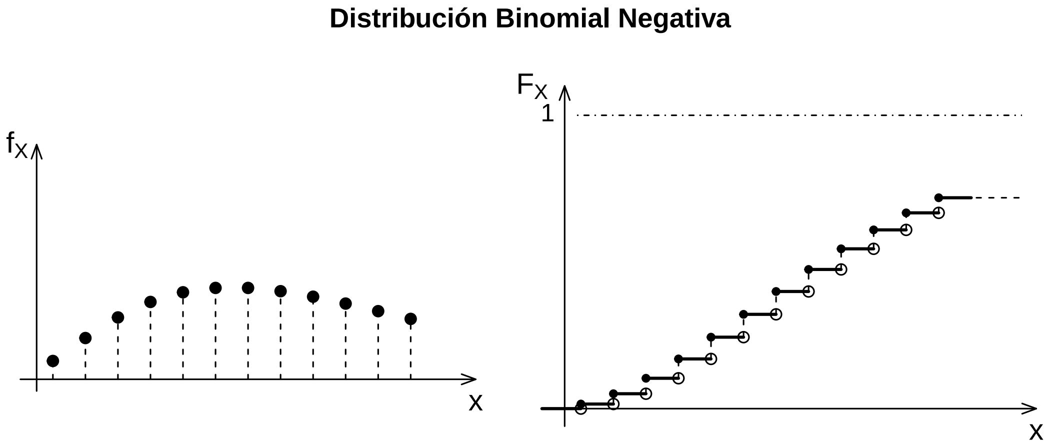


Figura 20: **Distribución binomial negativa**: Función masa de probabilidad f_X y correspondiente función de probabilidad acumulada F_X.

8.5. Distribución Binomial Negativa

Esta sección se apoya en resultados de Hoel et al. (1971) y Mood et al. (1974).

La distribución binomial negativa puede verse como una extensión a la distribución geométrica.[3] Y en ese caso, la distribución geométrica resulta un caos particular de la distribución binomial negativa. La figura 20 muestra el caso de una *f. m. p.* y la correspondiente función de distribución.

Considere una sucesión de ensayos Bernoulli con probabilidad de éxito p. Sea X la variable aleatoria que denota el número de fracasos hasta en el r-ésimo éxito, $r \in \mathbb{N}$.

- Si no hay fracazos antes de los r éxitos

$$\mathbb{P}(X = 0) = \underbrace{p \cdots p}_{r \text{ éxitos}} = p^r$$

- Si hay un fracaso antes del r-ésimo éxito, $X = 1$ y hay $\binom{r}{1} = r - 1$ posiciones donde localizar el fracaso

$$\mathbb{P}(X = 1) = \binom{r}{1} \underbrace{p \cdots p}_{(r-1) \text{ éxitos}} qp = rp^r q$$

- Si hay 2 fracaso antes del r-ésimo éxito, $X = 2$ y hay $\binom{r+1}{2}$ posiciones de donde localizar los 2 fracasos

$$\mathbb{P}(X = 2) = \binom{r+1}{2} p \cdots pq^2 p = \binom{r+1}{2} p^r q^2$$

- Si hay k fracasos antes del r-ésimo éxito, $X = k$ y hay $\binom{r+k-1}{k}$ posiciones de donde localizar los k fracasos

$$\mathbb{P}(X = k) = \binom{r+k-1}{k} p^r q^k$$

La variable aleatoria X se dice que sigue la **distribución binomial negativa** para $r \in \mathbb{N}$ y $0 \leq p \leq 1$, y se denota $X \sim \text{BinNeg}(r, p)$.

Parámetros: $r \in \mathbb{N}$, $0 \leq p \leq 1$.

[3]En el curso de Cálculo de Probabilidades II se verá que la suma de *v. a.*'s independientes distribuidas geométricamente se distribuye binomial negativa.

Soporte: $S_X = \{0, 1, 2, \dots\}$.

Función masa de probabilidad, sea $q = 1 - p$,

$$f_X(x) = \binom{r+x-1}{x} p^r q^x \mathbb{1}_{S_X}(x)$$

Corolario : Si $X \sim \text{Geom}(p)$, entonces $X \sim \text{BinNeg}(1, p)$.

Para $r \in \mathbb{R}$ y $x \in \{0, 1, 2, \dots\}$

$$\begin{aligned}\binom{-r}{x} &= \frac{(-r)_x}{x!} = \frac{(-r)(-r-1)\cdots(-r-x+1)}{x!} = \frac{(-1)^x (r+x-1)_x}{x!} \\ &= (-1)^x \binom{r+x-1}{x}\end{aligned}$$

Note además que si $g(t) = (1-t)^{-\alpha}$, para $|t| < 1$. Si se expande en serie de Taylor al rededor de 0,

$$g(t) = (1-t)^{-\alpha} = \sum_{k=1}^{\infty} \binom{-\alpha}{k} (-t)^k = \sum_{k=0}^{\infty} (-1)^k \binom{-\alpha}{k} t^k$$

Entonces, para mostrar que f_X es una *f. m. p.* propia,

ii) $f_X(x) = \binom{r+x-1}{x} p^r q^x \geq 0$, para todo $x \in \mathbb{R}$.

ii) $\sum_{x=0}^{\infty} f_X(x) = \sum_{x=o}^{\infty} \binom{r+x-1}{x} p^r q^x = p^r \sum_{x=0}^{\infty} (-1)^x \binom{-r}{x} q^x = p^r p^{-r} = 1$
aplicando el resultado anterior con $g(t) = (1-t)^{-r}$ y $t = q = 1 - p$.

Proposición : Sea $X \sim \text{BinNeg}(r, p)$. Entonces, su función generadora de momentos está dada por

$$m_X(t) = p^r (1 - qe^t)^{-r}. \quad t < -\ln q$$

Demostración: $m_X(t) = \mathbb{E}[e^{tX}]$.

$$\begin{aligned} m_X(t) &= \sum_{x=0}^{\infty} e^{tx} f_X(x) \\ &= \sum_{x=0}^{\infty} \binom{r+x-1}{x} p^r (qe^t)^x \\ &= p^r \sum_{x=0}^{\infty} (-1)^x \binom{-r}{x} (qe^t)^x \\ &= p^r (1 - qe^t)^{-r} \end{aligned}$$

Corolario : $X \sim \text{BinNeg}(r, p)$,

$$\mathbb{E}[X] = r\frac{1-p}{p}, \qquad \text{var}(X) = r\frac{1-p}{p^2}$$

Demostración: Se sigue del hecho $m_X^{(k)}(0) = \mathbb{E}[X^k]$, para $k = 1, 2, \dots$, aplicado a $m_X(t) = p^r (1 - qe^t)^{-r}$, con $q = 1 - p$.

Nota: Si $Y = X + r$, $S_Y = \{r, r_1, r+2, \dots\}$. Función masa de probabilidad

$$f_Y(y) = \binom{y-1}{r-1} p^r q^{y-r} \mathbb{1}_{S_Y}(y)$$

Función generadora de momentos

$$m_Y(t) = p^r e^{rt}(1 - qe^t)^{-r}. \quad t < -\ln(1-p)$$

Media y varianza

$$\mathbb{E}[Y] = \frac{r}{p}, \qquad \mathrm{var}(Y) = r\frac{q}{p^2}$$

Ejemplo : Suponga que an alguna población una proporción p tiene cierta característica. Si se muestrea de manera aleatoria hasta que haya exactamente r sujetos con la característica, entonces el número de elementos muestreados sigue una distribución binomial negativa.

8.6. Distribución Poisson

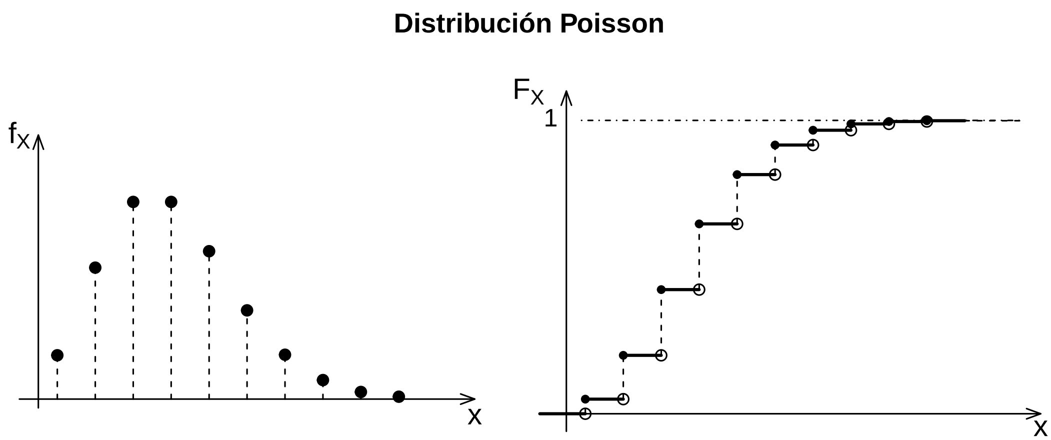


Figura 21: **Distribución Poisson**: Función masa de probabilidad f_X y correspondiente función de probabilidad acumulada F_X.

Sea X una variable aleatoria entera no negativa con tal que para $x = 0, 1, 2, \dots,$

$$\mathbb{P}(X = x) = \lambda^x \frac{e^{-\lambda}}{x!}$$

para algún $\lambda > 0$. X se dice que sigue una **distribución Poisson** de media λ y se denota por $X \sim \mathrm{Po}(\lambda)$. Luego:

Parámetro: $\lambda > 0$.

Soporte: $S_X = \mathbb{N}_0 = \{0, 1, 2, \dots\}$.

Función masa de probabilidad:

$$f_X(x) = \lambda^x \frac{e^{-\lambda}}{x!} \mathbb{1}_{\mathbb{N}_0}(x)$$

La *f. m. p.* es propia pues:

i) $f_X(x) = \lambda^x \frac{e^{-\lambda}}{x!} \geq 0$, para todo $x \in \mathbb{R}$.

ii) $\sum_{x \in S_X}^{\infty} f_X(x) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$.

Función generadora de momentos:

$$\begin{aligned} m_X(t) &= \sum_{x \in S_X} e^{tx} f_X(x) \\ &= \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^X e^{-\lambda}}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \\ &= e^{-\lambda} e^{\lambda e^t} \\ &= e^{\lambda(e^t-1)} \end{aligned}$$

Se sigue que

$$\begin{aligned} m'_X(t) &= \lambda e^t e^{\lambda(e^t-1)} \\ m'_X(0) &= \lambda \\ m''_X(t) &= \lambda e^t e^{\lambda(e^t-1)}[1+\lambda e^t] \\ m''_X(0) &= \lambda[1+\lambda] \end{aligned}$$

Por lo tanto,

Media:

$$\mathbb{E}[X] = m'_X(0) = \lambda$$

Varianza:

$$\mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X] = \lambda[1+\lambda] - \lambda^2 = \lambda$$

La distribución Poisson modela el número de ocurrencias de *"eventos raros E"*, conteos. Por ejemplo, el número de accidentes en cierto periodo; el número de defectos en una pieza; el número de errores en cierto procedimiento.

Al igual que la distribución binomial, algunos texto incluyen tablas de la distribución Poisson. La figura 22 muestra un extracto tomado de **?** (**?**).

Finalmente, considere el número de ocurrencias de E en un cierto intervalo de tiempo $[0, T]$. Suponga que el intervalo $[0, T]$ se divide en n subintervalos con las siguientes características:

Supuestos:

1. La probabilidad de más de una ocurrencia en un subintervalo es despreciable.
2. Los intervalos son tan pequeños que la ocurrencia o no de un evento E en el subintervalo no afecta en la ocurrencia o no de los otros subintervalos.
3. La probabilidad de ocurrencia (o no) del evento E en un subintervalo es p ($q = 1 - p$) y es la misma en todos los subintervalos.

2. Distribución Poisson

$$X \sim \text{Poisson}(\lambda)$$

$$p = P(X \leq x) = \sum_{k=0}^{x} \frac{\lambda^k e^{-\lambda}}{k!} = 1 - \alpha$$

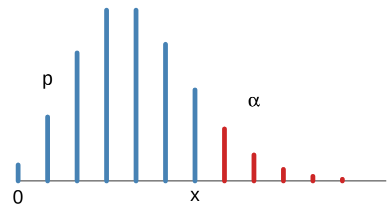


Tabla 2A. Probabilidades acumuladas p de la distribución Poisson.

	λ									
x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0.905	0.819	0.741	0.670	0.607	0.549	0.497	0.449	0.407	0.368
1	0.995	0.982	0.963	0.938	0.910	0.878	0.844	0.809	0.772	0.736
2	1.000	0.999	0.996	0.992	0.986	0.977	0.966	0.953	0.937	0.920
3	1.000	1.000	1.000	0.999	0.998	0.997	0.994	0.991	0.987	0.981
4	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.996
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Tabla 2B. Probabilidades acumuladas p de la distribución Poisson.

	λ										
x	2	3	4	5	6	7	8	9	10	15	20
0	0.135	0.050	0.018	0.007	0.002	0.001	0.000	0.000	0.000	0.000	0.000
1	0.406	0.199	0.092	0.040	0.017	0.007	0.003	0.001	0.000	0.000	0.000
2	0.677	0.423	0.238	0.125	0.062	0.030	0.014	0.006	0.003	0.000	0.000
3	0.857	0.647	0.433	0.265	0.151	0.082	0.042	0.021	0.010	0.000	0.000
4	0.947	0.815	0.629	0.440	0.285	0.173	0.100	0.055	0.029	0.001	0.000
5	0.983	0.916	0.785	0.616	0.446	0.301	0.191	0.116	0.067	0.003	0.000
6	0.995	0.966	0.889	0.762	0.606	0.450	0.313	0.207	0.130	0.008	0.000
7	0.999	0.988	0.949	0.867	0.744	0.599	0.453	0.324	0.220	0.018	0.001
8	1.000	0.996	0.979	0.932	0.847	0.729	0.593	0.456	0.333	0.037	0.002
9	1.000	0.999	0.992	0.968	0.916	0.830	0.717	0.587	0.458	0.070	0.005
10	1.000	1.000	0.997	0.986	0.957	0.901	0.816	0.706	0.583	0.118	0.011
11	1.000	1.000	0.999	0.995	0.980	0.947	0.888	0.803	0.697	0.185	0.021
12	1.000	1.000	1.000	0.998	0.991	0.973	0.936	0.876	0.792	0.268	0.039
13	1.000	1.000	1.000	0.999	0.996	0.987	0.966	0.926	0.864	0.363	0.066
14	1.000	1.000	1.000	1.000	0.999	0.994	0.983	0.959	0.917	0.466	0.105
15	1.000	1.000	1.000	1.000	0.999	0.998	0.992	0.978	0.951	0.568	0.157
16	1.000	1.000	1.000	1.000	1.000	0.999	0.996	0.989	0.973	0.664	0.221
17	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.995	0.986	0.749	0.297
18	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.993	0.819	0.381
19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.997	0.875	0.470
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.917	0.559
21	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.947	0.644
22	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.967	0.721
23	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.981	0.787
24	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.989	0.843
25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.994	0.888
26	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.922
27	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.948
28	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.966
29	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.978
30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.987
31	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.992
32	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995
33	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997
34	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
35	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
36	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Figura 22: Extracto de las tablas de probabilidad de la distribución Poisson.

Si Y denota el número de subintervalos con ocurrencia, entonces $Y \sim \text{Bin}(n, p)$. Sea ahora $\lambda = np$ y para $y = 0, 1, \ldots, n$,

$$\begin{aligned}
\mathbb{P}(Y = y) &= \binom{n}{y} p^y q^{n-y} \\
&= \frac{(n)_y}{y!} p^y q^{n-y} \\
&= \frac{n(n-1)\cdots(n-y+1)}{y!} \left(\frac{\lambda}{n}\right)^y \left(1 - \frac{\lambda}{n}\right)^{n-y} \\
&= \frac{\lambda^y}{y!} \frac{n}{n} \frac{n-1}{n} \cdots \frac{n-y+1}{n} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-y} \\
&\quad \text{haciendo } n \to \infty \\
&= \frac{\lambda^y}{y!} 1 \cdots e^{-\lambda} \cdot 1 \\
&= e^{-\lambda} \frac{\lambda^y}{y!}
\end{aligned}$$

Por lo que la *f. m. p.* de Y está dada por

$$f_Y(y) = e^{-\lambda} \frac{\lambda^y}{y!} \mathbb{1}_{\{0,1,2,\ldots\}}(y)$$

Aproximación de la distribución binomial por la Poisson

Como recién se ilustró, es posible derivar la distribución de Poisson a partir de la binomial. Luego, bajo ciertas condiciones es posible acercar la distribución de Poisson a la binomial. Esto fue relevante en la práctica pues no se disponía de tablas de probabilidad para valores grandes de n o cualquier valor de p.

Considere $X \sim \text{Bin}(n, p)$ y suponga que $Y \sim \text{Po}(\lambda)$, luego es razonable empatar $np = \mathbb{E}[X] = \mathbb{E}[Y] = \lambda$. En este caso, para $k = 0, 1, \ldots, n$,

$$\mathbb{P}(X \leq k) \approx \mathbb{P}(Y \leq k)$$

Algunos textos proponen las siguientes reglas:

- *Aproximación razonable:* si $n > 20$, $np \geq 5$ y $n(1 - p) \geq 5$.
- *Aproximación es buena:* si $n > 100$ o $p < 0.01$, siempre que $np = \lambda = 20$.

La siguiente tabla muestra el valor de la probabilidad $\mathbb{P}(X \leq x)$ calculada con base en la distribución binomial, $\mathbb{P}(Y \leq x)$ calculada con la distribución Poisson. $X \sim \text{Bin}(30, 0.1)$, $\lambda = np = 3$, $Y \sim \text{Po}(3)$.

x	0	1	2	3	4	5	6	7	8	9
$\mathbb{P}(X \leq x)$	0.042	0.184	0.411	0.647	0.825	0.927	0.974	0.992	0.998	1.000
$\mathbb{P}(Y \leq x)$	0.050	0.199	0.423	0.647	0.815	0.916	0.966	0.988	0.996	0.999

Ejemplo : (Wackerly et al. (2008) p. 135) Para analizar la capacidad de un nuevo bar se estudia la conglomeración de gente en una cuadricula de 2m^2. Se determina que en promedio por cada 2m^2 hay en promedio 4 personas en un día de operaciones.

a). Calcule la probabilidad de no haya personas en un espacio de 2m^2 elegido al azar.

b). Calcule la probabilidad de haya a lo menos tres personas en un espacio de 4m^2 elegido al azar.

Solución:

a). Sea X la *v. a.* que denota el número de personas en un espacio de 2m^2. Entonces $X \sim \text{Poisson}(\lambda = 4)$. Ahora,

$$\mathbb{P}(X=0) = \frac{(4)^0 e^{-4}}{0!} = e^{-4} = 0.018$$

b). "Si observamos un proceso Poisson [llamado X] y λ es el número medio de sucesos por unidad (longitud, área, etc.), entonces Y = número de sucesos en a unidades tiene una distribución Poisson con media $a\lambda$.".
Sea Y la *v. a.* que denota el número de personas en un espacio de 4m^2, se tiene que $Y \sim \text{Poisson}(\lambda' = 8)$. Por lo tanto,

$$\begin{aligned}\mathbb{P}(Y \geq 3) &= 1 - \mathbb{P}(Y < 3) \\ &= 1 - \mathbb{P}(Y \leq 2) \\ &= 1 - \big[\mathbb{P}(Y=0) + \mathbb{P}(Y=1) + \mathbb{P}(Y=2)\big] \\ &= 0.986\end{aligned}$$

8.7. Distribución Hipergeométrica

Esta sección se apoya en (Mood et al. 1974).

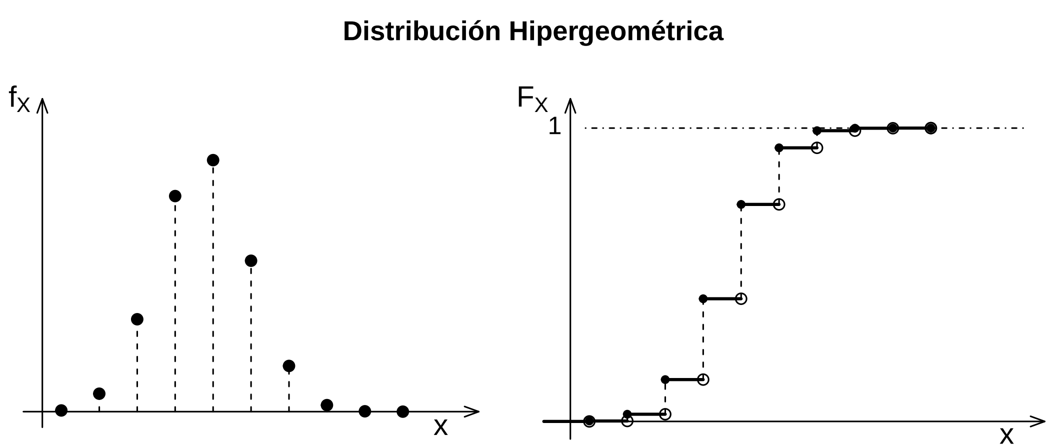
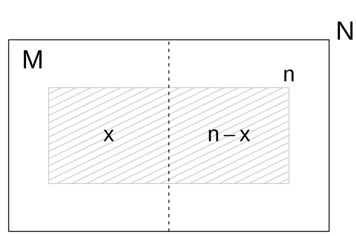


Figura 23: **Distribución hypergeométrica**: Función masa de probabilidad f_X y correspondiente función de probabilidad acumulada F_X.

Suponga que se tiene una población con N elementos de los cuales M son de clase $\mathcal{C}_1$ y $N-M$ de clase $\mathcal{C}_2$. Se toma una muestra aleatoria (no ordenada y sin reemplazo) de tamaño n.



Sea X la variable aleatoria que denota el número de elementos de la clase $\mathcal{C}_1$. Como se vio en Técnicas de Conteo, la distribución de X pude abordarse como un problema de particiones. Así, para $x = 0, 1, \ldots, n$,

$$\mathbb{P}(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{M+N}{n}}$$

La variable aleatoria X se dice que sigue la **distribución hipergeométrica**, de parámentros M, N y n y se denota como $X \sim \mathrm{HGeom}(N, M, n)$,

Parámetros: $N = 1, 2, 3, \ldots$; $M = 1, \ldots, N$; $n = 0, 1, \ldots, N$

Soporte: $S_X = \{0, 1, \ldots, n\}$

Función masa de probabilidad:

$$f_X(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \mathbb{1}_{S_X}(x)$$

Proposición : Sea X una variable aleatoria con distribución hipergeométrica $\mathrm{HGeom}(N, M, n)$. Entonces,

a). $E[X] = n\dfrac{M}{N}$

b). $\mathbb{E}[X(X-1)] = n(n-1)\,\dfrac{M(M-1)}{N(N-1)}$

c). $\mathrm{var}(X) = \dfrac{nM}{N}\,\dfrac{N-M}{N}\,\dfrac{N-n}{N-1}$

Demostración: Para da demostración de las propiedades recién enunciadas se apoya en el teorema de Binomial y corolarios, que se presentan al final de la sección.

a). $\mathbb{E}[X] = \sum_{x \in S_X} x f_X(x)$.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=0}^{n} x \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \\
&= \sum_{x=1}^{n} x \frac{\frac{M}{x}\binom{M-1}{x-1}\binom{N-M}{n-x}}{\frac{N}{n}\binom{N-1}{n-1}} \\
&= n\frac{M}{N} \sum_{y=0}^{n-1} \frac{\binom{M-1}{y}\binom{N-M}{n-1-y}}{\binom{N-1}{n-1}} \\
&= n\frac{M}{N}
\end{aligned}$$

La última suma es 1 pues es la suma sobre todo el soporte de la *f. m. p.*.

b).

$$\mathbb{E}[X(X-1)] = \sum_{x \in S_X} x(x-1) f_X(x) = n(n-1)\frac{M(M-1)}{N(N-1)}$$

Se muestra de manera similar al inciso anterior.

c). $\mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X]$,

$$\begin{aligned}
\mathrm{var}(X) &= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}^2[X] \\
&= \frac{nM}{N}\left[\frac{(N-M)(N-n)}{N(N-1)}\right]
\end{aligned}$$

Nota: Si denota $p = \frac{M}{N}$, entonces $\mathbb{E}[X]$ coincide con la esperanza de $Y \sim \text{Bin}(n,p)$, $\mathbb{E}[Y] = np$, y $\text{var}(Y) = np(1-p)$, cuando $\text{var}(X) = \frac{N-n}{N-1}np(1-p)$.

Teorema Binomial

$$(a+b)^n = \sum_{k=1}^{n} \binom{n}{k} a^k b^{n-k}$$

Nota: Se sigue el teorema que:

a). $(1+t)^n = \displaystyle\sum_{k=1}^{n} \binom{n}{k} t^k$

b). $(1-t)^n = \displaystyle\sum_{k=0}^{n} \binom{n}{k} (-1)^k t^k$

c). $2^n = (1+1)^n = \displaystyle\sum_{k=1}^{n} \binom{n}{k}$

d). $0 = (1-1)^n = \displaystyle\sum_{k=1}^{n} (-1)^k \binom{n}{k}$

e). $\displaystyle\binom{a+b}{n} = \sum_{k=0}^{n} \binom{a}{k} \binom{b}{n-k}$

Para esta última igualdad considere el siguiente desarrollo expandiendo ambos lados de la igualdad y empatando coeficientes de x^k,

$$\begin{aligned}
(1+x)^a (1=x)^b &= (1+x)^{a+b} \\
\sum_{i=0}^{a} \binom{a}{i} x^i \sum_{j=0}^{b} \binom{b}{j} x^j &= \sum_{k=0}^{a+b} \binom{a+b}{k} x^k \\
\sum_{i=0}^{a} \sum_{j=0}^{b} x^{i+j} &= \sum_{k=0}^{a+b} \binom{a+b}{k} x^k
\end{aligned}$$

a). $k = 0$,

$$\binom{a}{0}\binom{b}{0} = \binom{a+b}{0}$$

b). $k = 1$,

$$\binom{a}{1}\binom{b}{0} + \binom{a}{0}\binom{b}{1} = \binom{a+b}{1}$$

c). $k = 2$,

$$\binom{a}{2}\binom{b}{0} + \binom{a}{1}\binom{b}{1} + \binom{a}{0}\binom{b}{2} = \binom{a+b}{2}$$

$\vdots$

d). $k = n$,

$$\sum_{\ell=0}^{n} \binom{a}{\ell}\binom{b}{n-\ell} = \binom{a+b}{n}$$

Por lo tanto,

$$\sum_{k=0}^{n} \binom{a}{k}\binom{b}{n-k} = \binom{a+b}{n}$$

Esta última igualdad se conoce como la **Identidad o Convolución de Vandermonde**.

Ejercicio : Sea X el número de artículos defectuosos de una muestra de tamaño n cuando fue extraída sin reemplazo de una urna con N bolas, M de ellas defectuosas, entonces $X \sim \mathrm{HGeom}(N, M, n)$. Determine la probabilidad de encontrar al menos dos defectuosos en una muestra de tamaño $n = 5$, si $M = 10$ y $N = 50$.

8.8. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 8, Barrios and Heiras (2024).

9. Distribuciones Continuas Paramétricas

9.1. Distribución Uniforme

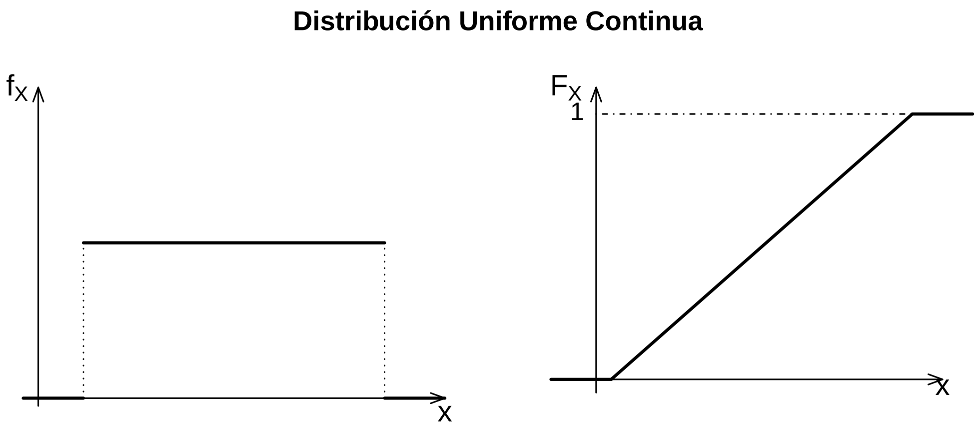
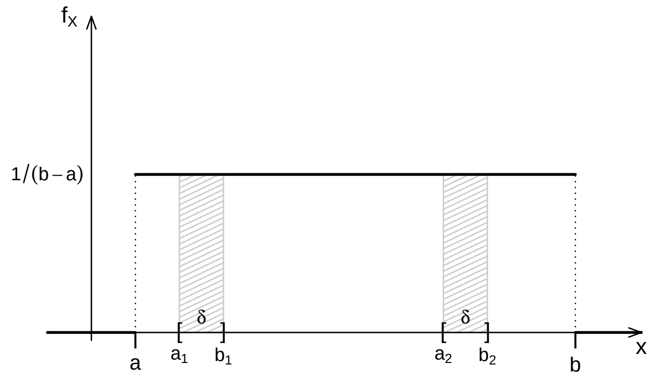


Figura 24: **Distribución uniforme**: Función de densidad de probabilidad f_X y correspondiente función de probabilidad acumulada F_X.

"Los puntos se localizan al azar". "Todos los puntos son igualmente posibles".

Sea U la variable aleatoria que sigue la **distribución uniforme (continua)**. La figura 25 muestra el caso de una *f. m. p.* y la correspondiente función de distribución.

Que todos los puntos sean igualmente posibles implica que intervalos de la misma longitud tienen asociado la misma probabilidad. Esto logra solamente si la *f. d. p.* es constante a lo largo del intervalo $[a, b]$. Luego, la función de densidad es $f(u) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(u)$.



La figura muestra los intervalos $[a_1, b_1]$ y $[a_2, b_2]$, ambos de longitud δ por lo que tienen asociada la misma probabilidad $\frac{\delta}{b-a}$.

Parámetros: $a < b \in \mathbb{R}$.

Soporte: $S_U = [a, b]$ o (a, b).

Función masa de probabilidad:

$$f_U(u) = \frac{1}{b-a}\mathbb{1}_{S_U}(u)$$

Función de probabilidad acumulada:

$$F_U(u) = \int_{-\infty}^{u} f_U(v)dv = \frac{1}{b-1}\int_{a]^u} du = \frac{u-a}{b-a}, \qquad a \leq u \leq b$$

Media:

$$\mathbb{E}[U] = \int_{\mathbb{R}} u f_U(u) du = \frac{1}{b-a} \int_a^b u du = \frac{a+b}{2}$$

Varianza:

$$\mathbb{E}[U^2] = \int_{\mathbb{R}} u^2 F_U(v) dv = \frac{1}{b-a} \int_a^b u^2 du = \frac{b^3 - a^3}{3(b-a)}$$
$$\text{var}(X) = \mathbb{E}[U^2] - \mathbb{E}^2[U] = \frac{(b-a)^2}{12}$$

Función generadora de momentos:

$$m_U(t) = \int_{\mathbb{R}} e^{tu} f_U(u) du = \frac{1}{b-a} \int_a^b e^{tu} du = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

La distribución uniforme se utiliza para modelar localización aleatoria *"al azar"*.

Ejercicio : Verifique que $\mathbb{E}[U] = (a+b)/2$, medianate el uso de la *f. g. m.*.

9.2. Distribución Exponencial

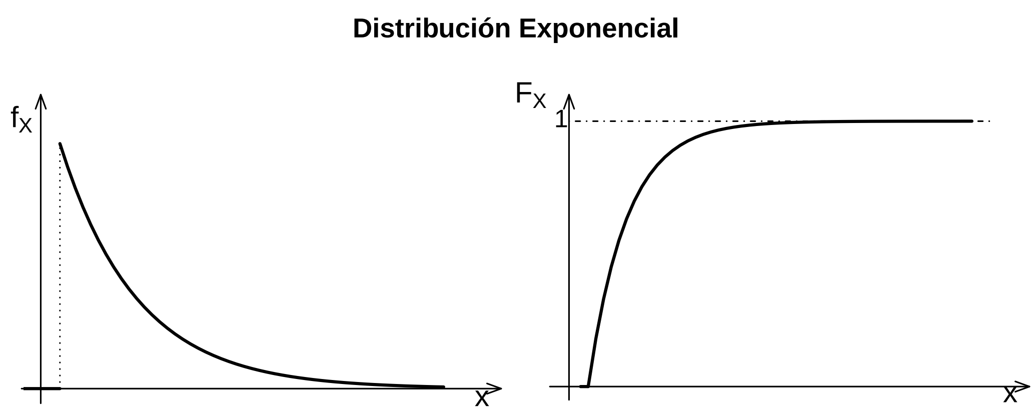


Figura 25: **Distribución exponencial**: Función de densidad de probabilidad f_X y correspondiente función de probabilidad acumulada F_X.

Sea T una variable aleatoria positiva continua con función de densidad $f(t) = \lambda e^{-\lambda t}$, para $t > 0$. T se dice que sigue la **distribución exponencial** con tasa λ o escala $\beta = 1/\lambda$, y se denota $T \sim \text{Exp}(\lambda)$ $\big(T \sim \text{Exp}(\beta)\big)$. La figura 25 muestra una función de densidad y la correspondiente función de distribución.

Parámetros: **tasa** $\lambda > 0$ o **escala** $\beta = 1/\lambda > 0$.

Soporte: $S_T = [0, \infty)$.

Función de densidad de probabilidad:

$$f_T(t) = \lambda e^{-\lambda t} \mathbb{1}_{S_T}(t) = \frac{1}{\beta} e^{-\frac{1}{\beta} t} \mathbb{1}_{S_T}(t)$$

Función de probabilidad acumulada:

$$F_T(t) = \int_{\infty}^{t} \lambda e^{-\lambda u} du = 1 - e^{-\lambda t}, \qquad t > 0$$

Función de sobrevivencia:

$$S_T(t) = \mathbb{P}(T \geq t) = e^{-\lambda t}, \qquad t > 0$$

Función generadora de momentos:

$$\begin{aligned} m_T(u) &= \int_{\mathbb{R}} e^{ut} f_T(u) du \\ &= \frac{\lambda}{\lambda - u} \int_0^{\infty} (\lambda - u) e^{-(\lambda-u)t} du \\ &= \frac{\lambda}{\lambda - u} \\ &= (1 - \beta u)^{-1} \end{aligned}$$

Pues la última integral es la de una densidad propia (Exp($\lambda - u$)) sobre todo su dominio. La última igualdad se sigue de que el parámetro de tasa λ es el recíproco del parámetro de escala β.

Media:

$$\mathbb{E}[T] = m'_T(0) = \frac{1}{\lambda} = \beta$$

Varianza:

$$\text{var}[T] = m''_T(0) - \left(m'_T(0)\right)^2 = \frac{1}{\lambda^2} = \beta^2$$

Proposición : La distribución exponencial *"no tiene memoria"*. Esto es, si $t_1, t_2 > 0$, se tiene

$$\mathbb{P}(T > t_1 + t_2 | T > t_1) = \mathbb{P}(T > t_2) \tag{3}$$

Demostración:

$$\begin{aligned} \mathbb{P}(T > t_1 + t_2 | T > t_1) &= \frac{\mathbb{P}(T > t_1 + t_2, T > t_1(}{\mathbb{P}(T > t_1)} \\ &= \frac{\mathbb{P}(T > t_1 + t_2)}{\mathbb{P}(T > t_1)} \\ &= \frac{e^{-\lambda(t1+t_2)}}{e^{-\lambda t_1}} \\ &= e^{-\lambda t_2} \\ &= \mathbb{P}(T > t_2) \end{aligned}$$

Proposición : (Tomada de Hoel et al. (1971)) Sea X una variable aleatoria continua que satisface que para todo $a, b \geq 0$, se tiene que

$$\mathbb{P}(X > a + b) = \mathbb{P}(X > a)\mathbb{P}(X > b)$$

Entonces, $\mathbb{P}(X > 0) = 0$ ó X sigue una distribución exponencial.

Demostración: Si $\mathbb{P}(X > 0) = 0$, entonces (9.2) se cumple trivialmente.

Suponga que (9.2) se cumple pero $\mathbb{P}(X > 0) > 0$.

Si $a = b = 0$, $\mathbb{P}(X > 0) = \mathbb{P}(X > 0)\mathbb{P}(X > 0) \Rightarrow \mathbb{P}(X > 0) = 1$. Esto es, X es una variable aleatoria positiva.

Sea ahora F la *f. p. a.* de X y $G = 1 - F$ su complemento. Luego, $G(x) = 1 - F(x) = \mathbb{P}(X > x)$ es una función no creciente y continua por la derecha.

$G(0) = 1 - F(0) = 1$, $G(+\infty) = 1 - F(+\infty) = 0$. De (9.2) se sigue $G(a+b) = G(a)G(b)$ para $a, b > 0$. Entonces, para $c > 0$, $m, n \in \mathbb{N}$, $G(nc) = \underbrace{G(c) \cdots G(c)}_{n} = \big[G(c)\big]^n$.

$$\begin{aligned} G(nc) &= \underbrace{G(c) \cdots G(c)}_{n} = \big[G(c)\big]^n \qquad (4) \\ G(c) &= \left[G\left(\frac{c}{m}\right) \cdots G\left(\frac{c}{m}\right)\right] = \left[G\left(\frac{c}{m}\right)\right]^m \end{aligned}$$

Además $0 < G(1) < 1$, pues si $G(1) = 1 \Rightarrow G(n) = \big[G(1)\big]^n = 1$, entonces $G(n) \to G(\infty) = 1$, que contradice el hecho que $G(\infty) = 0$. Si $G(1) = 0$, entonces $G\left(\frac{1}{m}\right) = \big[G(1)\big]^{1/m} = 0 \Rightarrow G(0) = 0$, por continuidad por la derecha y que contradice el hecho que $G(0) = 1$. Entonces, $0 < G(1) < 1$. Sea entonces $G(1) = e^{-\lambda}$, para algún $\lambda > 0$. Se sigue de (4) que $G\left(\frac{1}{m}\right) = \big[G(1)\big]^{\frac{1}{m}} = e^{-\frac{1}{m}\lambda}$ y $G\left(\frac{n}{m}\right) = \left[G\left(\frac{1}{m}\right)\right]^n = e^{-\frac{n}{m}\lambda}$. Luego, $G(q) = e^{-\lambda q}$, para todo $q \in ||$ y se concluye que $G(x) = e^{\lambda x}$, para todo $x \in \mathbb{R}^+$, por continuidad por la derecha. Por lo tanto, $F(x) = 1 - G(x) = 1 - e^{\lambda x}$ y $X \sim \text{Exp}(\lambda)$.

Los dos resultados anteriores se resumen en el siguiente teorema:

Teorema : Sea X una variable aleatoria positiva continua. X tiene la propiedad *sin memoria* (3) si y solo si X se distribuye exponencialmente.

Relación de la distribución Poisson y la distribución Exponencial

Ejemplo : Suponga que el número de automóviles que van a un exceso de velocidad son 8.4 autos/hora, sigue una distribución Poisson.

a). Determine la probabilidad de que transcurran al menos 10 minutos entre automóviles a alta velocidad.

b). Calcule la probabilidad de que transcurran más de 10 minutos si ya han transcurrido más de 5.

Solución: Sea N_t la *v. a.* que denota el número de automóviles a exceso de velocidad en $[0, t]$. Luego

$$N_t \sim \text{Po}(\lambda_t), \qquad \lambda_1 = \frac{8.4}{60} = 0.14\frac{\text{autos}}{\text{min}}, \qquad \lambda_t = \lambda_1 t$$

a). Que transcurran al menos 10 minutos entre autos con alta velocidad se interpreta como que en un intervalo de 10 minutos no haya automóviles a tal velocidad. Así,

$$\mathbb{P}(N_{10} = 0) = \lambda_{10}^0 \frac{e^{-\lambda_{10}}}{0!} = e^{\lambda_1(10)} = 0.2466$$

b). Sea T el tiempo entre autos con exceso de velocidad.

$$\mathbb{P}(T \leq 10) = \mathbb{P}(N_{10} > 10) = 1 - e^{\lambda_1(10)} = F_T(10) = 0.7534$$

Por lo tanto, $T \sim \text{Exp}(\lambda_1)$ y se sigue de la propiedad sin memoria de la distribución exponencial

$$\mathbb{P}(T > 10 | T > 5) = P(T > 5) = e^{-\lambda_1(5)} = e^{-0.14(5)} = 0.4965$$

9.3. Distribución Gamma

9.3.1. La función matemática Gamma

Proposición Para todo $u > 0$, se tiene que

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}uy^2} dy = \frac{1}{\sqrt{u}}$$

Demostración. Sea $I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}uy^2} dy$. Entonces,

$$\begin{aligned}
I^2 &= \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}ux^2} dx\right] \cdot \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}uy^2} dy\right] \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}ux^2} \cdot e^{-\frac{1}{2}uy^2} dx\, dy \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{u}{2}(x^2+y^2)} dx\, dy \\
&= \frac{1}{2\pi} \int_{0}^{2\pi} \int_{0}^{\infty} re^{-\frac{u}{2}r^2} dr\ d\theta \\
&= \int_{0}^{\infty} re^{-\frac{u}{2}r^2} dr \\
&= -\frac{1}{u}\ e^{-\frac{u}{2}v}\Big|_0^{\infty} \\
&= \frac{1}{u}
\end{aligned}$$

donde se ha empleado la transformación a *coordenadas polares*, a saber, $x = r\cos\theta$ y $y = r\sin\theta$, y $v = r^2$. Se concluye entonces que

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}uy^2} dy = \frac{1}{\sqrt{u}}$$

Proposición

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy = \sqrt{2\pi}$$

Demostración. Se sigue de la proposición anterior para $u = 1$.

Definición Se define la **función (matemática) gamma** por

$$\Gamma(x) = \int_{0}^{\infty} u^{x-1} e^{-u} du \tag{5}$$

para todo $x \in \mathbb{R} \backslash \{0, -1, -2, \ldots\}$.

La figura 26 muestra la forma de la función gamma.

Propiedades La función gamma cumple las siguientes propiedades:

1. $\Gamma(x+1) = x\Gamma(x)$ para todo $x \notin \mathbb{Z}^-$.
2. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
3. $\Gamma(1) = 1$

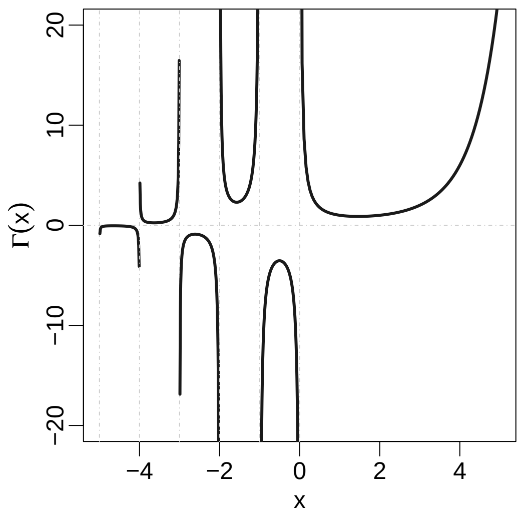


Figura 26: Función matemática gamma.

4. $\Gamma(n+1) = n!, \;$ para todo $n \in \mathbb{N}$
5. $0! = 1$
6. $\Gamma\left(\frac{n+1}{2}\right) = \frac{1}{2^{\frac{n+1}{2}}} \int_0^\infty v^n e^{-\frac{1}{2}v^2} dv$

Demostración.

1. Se sigue de aplicar un paso de la integración por partes.
2. En la definición de la función $\Gamma(1/2)$ utilice el cambio de variable $u \equiv y^2$ y aplique la primera proposición.
3. Se obtiene mediante cálculo directo de la integral.
4. Se sigue de las propiedades 1 y 3 para n entero positivo.
5. Consecuencia de las dos propiedades anteriores.
6. Se sigue de la aplicación sucesiva de n pasos en la integración por partes de $\Gamma\left(\frac{n+1}{2}\right)$.

9.3.2. La distribución Gamma

Definición Sea Y la variable aleatoria continua positiva con función de densidad dada por

$$f_Y(y) = \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y} \mathbb{1}_{\mathbb{R}^+}(y) \tag{6}$$

Se dice entonces que Y sigue una **distribución gamma con parámetro de forma** $\alpha(>0)$ **y parámetro tasa** $\lambda(>0)$, y se denota por $Y \sim \text{Ga}(\alpha, \lambda)$.

La figura 27 muestra el caso de una *f. d. p.* correspondiente función de distribución gamma.

Proposición La función de densidad anterior (6) es propia o legítima ya que para $\alpha > 0$ y $\lambda > 0$ fijas,

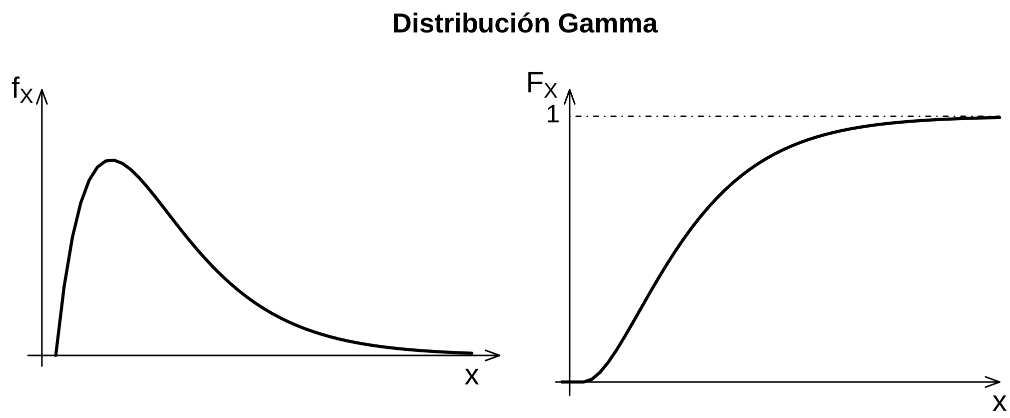


Figura 27: **Distribución gamma**: Función de densidad de probabilidad f_X y correspondiente función de probabilidad acumulada F_X.

i) $f_Y(y) = \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y} \geq 0$, para toda $y \in \mathbb{R}$.

ii) $\int_{\mathbb{R}} f_Y(y)dy = \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y} dy = 1$.

Para mostrar la igualdad anterior defina $u = \lambda y$ y utilice la definición de $\Gamma(\alpha)$.

Definición La constante $\beta = 1/\lambda$ se conoce como **parámetro de escala** de la distribución gamma. Luego, la densidad (6) se puede escribir también como

$$f_Y(y) = \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-\frac{1}{\beta} y} \mathbb{1}_{\mathbb{R}^+}(y) \tag{7}$$

Proposición : Sea Y variable aleatoria que sigue una distribución gamma con parámetro de forma α y de escala β, $Y \sim \text{Gamma}(\alpha, \beta)$. Entonces, para $c > 0$ fijo, $X = cY \sim \text{Gamma}(\alpha, c\beta)$.

Demostración: Considere $c > 0$ y sea $X = cY$. Las funciones F_Y, F_X, f_Y, f_X las correspondientes *f. p. a.* y *f. d. p.* de Y y X respectivamente. Entonces, para $x > 0$ se tiene

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(cY \leq x) = \mathbb{P}(Y \leq x/c) = F_Y(x/c)$$

Luego,

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) \\ &= F_Y'(x/c)\frac{1}{c} \\ &= \frac{1}{c} f_X(x/c) \\ &= \frac{1}{c} \frac{(x/c)^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{1}{\beta}(x/c)} \\ &= \frac{x^{\alpha-1}}{(c\beta)^\alpha \Gamma(\alpha)} e^{-\frac{1}{c\beta} x} \end{aligned}$$

que corresponde a la *f. d. p.* de una $\text{Gamma}(\alpha, c\beta)$. La proposición justifica a β como parámetro de escala.

Corolario : Sea $X \sim \text{Ga}(\alpha, 1)$, entonces $Y = \beta X \sim \text{Ga}(\alpha, \beta)$.

Proposición Sea Y una variable aleatoria (*v. a.*) distribuida $\text{Ga}(\alpha, \lambda)$ con función de densidad de probabilidad dada por la expresión (6). Entonces, para $r > -\alpha$,

$$\mathbb{E}[Y^r] = \frac{\Gamma(\alpha + r)}{\Gamma(\alpha)} \frac{1}{\lambda^r} = (\alpha + r - 1)_r \beta^r$$

donde $(\alpha + r - 1)_r = (\alpha + r - 1)(\alpha + r - 2) \cdots (\alpha)$.

Demostración Sea Y con *f. d. p.* dada por la expresión (6). Entonces,

$$\begin{aligned}
\mathbb{E}[Y^r] &= \int_0^\infty y^r \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y} dy \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(r+\alpha)}{\lambda^{r+\alpha}} \int_0^\infty \frac{\lambda^{r+\alpha}}{\Gamma(r+\alpha)} y^{(r+\alpha)-1} e^{-\lambda y} dy \\
&= \frac{\Gamma(\alpha + r)}{\Gamma(\alpha)} \cdot \frac{\lambda^\alpha}{\lambda^{r+\alpha}} \cdot 1 \\
&= \frac{\Gamma(\alpha + r)}{\Gamma(\alpha)} \frac{1}{\lambda^r}
\end{aligned}$$

válido para $(\alpha + r) > 0$. Note finalmente que $\Gamma(\alpha + r) = (\alpha + r - 1) \cdots \alpha \cdot \Gamma(\alpha) = (\alpha + r - 1)_r$ y $\beta = 1/\lambda$.

Proposición Sea Y una variable aleatoria distribuida gamma con parámetro de forma α, parámetro tasa λ y parámetro de escala $\beta = 1/\lambda$. Entonces,

$$\begin{aligned}
\mathbb{E}[Y] &= \alpha/\lambda = \alpha\beta \\
\text{var}[Y] &= \alpha/\lambda^2 = \alpha\beta^2
\end{aligned}$$

Demostración Considere la proposición anterior para $r = 1, 2$ y el hecho que $\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X]$.

Corolario : Sea $Y \sim \text{Ga}(\alpha, \lambda)$, tal que $\mathbb{E}[Y] = \alpha/\lambda$. Entonces $X = \lambda Y \sim \text{Ga}(\alpha, 1)$.

Proposición Sea Y una variable aleatoria distribuida $\text{Ga}(\alpha, \lambda)$ con función de densidad de probabilidad dada por la expresión (6). Entonces, Y tiene función generadora de momentos dada por

$$m_Y(t) = (1 - t/\lambda)^{-\alpha}, \quad t < \lambda$$

Demostración: Sea $Y \sim \text{Gamma}(\alpha, \lambda)$. Entonces,

$$\begin{aligned}
m_Y(t) = \mathbb{E}[e^{tY}] &\stackrel{\text{TEI}}{=} \int_0^\infty e^{ty} \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y} dy \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{1}{K} \int_0^\infty K y^{\alpha-1} e^{-(\lambda-t)y} dy \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\lambda + t)^\alpha} \\
&= \frac{1}{(1 - t/\lambda)^\alpha} \\
&= (1 - t/\lambda)^{-\alpha}
\end{aligned}$$

siempre que $(\lambda - t) > 0$, o equivalentemente, $t < \lambda$.

Corolario : Sea $Y \sim \text{Gamma}(\alpha, \beta)$, con $\mathbb{E}[Y] = \alpha\beta$, entonces la *f. g. m.* de Y está dada por $m_Y(t) = (1 - \beta t)^{-\alpha}$, para $t < 1/\beta$.

9.3.3. La distribución exponencial.

Definición : Sea T una variable aleatoria que sigue una distribución Gamma con parámetro de forma $\alpha = 1$ y parámetro tasa λ. En este caso, la *v. a.* T tiene una *f. d. p.* dada por

$$f_T(t) = \lambda e^{-\lambda t} \mathbb{1}_{(0,\infty)}(t)$$

y se dice que sigue una **distribución exponencial** con media $1/\lambda$. Esto es, $T \sim \text{Exp}(\lambda) \equiv \text{Gamma}(1, \lambda)$.

Corolario Sea $T \sim \text{Exp}(\lambda)$. Entonces $\mathbb{E}[T] = 1/\lambda$, $\text{var}(Y) = 1/\lambda^2$ y *f. g. m.* $m_Y(t) = (1 - t/\lambda)^{-1}$, para $t < \lambda$.

9.3.4. La distribución Ji-cuadrada χ^2.

Definición : El caso particular de la distribución Gamma con parámetro de forma $\alpha = n/2$, para algún $n \in \mathbb{N}$ y de escala $\beta = 2 = 1/\lambda$ se define como la **distribución Ji-cuadarada con n grados de libertad** y se denota por χ^2_n.

Corolario Sea $Y \sim \chi^2_n$. Entonces $\mathbb{E}[Y] = n$ y $\text{var}(Y) = 2n$ y *f. g. m.* $m_Y(t) = (1-2t)^{-n/2}$, para $t < \lambda$.

En algunos textos se incluyen tablas de probabilidad para la distribución Ji-cuadrada entre otras. La figura 28 muestra un extracto tomado de **?** (**?**).

9.4. Distribuciones Normal

Considere la función $h(y) = e^{-\frac{1}{2}y^2}$, para todo $y \in \mathbb{R}$. Se ha visto ya que la función es integrable sobre los reales y que

$$\sqrt{2\pi} = \int_{\mathbb{R}} e^{-\frac{1}{2}y^2} dy$$

por lo que $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \mathbb{1}_{\mathbb{R}}(y)$ define una función de densidad propia.

Definición : Sea Z una variable aleatoria con función de densidad dada por

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \mathbb{1}_{\mathbb{R}}(z)$$

Z se dice que sigue una **distribución normal estándar** y se denota por $Z \sim \text{N}(0, 1)$.

La *f. d. p.* de la distribución normal estándar se acostumbra a denotar por ϕ y su correspondiente *f. p. a.* por Φ. Así, para $z \in \mathbb{R}$,

$$\Phi(z) = \int_{-\infty}^{z} \phi(u) du = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$

Sea ahora $\mu \in \mathbb{R}$, $\sigma > 0$ y defina $X = \mu + \sigma Z$. Por determinar F_X y f_X, la *f. p. a.* y *f. d. p.* de X, respectivamente. Sea $x \in \mathbb{R}$, luego

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\mu + \sigma Z \leq x) = \mathbb{P}\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

$$f_X(x) = \frac{d}{dx} F_X(x) = \Phi'\left(\frac{x - \mu}{\sigma}\right) \frac{1}{\sigma} = \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Barrios et al. Tablas de Probabilidades 31

7. Distribución χ^2 Ji-Cuadrada

$$Y \sim \chi^2_n$$

siendo n los grados de libertad.

$$p = P(Y \leq y) = \int_0^y f_Y(u)du = 1 - \alpha$$

donde, para $u \geq 0$,

$$f_Y(u) = \frac{1}{2^{n/2}\Gamma(n/2)} u^{n/2-1} e^{-u/2}$$

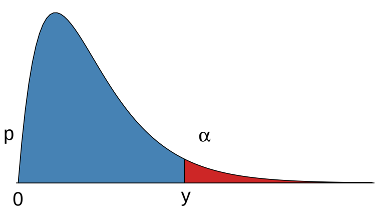


Tabla 7. Valores críticos $\chi^2_{(\alpha;n)}$ de la distribución χ^2_n Ji-Cuadrada.

	p									
	0.005	0.01	0.025	0.05	0.1	0.90	0.95	0.975	0.99	0.995
	α									
n	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
31	14.458	15.655	17.539	19.281	21.434	41.422	44.985	48.232	52.191	55.003
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328
33	15.815	17.074	19.047	20.867	23.110	43.745	47.400	50.725	54.776	57.648
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061	58.964
35	17.192	18.509	20.569	22.465	24.797	46.059	49.802	53.203	57.342	60.275
36	17.887	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619	61.581
37	18.586	19.960	22.106	24.075	26.492	48.363	52.192	55.668	59.893	62.883
38	19.289	20.691	22.878	24.884	27.343	49.513	53.384	56.896	61.162	64.181
39	19.996	21.426	23.654	25.695	28.196	50.660	54.572	58.120	62.428	65.476
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
75	47.206	49.475	52.942	56.054	59.795	91.061	96.217	100.839	106.393	110.286
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

ITAM Departamento de Estadística v.1.07

Figura 28: Extracto de las tablas de probabilidad de la distribución Ji-cuadrada (χ^2).

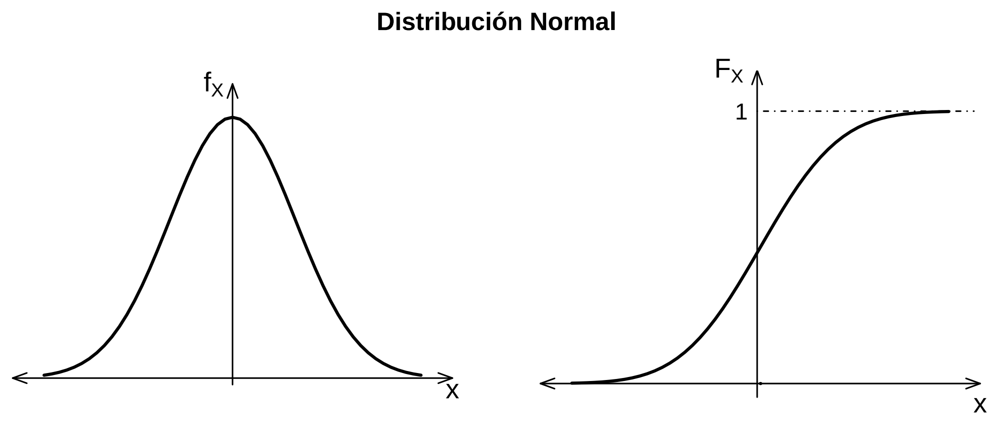


Figura 29: **Distribución normal**: Función de densidad de probabilidad f_X y correspondiente función de probabilidad acumulada F_X.

Así, f_X es una *f. d. p.* propia. La figura 29 muestra la *f. d. p.* y la correspondiente *f. p. a.* de una distribución normal.

Definición : Se dice que la variable aleatotia X sigue la **distribución normal** con parámetros μ y σ^2 y se denota por $X \sim \mathrm{N}(\mu, \sigma^2)$.

Parámetros: $\mu \in \mathbb{R}$, $\sigma > 0$.

Soporte: $S_X = \mathbb{R}$.

Función de densidad de probabilidad:

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

Considere nuevamente la variable aleatoria Z distribuida normal estándar. Por determinar su función generadora de momentos $m_Z(t) = \mathbb{E}[e^{tZ}]$.

$$\begin{aligned}
m_Z(t) &\overset{\text{TEI}}{=} \int_{\mathbb{R}} e^{tz}\phi(z)dz \\
&= \int_{\mathbb{R}} \frac{e^{tz}}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}z^2} dz \\
&= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2-2tz)} dx \\
&= e^{\frac{1}{2}t^2} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2} dz \qquad \text{completando cuadrados} \\
&= e^{\frac{1}{2}t^2}
\end{aligned}$$

pues la última integral es 1 por ser la integral sobre todo su dominio de una función de densidad $\mathrm{N}(t, 1)$.

Se sigue que

Media:

$$\mathbb{E}[Z] = m'_Z(0) t e^{\frac{1}{2}t^2}\Big|_{t=0} = 0$$

Varianza:

$$\operatorname{var}(Z) = m''_Z(0) - \left(m'_Z(0)\right)^2 = e^{\frac{1}{2}t^2}[1+t^2]\Big|_{t=10} = 1$$

Por lo que Z es una variable aleatoria estandarizada al tener media 0 y varianza 1, de ahí su nombre de *normal estándar*.

Proposición : Sea $Z \sim \mathrm{N}(0,1)$. Entonces:

a). $\mathbb{E}[Z] = 0$

b). $\mathrm{var}(Z) = 1$

c). $m_Z(t) = e^{\frac{1}{2}t^2}$.

Proposición : Sea $X \sim \mathrm{N}(\mu, \sigma^2)$. Entonces:

a). $\mathbb{E}[X] = \mu$

b). $\mathrm{var}(X) = \sigma^2$

c). $m_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$.

Demostración: La proposición es consecuencia de la proposición anterior al ver $X = \mu + \sigma Z$ y las propiedades de media, varianza y *f. g. m.*. Así, se refiere uno también a $X \sim \mathrm{N}(\mu, \sigma^2)$ como **distribución nomal de media μ y varianza σ^2**.

Proposición : Sea X una variable aleatoria distribuida normal de media μ y varianza σ^2. Entonces, su función de densidad de probabilidad es

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

tal que:

a). f_X es una distribución simétrica alrededor de μ.

b). La media, mediana y moda se localizan en mu.

c). f_X tiene dos puntos de inflexión en $\mu \pm \sigma$.

La distribución normal, aunque no con ese nombre fue empleada por Laplace y Gauss. Su importancia es teórica y práctica y resulta fundamental en el desarrollo de la Probabilidad y la Estadística.

El cálculo de la probabilidad acumulada $\Phi(Z)$ implica necesariamente métodos numéricos al no tener la integral

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2} du$$

una expresión analítica cerrada, por lo que se construyeron *"tablas de probabilidad"* y que cada vez están en más desuso por la disponibilidad de computadoras y aplicaciones como MATLAB, PQRS, R, PYTHON, etcétera.

La figura 30 muestra unas tablas de la probabilidad acumulada para la parte negativa de la distribución normal estándar. Si $z > 0$, por simetría de la distribución, $\Phi(z) = 1 - \Phi(-z)$, por lo que las tablas cubren toda la distribución.

Sea $Z \sim \mathrm{N}(0,1)$. De tablas se sigue que:

a). $\mathbb{P}(Z \leq -1.54) = \Phi(-1.54) = 0.0618$.

b). $\mathbb{P}(Z \leq -3) = \Phi(-3) = 0.0013 = \mathbb{P}(Z \geq 3)$.

c). $\mathbb{P}(Z \leq -2) = \Phi(-2) = 0.0228 = \mathbb{P}(Z \geq 2)$.

d). $\mathbb{P}(Z \leq -1) = \Phi(-1) = 0.1587 = \mathbb{P}(Z \geq 1)$.

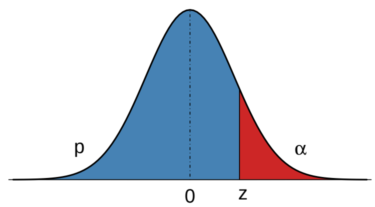
3. Distribución Normal Estándar

$$Z \sim \text{Normal}(0,1)$$

$$p = P(Z \leq z) = \Phi(z) = \int_{-\infty}^{z} \phi(u)du = 1 - \alpha$$

donde

$$\phi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$$



Nota: Si $X \sim \text{N}(\mu, \sigma^2)$, entonces $Z = (X-\mu)/\sigma \sim \text{N}(0,1)$. Luego,

$$P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Tabla 3A. Probabilidades acumuladas p de la distribución normal estándar.

z	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00
−3.4	0.0002	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
−3.3	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005	0.0005	0.0005
−3.2	0.0005	0.0005	0.0005	0.0006	0.0006	0.0006	0.0006	0.0006	0.0007	0.0007
−3.1	0.0007	0.0007	0.0008	0.0008	0.0008	0.0008	0.0009	0.0009	0.0009	0.0010
−3.0	0.0010	0.0010	0.0011	0.0011	0.0011	0.0012	0.0012	0.0013	0.0013	0.0013
−2.9	0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019
−2.8	0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026
−2.7	0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035
−2.6	0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047
−2.5	0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062
−2.4	0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082
−2.3	0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107
−2.2	0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139
−2.1	0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179
−2.0	0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228
−1.9	0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287
−1.8	0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359
−1.7	0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446
−1.6	0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548
−1.5	0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668
−1.4	0.0681	0.0694	0.0708	0.0721	0.0735	0.0749	0.0764	0.0778	0.0793	0.0808
−1.3	0.0823	0.0838	0.0853	0.0869	0.0885	0.0901	0.0918	0.0934	0.0951	0.0968
−1.2	0.0985	0.1003	0.1020	0.1038	0.1056	0.1075	0.1093	0.1112	0.1131	0.1151
−1.1	0.1170	0.1190	0.1210	0.1230	0.1251	0.1271	0.1292	0.1314	0.1335	0.1357
−1.0	0.1379	0.1401	0.1423	0.1446	0.1469	0.1492	0.1515	0.1539	0.1562	0.1587
−0.9	0.1611	0.1635	0.1660	0.1685	0.1711	0.1736	0.1762	0.1788	0.1814	0.1841
−0.8	0.1867	0.1894	0.1922	0.1949	0.1977	0.2005	0.2033	0.2061	0.2090	0.2119
−0.7	0.2148	0.2177	0.2206	0.2236	0.2266	0.2296	0.2327	0.2358	0.2389	0.2420
−0.6	0.2451	0.2483	0.2514	0.2546	0.2578	0.2611	0.2643	0.2676	0.2709	0.2743
−0.5	0.2776	0.2810	0.2843	0.2877	0.2912	0.2946	0.2981	0.3015	0.3050	0.3085
−0.4	0.3121	0.3156	0.3192	0.3228	0.3264	0.3300	0.3336	0.3372	0.3409	0.3446
−0.3	0.3483	0.3520	0.3557	0.3594	0.3632	0.3669	0.3707	0.3745	0.3783	0.3821
−0.2	0.3859	0.3897	0.3936	0.3974	0.4013	0.4052	0.4090	0.4129	0.4168	0.4207
−0.1	0.4247	0.4286	0.4325	0.4364	0.4404	0.4443	0.4483	0.4522	0.4562	0.4602
−0.0	0.4641	0.4681	0.4721	0.4761	0.4801	0.4840	0.4880	0.4920	0.4960	0.5000

Figura 30: Extracto de las tablas de probabilidad de la distribución Normal Estándar N(0, 1)
.

Considere ahora $X\mathrm{N}(\mu, \sigma^2)$ y se desea determinar $\mathbb{P}(X \leq x)$. Entonces, se estandariza X y se lee de tablas de la normal estándar. A saber,

$$\mathbb{P}(X \leq x) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = \mathbb{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Por ejemplo,

$$\begin{aligned}
\mathbb{P}(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &= \mathbb{P}\left(-3 \leq \frac{X-\mu}{\sigma} \leq 3\right) \\
&= \mathbb{P}(-3 \leq Z \leq 3) \\
&= \Phi(3) - \Phi(-3) \\
&= 1 - 2\Phi(-3) \\
&= 0.9973 \\
\mathbb{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &= 1 - 2\Phi(-2) = 0.9545 \\
\mathbb{P}(\mu - \sigma \leq X \leq \mu + \sigma) &= 1 - 2\Phi(-1) = 0.6827
\end{aligned}$$

Aproximación a la distribución binomial mediante la normal

Al igual que en el caso de la aproximación a la distribución binomial por medio de la de Poisson, así bajo ciertas condiciones la aproximación por medio de la distribución resulta razonable. La justificación teórica del resultado es el Teorema Central del Límite)TCL), que se discute en el curso siguiente.

Algunos texto sugieren que si $X \sim \mathrm{Bin}(n, p)$, para que la aproximación sea razonable se debe tener que $np > 5$ y $n(1-p) > 5$. Otros, piden que se cumpla que $np - 3\sqrt{np(1-p)} > 0$ y $np + 3\sqrt{np(1-p)} < n$, lo que lleva a la condición $n > 9(1-p)/p$.

Si alguna de las condiciones anteriores se cumple, considere que si las distribuciones de X y $Y \sim \mathrm{N}(\mu, \sigma^2)$ se aproximan, resulta razonable que $np = \mathbb{E}[X] \approx \mathbb{E}[Y] = \mu$ y $np(1-p) = \mathrm{var}(X) \approx \mathrm{var}(Y) = \sigma^2$ por lo que, para $x = 0, 1, \ldots, n$,

$$\mathbb{P}(X \leq x) \approx \mathbb{P}(Y \leq x) = \Phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right)$$

Algunos libros de Estadística sugieren usar $\mathbb{P}(X \leq x) \approx \mathbb{P}(Y \leq x + \frac{1}{2})$ pues se está aproximando una distribución discreta mediante una continua. La constante $\frac{1}{2}$ le llaman **factor de corrección por continuidad**.

Sin embargo, como se comentó en el caso de la aproximación por la distribución Poisson, estas aproximaciones están a desuso por la disponibilidad de aplicaciones computacionales.

Tabla 1: Probabilidades acumuladas de $X \sim \mathrm{Bin}(30, 0.2)$ y $Y \sim \mathrm{N}(6, 4.8)$.

x	2	3	4	5	6	7	8	9
$\mathbb{P}(X \leq x)$	0.0442	0.1227	0.2552	0.4275	0.6070	0.7608	0.8713	0.9389
$\mathbb{P}(Y \leq x)$	0.0339	0.0855	0.1807	0.3240	0.5000	0.6760	0.8193	0.9145
$\mathbb{P}(Y \leq x + 1/2)$	0.0551	0.1269	0.2468	0.4097	0.5903	0.7532	0.8731	0.9449

Ejercicio : Para el ejemplo mostrado en la tabla, compare el valor exacto de $\mathbb{P}(3 \leq X \leq 6)$ con la aproximación por la distribución normal con y sin corrección por continuidad.

Aproximación a la distribución Poisson mediante la normal

Al igual que la aproximación a la distribución binomial mediante la normal. suponga que $N \sim \mathrm{Po}(\lambda)$, entonces, $\lambda = \mathbb{E}[N] \approx \mathbb{E}[Y] = \mu$ y $\lambda = \mathrm{var}(N) \approx \mathrm{var}(Y) = \sigma^2$, donde $Y \sim \mathrm{N}(\mu, \sigma^2)$. Así, para $n = 0, 1, 2, \ldots,$

$$\mathbb{P}(N \leq n) \approx \mathbb{P}(Y \leq n) = \Phi\left(\frac{x-\lambda}{\sqrt{\lambda}}\right)$$

Como regla empírica se emplea la aproximación si $\lambda > 5$. Nuevamente, la justificación teórica es el teorema central de límite (TCL). También este caso, al aproximar la distribución discreta Poisson por la distribución continua normal se puede emplear el factor de corrección por continuidad.

Tabla 2: Probabilidades acumuladas de $N \sim \mathrm{Po}(8)$ y $Y \sim \mathrm{N}(8, 8)$.

n	2	3	4	5	6	7	8	9
$\mathbb{P}(N \leq n)$	0.0138	0.0424	0.0996	0.1912	0.3134	0.4530	0.5925	0.7166
$\mathbb{P}(Y \leq n)$	0.0169	0.0385	0.0786	0.1444	0.2398	0.3618	0.5000	0.6382
$\mathbb{P}(Y \leq n + 1/2)$	0.0259	0.0558	0.1080	0.1884	0.2979	0.4298	0.5702	0.7021

Ejercicio : Para el ejemplo mostrado en la tabla, compare el valor exacto de $\mathbb{P}(3 \leq N \leq 6)$ con la aproximación por la distribución normal con y sin corrección por continuidad.

9.5. Distribución Cauchy

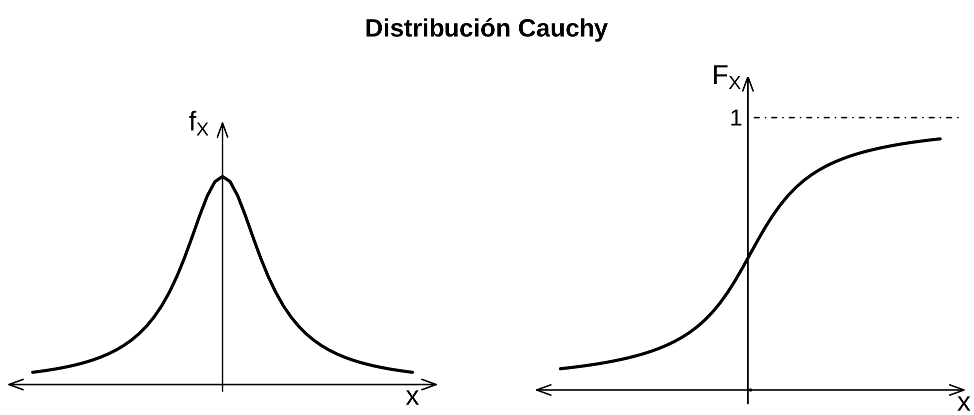


Figura 31: **Distribución Cauchy**: Función de densidad de probabilidad f_X y correspondiente función de probabilidad acumulada F_X.

Sea X una variable continua con función de densidad de probabilidad dada por $f(x) = \left\{\pi\beta\left[1 + \left(\frac{x-\alpha}{\beta}\right)^2\right]\right\}^{-1}$, para todo $x \in \mathbb{R}$. Se dice entonces que X sigue la **distribución Cauchy con parámetro de localización** α **y parámetro de escala** β y se denota por $X \sim \mathrm{Cauchy}(\alpha, \beta)$. La figura 31 muestra el caso de una *f. d. p.* y la correspondiente función de distribución.

Parámetros: localización $\alpha \in \mathbb{R}$, escala $\beta > 0$.

Soporte: $S_X = \mathbb{R}$.

Función de densidad de probabilidad:

$$f_X(x) = \frac{1}{\pi\beta\left[1 + \left(\frac{x-\alpha}{\beta}\right)^2\right]} \mathbb{1}_{\mathbb{R}}(x)$$

La función de densidad es propia, es una mera traslación y escalamiento de la Cauchy(0,1), vista ya anteriormente. Así también se mostró que la distribución no tienen ningún momento ni tampoco función generadora de momentos. Sin embargo, sí tiene función característica. A saber,

Función característica:

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \exp\left\{i\alpha t + \beta|t|\right\}, \quad t \in \mathbb{R}$$

Propiedades :

a). El parámetro de escala es tal que $\beta = q_3 - q_1$, donde los q_i denotan el tercer y primer cuartil respectivamente.

b). La función f_X es simétrica alrededor de α.

c). Ld densidad es de *"colas pesadas"*. Como ya se mencionó, la distribución no tiene momentos.

d). La distribución de Cauchy resulta del cociente de variables aleatorias normales independientes.

e). La distribución tiene aplicaciones prácticas en la Física y teóricas en la misma Teoría de la Probabilidad.

f). Si la distribución se centra en $\alpha = 0$, la función característica φ_X es real y simétrica.

9.6. La distribución Beta

9.6.1. La función Beta

La **función (matemática) Beta** se define para $\alpha_1 > 0,\ \alpha_2 > 0$, por

$$B(\alpha_1, \alpha_2) = \int_0^1 t^{\alpha_1 - 1}(1-t)^{\alpha_2 - 1}\, dt \tag{8}$$

y satisface la siguiente identidad

$$B(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}$$

donde $\Gamma(\cdot)$ denota la función matemática Gamma definida anteriormente dada por

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$$

De hecho con base a la función Beta se define la **distribución Beta**.

9.6.2. La distribución Beta

Definición La variable aleatoria X se dice que sigue una **distribución Beta con parámetros** α_1 **y** α_2, denotada como $X \sim B(x; \alpha_1, \alpha_2)$, si X tiene como función de densidad de probabilidad

$$f(x; \alpha_1, \alpha_2) = \frac{1}{B(\alpha_1, \alpha_2)}\, x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1}, \quad 0 \le x \le 1$$

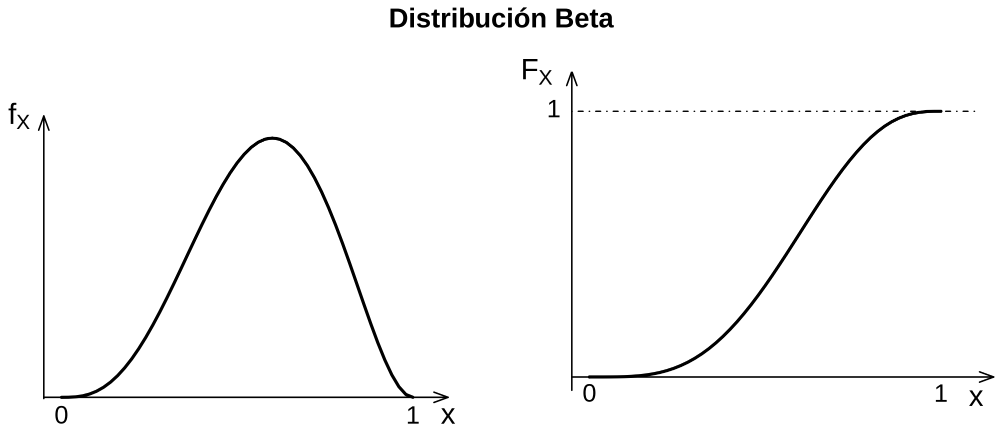


Figura 32: **Distribución Beta**: Función de densidad de probabilidad f_X y correspondiente función de probabilidad acumulada F_X.

con la función $B(\alpha_1, \alpha_2)$ definida en (8). La figura 32 muestra una función de densidad y la correspondiente función de distribución.

Parámetros: $\alpha_1 > 0$, $\alpha_2 > 0$.

Soporte: $S_X = [0, 1]$.

Función de densidad de probabilidad:

$$f_X(x) = \frac{1}{B(\alpha_1, \alpha_2)}\, x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1}\mathbb{1}_{S_X}(x)$$

Proposición : El valor esperado y varianza de $X \sim B(x; \alpha_1, \alpha_2)$ están dados por:

$$\mathbb{E}[X] = \frac{\alpha_1}{\alpha_1 + \alpha_2}, \qquad \text{var}(X) = \frac{\alpha_1\alpha_2}{(\alpha_1+\alpha_2)^2(\alpha_1+\alpha_2+1)}$$

Demostración:

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^1 x f(x; \alpha_1, \alpha_2) dx \\
&= \frac{1}{B(\alpha_1, \alpha_2)} \int_0^1 x^{(\alpha_1+1)-1}(1-x)^{\alpha_2} dx \\
&= \frac{B(\alpha_1+1, \alpha_2)}{B(\alpha_1, \alpha_2)} \\
&= \frac{\Gamma(\alpha_1+1)\Gamma(\alpha_2)/\Gamma(\alpha_1+1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)/\Gamma(\alpha_1+\alpha_2)} \\
&= \frac{\alpha_1}{\alpha_1+\alpha_2}
\end{aligned}$$

utilizando el hecho que $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$. Similarmente se puede ver que

$$\mathbb{E}[X^2] = \frac{(\alpha_1+1)\alpha_1}{(\alpha_1+\alpha_2+1)(\alpha_1+\alpha_2)}$$

de donde se sigue el resultado para $\text{var}(X)$.

Notas:

- Si $\alpha_1 = \alpha_2 > 0$ la distribución Beta es simétrica alrededor de $1/2$.
- Si $\alpha_1 = 1 = \alpha_2$, la distribución Beta resulta la distribución uniforme en el intervalo $(0,1)$.
- La distribución Beta, entre varias aplicaciones, se utiliza para modelar el valor de *probabilidades*.

9.7. Distribución Weibull

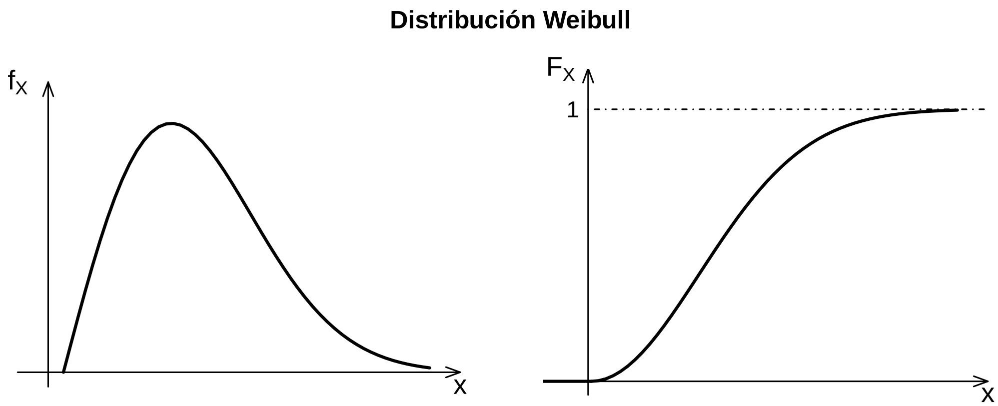


Figura 33: **Distribución Weibull**: Función de densidad de probabilidad f_X y correspondiente función de probabilidad acumulada F_X.

Sea W una variable aleatoria continua positiva con función de densidad

$$f(w) = \frac{\alpha}{\beta}\left(\frac{w}{\beta}\right)^{\alpha-1} e^{-\left(\frac{w}{\beta}\right)^{\alpha}} \qquad x > 0$$

con $\alpha, \beta > 0$ fijos. W se dice que sigue la **distribución Weibull con parámetro de forma** α **y parámetro de escala** β y se denota por $W \sim \text{Weibull}(\alpha, \beta)$. La figura 33 muestra una función de densidad y la correspondiente función de distribución.

Parámetros: $\alpha > 0$, $\beta > 0$.

Soporte: $(0, \infty)$.

Función de densidad de probabilidad:

$$f_W(w) = \frac{\alpha}{\beta}\left(\frac{w}{\beta}\right)^{\alpha-1} e^{-\left(\frac{w}{\beta}\right)^{\alpha}} \mathbb{1}_{(0,\infty)}(w)$$

Función de probabilidad acumulada:

$$F_W(w) = \mathbb{P}(W \le w) = \int_{-\infty}^{w} f_W(u)du = 1 - \exp\left\{-\left(\frac{w}{\beta}\right)^{\alpha}\right\}, \quad w \ge 0$$

Función de sobrevivencia:

$$S_W(w) = \mathbb{P}(W \ge w) = \exp\left\{-\left(\frac{w}{\beta}\right)^{\alpha}\right\}, \quad w \ge 0$$

Propiedades : Sea $W \sim \text{Weibull}(\alpha, \beta)$, entonces W tiene las siguientes propiedades:

a). $\mathbb{E}[W^r] = \beta^r \Gamma\left(1 + \frac{r}{\alpha}\right)$, $r > 0$

b). $\mathbb{E}[W] = \beta \Gamma\left(1 + \frac{1}{\alpha}\right)$

c). $\text{var}[W] = \beta^2 \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right]$

d). $m_W(t) = \displaystyle\sum_{k=1}^{\infty} \frac{t^k \alpha^k}{k!} \Gamma\left(1 + \frac{k}{\alpha}\right)$

e). La distribución de Weibull es empleada para modelar tiempos de vida, tiempos de falla, fatiga de materiales, etcétera.

Proposición : Sea $W \sim \text{Weibull}(\alpha, \beta)$.

a). Si $T \sim \text{Exp}(\lambda)$, con λ parámetro tasa, entonces $W = T^r \sim \text{Weibull}(\alpha = 1/r, \beta = \lambda^{-r})$.

b). La *función de riesgo* de W es

$$h_W(w) = -\frac{d}{dw} \log S_W(w) = \frac{w}{\beta}\left(\frac{w}{\beta}\right)^{\alpha-1}$$

por lo que la tasa de riesgo es h_W es:

- Creciente en el tiempo si $\alpha > 1$.
- Constante en el tiempo si $\alpha = 1$.
- Decreciente en el tiempo si $\alpha < 1$.

9.8. Distribución Lognormal

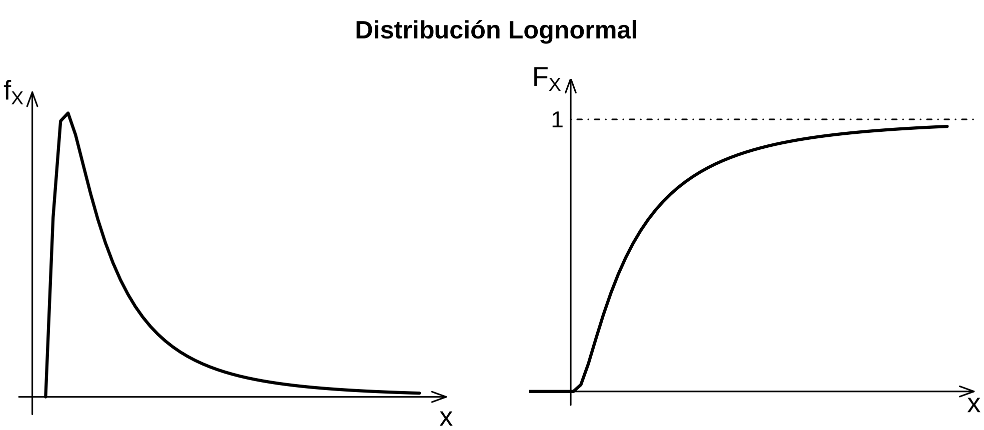


Figura 34: **Distribución Lognormal**: Función de densidad de probabilidad f_X y correspondiente función de probabilidad acumulada F_X.

Sea Y una variable aleatoria continua positiva con función de densidad dada por

$$f(y) = \frac{1}{y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln y - \mu}{\sigma}\right)^2}, \quad y > 0$$

Y se dice que sigue la **distribución lognormal de parámetros** μ **y** σ y se denota por $Y \sim \text{Lognormal}(\mu, \sigma)$. La figura 34 muestra una función de densidad y la correspondiente función de distribución.

Parámetros: $\mu \in \mathbb{R}$ y $\sigma > 0$.

Soporte: $S_Y = (0, \infty)$.

Función de densidad de probabilidad:

$$f_Y(y) = \frac{1}{y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln y - \mu}{\sigma}\right)^2} \mathbb{1}_{(0,\infty)}(y)$$

Proposición : Sea $X \sim \mathrm{N}(\mu, \sigma^2)$, entonces $Y = e^X \sim \mathrm{Lognormal}(\mu, \sigma)$.

Corolario : Sean Z distribuida normal estándar, $\mu \in \mathbb{R}$, $\sigma > 0$, entonces $Y = \exp\{\mu + \sigma Z\} \sim \mathrm{Lognormal}(\mu, \sigma)$.

Media: $\mathbb{E}[Y] = \exp\{\mu + \sigma^2/2\}$.

Varianza: $\mathrm{var}(Y) = \exp\{2\mu + 2\sigma^2\} - \exp\{2\mu + \sigma^2\}$.

Momentos: $\mathbb{E}[Y^r] = \exp\{r\mu + \frac{1}{2} r^2\sigma^2\}$.

Aplicaciones: La distribución *lognormal* tiene aplicaciones en la Biología, en la Física y la Actuaría, entre otras áreas. Se utiliza en la modelación de pesos, volúmenes, tiempos de vida, etcétera.

Ejercicio . Para $r \in \mathbb{N}$, muestre el resultado anterior para $\mathbb{E}[Y^r]$.

Proposición : Sea Y variable aleatoria que sigue una distribución lognormal. Entonces Y tiene todos sus momentos finitos pero no tiene función generadora de momentos $\mathbb{E}[e^{tY}]$ para $t > 0$.

9.9. Distribución Laplace o Doble Exponencial

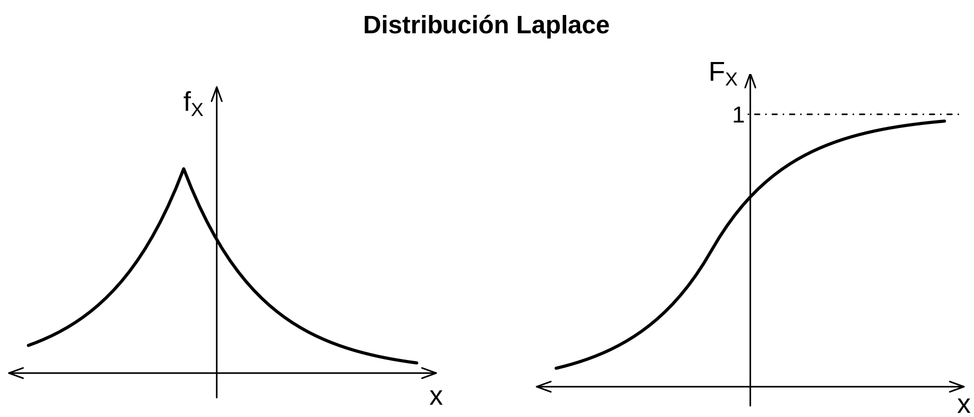


Figura 35: **Distribución Laplace o doble exponencial**: Función de densidad de probabilidad f_X y correspondiente función de probabilidad acumulada F_X.

Sea X una variable aleatoria continua con función de densidad dada por

$$f(x) = \frac{1}{2\beta} e^{-\frac{|x-\mu|}{\beta}}, \quad \text{para todo } x \in \mathbb{R}$$

X se dice que sigue la **distribución Laplace o doble exponencial con parámetro de localización** μ **y parámetro de escala** β y se denota por $X \sim \mathrm{DExp}(\mu, \beta)$. La figura 35 muestra una función de densidad y la correspondiente función de distribución.

Parámetros: localización $\mu \in \mathbb{R}$ y escala $\beta > 0$.

Soporte: $S_X = \mathbb{R}$.

Función de densidad de probabilidad:

$$f_X(x) = \frac{1}{2\beta} e^{-\frac{|x-\mu|}{\beta}} \mathbb{1}_{\mathbb{R}}$$

Media: $\mathbb{E}[X] = \mu$.

Varianza: $\mathrm{var}(X) = 2\beta^2$.

Función generadora de momentos:

$$m_X(t) = \frac{e^{\mu t}}{1-\beta^2 t^2}, \quad |t| < 1/\beta$$

Aplicaciones:

- Surge como la diferencia de variables aleatorias independientes distribuidas exponencialmente.
- Por mucho tiempo se empleó como la distribución del *error experimental*.
- Como la distribución del término de error en modelos de regresión.

Proposición : Sean Y_1 y Y_2 dos variables aleatorias independientes con distribución común exponencial de tasa λ, entonces $X = Y_1 - Y_2 \sim \mathrm{DExp}(\mu = 0, \beta = 1/\lambda)$.

9.10. Distribución Pareto

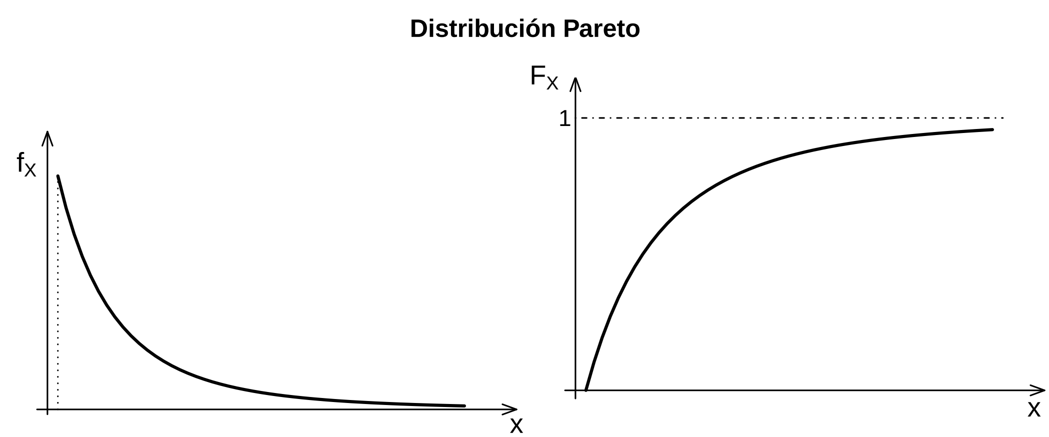


Figura 36: **Distribución Pareto**: Función de densidad de probabilidad f_X y correspondiente función de probabilidad acumulada F_X.

Sea W una variable aleatoria continua con función de densidad dada por

$$f(w) = \alpha\beta^\alpha w^{-(\alpha+1)}, \quad w > \beta$$

W se dice que sigue la **distribución Pareto con parámetro de forma** α **y parámetro de escala** β y se denota por $W \sim \mathrm{Pareto}(\alpha, \beta)$. La figura 36 muestra una función de densidad y la correspondiente función de distribución.

Parámetros: forma $\alpha > 0$ y escala $\beta > 0$.

Soporte: $S_W = (\beta, \infty)$.

Función de densidad de probabilidad:

$$f_W(w) = \alpha\beta^\alpha w^{-(\alpha+1)} \mathbb{1}_{(\beta,\infty)}(w)$$

Función de probabilidad acumulada:

$$F_W(w) = \mathbb{P}(W \leq w) = 1 - \left(\frac{\beta}{w}\right)^\alpha, \quad w \geq \beta$$

Media:

$$\mathbb{E}[W] = \begin{cases} \frac{\alpha\beta}{\alpha-1} & \text{si } \alpha > 1 \\ +\infty & \text{si } 0 < \alpha \leq 1 \end{cases}$$

Varianza:

$$\text{var}[W] = \begin{cases} \frac{\alpha\beta^2}{(\alpha-1)(\alpha-2)} & \text{si } \alpha > 2 \\ +\infty & \text{si } 0 < \alpha \leq 2 \end{cases}$$

Función generadora de momentos: No existe.

Aplicaciones: La distribución de Pareto se emplea para modelar la distribución del empleo.

Ejercicio : Verifique $\mathbb{E}[w]$ y $\text{var}(W)$.

9.11. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 9, Barrios and Heiras (2024).

Referencias

Ash, R. B. (1972). *Real Analysis and Probability*. New York, NY: Acaemic Press.

Ash, R. B. and C. A. Doléans-Dade (2000). *Real Analysis and Probability*. San Diego, CA: Harcourt/Academic Press.

Barrios, E. (2024). Apuntes para el curso de Cálculo de Probabilidades II. https://gente.itam.mx/ebarrios/docs/apuntes_CP2.pdf. (2 de enero de 2024).

Barrios, E. and M. Heiras (2024). Cálculo de Probabilidades I – Cuaderno de Ejercicios. https://gente.itam.mx/ebarrios/docs/CuadernoEjercicios_CP1.pdf. (2 de enero de 2024).

Bartle, R. G. (1966). *The Elements of Integration and Lebesgue Measur*. New York, NY: John Wiley & Sons.

Blitzstein, J. K. and J. Hwang (2014). *Intorduction to Probability*. Boca Raton, FL: CRC Press.

Grabinsky, G. (2000). *Teoría de la Medida*. DF, México: Faultad de Ciencias, UNAM.

Harris, B. (1966). *Theory of Probability*. Reading, MA.: Addison-Wedsley.

Hoel, P. G., S. C. Port, and C. J. Stone (1971). *Introduction to Probability Theory*. Boston: Houghton Miffling Company.

León-García, A. (2008). *Probability, Statistics and Random Processes for Electrical Engineering* (3 ed.). Reading, Massachusetts: Addison-Wesley.

Mood, A. M., F. A. Graybill, and D. C. Boes (1974). *Introduction to the Theory of Statistics* (3rd ed.). Singapore: McGraw-Hill.

Rincón, L. (2014). Introducción a la Probabilidad. https://filedn.com/lwDqCORMczI54WhtiTmXwGp/flip-proba1/mobile/index.html. (2 de enero de 2024).

Ross, S. (2006). *A First Course in Probability* (7th ed.). Upper Saddle River, N.J.: Prentice Hall.

Ross, S. (2010). *A First Course in Probability* (8th ed.). Upper Saddle River, N.J.: Prentice Hall.

Wackerly, D. D., W. Mendenhall III, and R. L. Scheaffer (2008). *Mathematical Statistics with Applications* (7 ed.). Australia: Thomson.