

Apuntes de Probabilidades y Estadística

Federico Yulita

Verano, 2023

[Esta materia](#) la cursé el primer cuatrimestre del 2019 con Matthieu Jonckheere como docente de las clases teóricas y Pablo Groisman como JTP. Si querés ver el resto de mis apuntes los podés encontrar en [mi blog](#).

Índice

| | | |
|-----------|------------------------------------------------------------|-----------|
| I | Probabilidades | 3 |
| 1. | Espacios y Probabilidades | 4 |
| 1.1. | Introducción | 4 |
| 1.2. | Probabilidad Condicional | 5 |
| 2. | Variables Aleatorias Discretas | 8 |
| 2.1. | Definiciones | 8 |
| 2.2. | Distribución de Bernoulli | 12 |
| 2.3. | Distribución Geométrica | 13 |
| 2.4. | Distribución Binomial | 15 |
| 2.5. | Distribución Binomial Negativa | 17 |
| 2.6. | Distribución de Poisson | 17 |
| 2.7. | Distribución Hipergeométrica | 20 |
| 2.8. | Distribución Empírica | 20 |
| 3. | Variables Aleatorias Continuas | 21 |
| 3.1. | Definiciones | 21 |
| 3.2. | Distribución Normal | 22 |
| 3.3. | Distribución Uniforme | 23 |
| 3.4. | Distribución Exponencial | 24 |
| 3.5. | Distribución Gamma | 26 |
| 3.6. | Distribución Chi Cuadrado | 27 |
| 3.7. | Distribución t-Student | 27 |
| 4. | Vectores Aleatorios | 28 |
| 4.1. | Distribuciones Marginales | 28 |
| 4.2. | Esperanza y Covarianza | 32 |
| 5. | Ley de Grandes Números y Teorema Central del Límite | 36 |
| 5.1. | Definiciones | 36 |
| 5.2. | Ley de Grandes Números | 37 |
| 5.3. | Función Generadora de Momentos | 38 |
| 5.4. | Teorema Central del Límite | 41 |
| II | Estadística | 45 |

| | |
|-------------------------------------------------|-----------|
| 6. Estadística Paramétrica | 46 |
| 6.1. Estimación Puntual | 46 |
| 6.1.1. Método de Momentos | 46 |
| 6.1.2. Método de Máxima Verosimilitud | 47 |
| 6.2. Intervalo de Confianza | 50 |
| 7. Tests de Hipótesis | 53 |
| 7.1. Tests Paramétricos | 53 |
| 7.2. Tests No Paramétricos | 56 |
| 7.2.1. Test de Adherencia | 56 |
| 7.2.2. Test de Independencia | 56 |
| 7.2.3. Test de Bondad de Ajuste | 57 |
| 8. Correlación y Regresión Lineal | 60 |

Parte I
Probabilidades



1. Espacios y Probabilidades

1.1. Introducción

Definimos al **Espacio Muestral** como un conjunto S que contiene todos los posibles resultados de un experimento. A cada subconjunto del espacio muestral lo llamamos **Evento**. Definimos el **Espacio de Eventos** como el conjunto \mathcal{F} que contiene todos los posibles eventos. Si el espacio muestral es numerable entonces $\mathcal{F} = \mathcal{P}(S)$, donde \mathcal{P} denomina el conjunto de partes.

Definimos la probabilidad como:

$$\mathbb{P} : \mathcal{F} \rightarrow [0, 1] / (\mathbb{P}(S) = 1) \wedge \left(\{A_n\}, n \in [1, N] \subseteq \mathbb{N} / A_i \cap A_j = \emptyset \forall i \neq j \implies \mathbb{P}\left(\bigcup_{n=1}^N A_n\right) = \sum_{n=1}^N \mathbb{P}(A_n) \right)$$

Con esto decimos que la probabilidad es una función que toma eventos y les asigna un valor entre 0 y 1 y que además tiene dos propiedades fundamentales:

1. La probabilidad de que suceda cualquier eventos del espacio muestral es 1 (o sea, si o si va a suceder un evento dentro del espacio muestral)
2. La probabilidad de que sucedan algunos de los eventos A_n ($n \in [1, N] \subseteq \mathbb{N}$) es igual a la suma de las probabilidades de cada evento siempre y cuando los eventos sean independientes ($A_i \cap A_j = \emptyset \forall i \neq j$).

Veamos unas propiedades de la probabilidad. Sean A y B eventos entonces:

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}((A \setminus B) \cup (A \cap B)) \\ (A \setminus B) \cap (A \cap B) &= \emptyset \rightarrow = \mathbb{P}(A \setminus B) + \underbrace{\mathbb{P}(A \cap B)}_{\geq 0} \\ &\geq \mathbb{P}(A \setminus B) \\ \therefore \mathbb{P}(A \setminus B) &= \mathbb{P}(A) - \mathbb{P}(A \cap B) \end{aligned}$$

En particular, notemos:

$$\begin{aligned} \mathbb{P}(A') &= \mathbb{P}(S \setminus A) \\ &= \mathbb{P}(S) - \mathbb{P}(S \cap A) \\ &= 1 - \mathbb{P}(A) \end{aligned}$$

Supongamos que tenemos una sucesión de eventos $\{A_i \subseteq S / A_{i+1} \cap A_i = A_i\}$, $i \in [1, n] \subseteq \mathbb{N}$. Definamos entonces a la sucesión $\{B_i\}_{i \leq n}$ tal que $B_i = A_i \setminus B_{i-1}$ y $B_1 = A_1$. Notemos que estos eventos no intersecan ($B_{i+1} \cap B_i = \emptyset \forall i$) y que además $\bigcup_{i=1}^n A_i = \bigcup_{i=1}^n B_i$. Entonces:

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &= \mathbb{P}\left(\bigcup_{i=1}^n B_i\right) \\ B_i \cap B_j &= \emptyset \forall i \neq j \rightarrow = \sum_{i=1}^n \mathbb{P}(B_i) \\ &= \sum_{i=1}^n \mathbb{P}(A_i \setminus B_{i-1}) \\ &\leq \sum_{i=1}^n \mathbb{P}(A_i) \\ \therefore \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &\leq \sum_{i=1}^n \mathbb{P}(A_i) \end{aligned}$$

Ahora veamos un ejemplo. Consideremos un experimento donde se tira un dado. El espacio muestral es $S = \{1, 2, 3, 4, 5, 6\}$, donde cada número representa el resultado del dado. En este caso como la moneda se tira una sola

vez el espacio de eventos es $\mathcal{F} = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$. Definamos entonces cada evento A_n en base al resultado n del dado. Notemos que $A_i \cap A_j = \emptyset$ ya que cada evento es independiente. Por lo tanto, $\cup_{n=1}^6 A_n = \mathcal{S}$. Entonces:

$$\begin{aligned} \mathbb{P}(A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6) &= \sum_{n=1}^6 \mathbb{P}(A_n) \\ &= \mathbb{P}(\mathcal{S}) \\ &= 1 \end{aligned}$$

Si asumimos que el dado está balanceado eso significa que cada evento es equiprobable. Por lo tanto, si consideramos algún evento A_m tenemos que:

$$\begin{aligned} \sum_{n=1}^6 \mathbb{P}(A_n) &= 6 \cdot \mathbb{P}(A_m) \\ &= 1 \end{aligned}$$

$$\therefore \mathbb{P}(A_m) = \frac{1}{6}$$

Es decir, la probabilidad de que salga algún número específico $m \in [1, 6]$ en el dado es de $1/6$. Esto es lo que esperábamos obtener.

1.2. Probabilidad Condicional

Sean A y B dos eventos los llamamos **Independientes** $\iff \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Ya mencioné eventos independientes antes y a lo que me refería es a esta propiedad. Cuando dos eventos son independientes notamos $A \perp B$. Sean A y B dos eventos llamamos la **Probabilidad Condicional** de A dado B ($A|B$) a la probabilidad de que el evento A suceda dado que sucede el evento B . Vale que:

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

Por ejemplo, si quiero saber cuál es la probabilidad de que al tirar el dado salga un 2 dado que previamente salió el 5 entonces:

$$\begin{aligned} \mathbb{P}(2|5) &= \frac{\mathbb{P}(2 \cap 5)}{\mathbb{P}(5)} \\ \text{son independientes} \rightarrow &= \frac{\mathbb{P}(2) \cancel{\mathbb{P}(5)}}{\cancel{\mathbb{P}(5)}} \\ &= \mathbb{P}(2) \\ &= \frac{1}{6} \end{aligned}$$

Sean A y B eventos independientes vale que $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Con estas propiedades podemos definir la **Fórmula de Probabilidad Total**. Sea $I = \{B_i\}$, $i \in [1, N]$ una partición del espacio muestral y A un evento entonces:

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

Demostremosla:

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap \mathcal{S}) \\ &= \mathbb{P}(A \cap (\cup_{i=1}^N B_i)) \\ &= \mathbb{P}(\cup_{i=1}^N (A \cap B_i)) \end{aligned}$$

Como $I = \{B_i\}$ es una partición de \mathcal{S} entonces $i \neq j \iff B_i \cap B_j = \emptyset$. Por lo tanto:

$$\begin{aligned}(A \cap B_i) \cap (A \cap B_j) &= (A \cap A) \cap \underbrace{(B_i \cap B_j)}_{=\emptyset} \\ &= \emptyset\end{aligned}$$

$$\therefore \mathbb{P}\left(\bigcup_{i=1}^N (A \cap B_i)\right) = \sum_{i=1}^N \mathbb{P}(A \cap B_i)$$

Esta fórmula es útil para calcular una probabilidad en función de las probabilidades de eventos que ya conocemos. Veamos un ejemplo. Consideremos que el país está viviendo una pandemia debida a un nuevo virus sumamente contagioso. Se hace una estadística y se encuentra que el 80 % de la población está vacunada y que de los vacunados el 2 % está enferma y de los no vacunados el 10 % está enferma. Es decir, $\mathbb{P}(V) = 0.8$, $\mathbb{P}(E|V) = 0.02$ y $\mathbb{P}(E|V') = 0.1$. Calculemos la probabilidad de tomar a una persona enferma del grupo estadístico usando la fórmula de probabilidad total:

$$\begin{aligned}\mathbb{P}(E) &= \mathbb{P}(E|V) \mathbb{P}(V) + \mathbb{P}(E|V') \mathbb{P}(V') \\ &= 0.02 \cdot 0.8 + 0.1 \cdot (1 - 0.8) \\ &= 0.016 + 0.02 \\ &= 0.036\end{aligned}$$

También, podemos definir la **Fórmula de Bayes**. Sea $I = \{B_i\}$, $i \in [1, N]$ una partición del espacio muestral y A un evento entonces:

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i) \mathbb{P}(B_i)}{\mathbb{P}(A)}$$

Veamos un ejemplo. Consideremos que estamos en un juego de televisión donde tenemos tres puertas. Detrás de una de las puertas hay un premio y detrás de las otras dos no hay nada. En este juego debemos elegir una puerta, luego de elegirla el conductor abre una de las puertas donde no hay nada y nos pregunta si queremos cambiar nuestra elección a la otra puerta o quedarnos con nuestra elección inicial. La pregunta es, ¿conviene cambiar nuestra elección? Sin perder generalidad consideremos que la primera puerta que elegimos es la primera puerta. La probabilidad de que esta sea la puerta ganadora es de $\mathbb{P}(B_1) = 1/3$. Luego, el conductor nos abre la puerta 3 para revelar que no hay nada detrás (llamemos a este evento A). Sabemos que $\mathbb{P}(B_3|A) = 0$, ya que detrás de la tercera puerta no hay nada. También sabemos que $\mathbb{P}(A|B_3) = 0$ ya que el conductor no abriría la tercera puerta si sabe que detrás está el premio. Como nosotros elegimos la primera puerta sabemos que $\mathbb{P}(A|B_2) = 1$, ya que no puede abrir la primera puerta porque es la que elegimos. Usando la fórmula de probabilidad total podemos calcular que:

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A|B_1) \mathbb{P}(B_1) + \mathbb{P}(A|B_2) \mathbb{P}(B_2) + \mathbb{P}(A|B_3) \mathbb{P}(B_3) \\ &= \frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} \\ &= \frac{1}{2}\end{aligned}$$

Entonces:

$$\begin{aligned}\mathbb{P}(B_1|A) &= \frac{\mathbb{P}(A|B_1) \mathbb{P}(B_1)}{\mathbb{P}(A)} \\ &= \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} \\ &= \frac{1}{3}\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(B_2|A) &= \frac{\mathbb{P}(A|B_2)\mathbb{P}(B_2)}{\mathbb{P}(A)} \\
&= \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} \\
&= \frac{2}{3}
\end{aligned}$$

Por lo tanto, nos conviene cambiar de elección a la segunda puerta ya que duplicamos nuestras probabilidades de ganarnos un auto.

Veamos una última definición. Sea $\{A_i\}$, $i \in [1, N] \subseteq N$ un conjunto de eventos lo llamamos una **Familia de Eventos Independientes** si:

$$\mathbb{P}\left(\bigcap_{n=1}^N A_n\right) = \prod_{n=1}^N \mathbb{P}(A_n)$$

Notemos que no es suficiente que cada evento sea independiente a cada otro, ya que deben ser todos independientes entre sí. Por ejemplo, consideremos que se tiran dos monedas y definamos los siguientes tres eventos:

- A : La primera moneda es cara.
- B : La segunda moneda es cara.
- C : Ambas monedas son iguales.

Notemos que $A \perp B$, $A \perp C$ y $B \perp C$, pero estos eventos no forman una familia de eventos independientes. El espacio de eventos es $\mathcal{F} = \{\{\ominus, \ominus\}, \{\ominus, \times\}, \{\times, \ominus\}, \{\times, \times\}\}$, por lo tanto $A = \{\ominus, \ominus\} \cup \{\ominus, \times\}$, $B = \{\ominus, \ominus\} \cup \{\times, \ominus\}$ y $C = \{\ominus, \ominus\} \cup \{\times, \times\}$:

$$\begin{aligned}
\mathbb{P}(A \cap B \cap C) &= \mathbb{P}((\{\ominus, \ominus\} \cup \{\ominus, \times\}) \cap (\{\ominus, \ominus\} \cup \{\times, \ominus\}) \cap (\{\ominus, \ominus\} \cup \{\times, \times\})) \\
&= \mathbb{P}(\{\ominus, \ominus\}) \\
&= \frac{1}{4}
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) &= \mathbb{P}(\{\ominus, \ominus\} \cup \{\ominus, \times\})\mathbb{P}(\{\ominus, \ominus\} \cup \{\times, \ominus\})\mathbb{P}(\{\ominus, \ominus\} \cup \{\times, \times\}) \\
&= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \\
&= \frac{1}{8}
\end{aligned}$$

$$\therefore \mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$$

2. Variables Aleatorias Discretas

2.1. Definiciones

Definimos una **Variable Aleatoria** como una función $X : \mathcal{S} \rightarrow \mathbb{R}$ que sirve para describir eventos en \mathcal{S} . Notamos $\{X \in A\} = \{s \in \mathcal{S} / X(s) \in A\}$. Las variables aleatorias discretas asumen numerables valores con probabilidad positiva. El rango de una variable aleatoria es $\mathcal{R}(x) = \{x \in \mathbb{R} / \mathbb{P}(X = x) > 0\}$. Definimos también una **Distribución de Probabilidad Puntual** (también conocida como **Función de Probabilidad Puntual**) como $p_X(x) = \mathbb{P}(X = x)$. Sea X una variable aleatoria y p su distribución de probabilidad notamos $X \sim p$.

Veamos un ejemplo. Primero, consideremos que se tiran dos monedas (notemos al resultado de cara como 1 y cruz como 0). Entonces, $\mathcal{S} = \{\{0, 0\}, \{0, 1\}, \{1, 0\}, \{1, 1\}\}$. Definamos la variable aleatoria X como el número de caras obtenidas. Por lo tanto, $X(0, 0) = 0$, $X(0, 1) = X(1, 0) = 1$ y $X(1, 1) = 2$. Como la probabilidad de obtener cara o cruz en cada moneda es de $1/2$ sabemos que $p_X(0) = p_X(2) = 1/4$ y $p_X(1) = 1/2$. Por lo tanto, podemos armar la siguiente tabla:

| | | | |
|-------|---------------|---------------|---------------|
| X | 0 | 1 | 2 |
| p_X | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ |

Este tipo de tablas es muy útil a la hora de resolver ejercicios porque nos provee una descripción clara y simple de los posibles valores de la variable junto con la probabilidad de obtener cada valor.

Veamos ahora un ejemplo más complejo. Consideremos que se tiran dos dados y definamos la variable aleatoria X como la suma de ambos resultados. Entonces, podemos armar la siguiente tabla:

| | | | | | | | | | | | |
|-------|----------------|----------------|----------------|---------------|----------------|---------------|----------------|---------------|----------------|----------------|----------------|
| X | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| p_X | $\frac{1}{36}$ | $\frac{1}{18}$ | $\frac{1}{12}$ | $\frac{1}{9}$ | $\frac{5}{36}$ | $\frac{1}{6}$ | $\frac{5}{36}$ | $\frac{1}{9}$ | $\frac{1}{12}$ | $\frac{1}{18}$ | $\frac{1}{36}$ |

Definamos ahora la **Distribución de Probabilidad Puntual Acumulada** como una función $P_X(x) = \mathbb{P}(X \leq x)$. Notemos que en el caso de una variable aleatoria discreta vale que:

$$P_X(x) = \sum_{X=t}^x p_X(t)$$

Algunas propiedades de la distribución acumulada son:

- $P_X(x) \leq P_X(y) \iff x \leq y$.
- $\lim_{x \rightarrow x_0^+} P_X(x) = P_X(x_0)$ (o sea que es continua por derecha).
- $\lim_{x \rightarrow -\infty} P_X(x) = 0$.
- $\lim_{x \rightarrow \infty} P_X(x) = 1$.

Demostremoslos:

$$P_X(x) \leq P_X(y) \iff x \leq y$$

$$\begin{aligned}
 P_X(y) &= P_X(x + (y - x)) \\
 &= \mathbb{P}(X \leq x + (y - x)) \\
 x \leq y &\rightarrow = \mathbb{P}((X \leq x) \cup (x < X \leq y - x)) \\
 (X \leq x) \cap (x < X \leq y - x) &= \emptyset \rightarrow = \mathbb{P}(X \leq x) + \mathbb{P}(x < X \leq y - x) \\
 &= P_X(x) + \underbrace{\mathbb{P}(x < X \leq y - x)}_{\geq 0} \\
 &\geq P_X(x)
 \end{aligned}$$

$\lim_{x \rightarrow x_0^+} P_X(x) = P_X(x_0)$ Veamos que $\exists \delta > 0 / |x - x_0| < \delta \implies |P_X(x) - P_X(x_0)| < \varepsilon \forall \varepsilon > 0, x_0 \in \mathbb{R}$.
Notemos:

$$\begin{aligned} |P_X(x) - P_X(x_0)| &= |\mathbb{P}(X \leq x) - \mathbb{P}(X \leq x_0)| \\ x > x_0 &\rightarrow = |\mathbb{P}((X \leq x_0) \cup (x_0 < X \leq x)) - \mathbb{P}(X \leq x_0)| \\ (X \leq x_0) \cap (x_0 < X \leq x) &= \emptyset \rightarrow = |\mathbb{P}(X \leq x_0) + \mathbb{P}(x_0 < X \leq x) - \mathbb{P}(X \leq x_0)| \\ &= \mathbb{P}(x_0 < X \leq x) \\ &< \varepsilon \end{aligned}$$

Entonces, si definimos a x como la sucesión $x_n = x_0 + 1/n$ tenemos que $\lim_{n \rightarrow \infty} x_n = x_0$ y $x_n > x_0 \forall n \in \mathbb{N}$.
Entonces:

$$\begin{aligned} \mathbb{P}(x_0 < X \leq x_n) &= \mathbb{P}\left(x_0 < X \leq x_0 - \frac{1}{n}\right) \\ &< \varepsilon \end{aligned}$$

Por lo tanto, $\exists n_0 \in \mathbb{N} / \mathbb{P}(x_0 < X \leq x_0 - \frac{1}{n}) < \varepsilon \forall n > n_0$. Entonces:

$$\begin{aligned} |x_n - x_0| &= \frac{1}{n} \\ &< \frac{1}{n_0} \\ &= \delta \end{aligned}$$

$\lim_{x \rightarrow -\infty} P_X(x) = 0$ Consideremos una sucesión de eventos $\{A_n\} \in \mathcal{S}$ donde $A_n = \{X \leq -x\}$. Notemos que $A_n \subset A_{n-1}$ y que $\lim_{x \rightarrow \infty} A_n = \{X \notin \mathbb{R}\}$ (ya que X no puede ser menor que $-\infty$). Entonces:

$$\begin{aligned} \lim_{x \rightarrow \infty} \mathbb{P}(A_n) &= \lim_{x \rightarrow \infty} \mathbb{P}(X \leq -x) \\ &= \lim_{x \rightarrow \infty} P_X(-x) \\ &= \lim_{x \rightarrow -\infty} P_X(x) \\ &= \mathbb{P}(X \notin \mathbb{R}) \\ &= 0 \end{aligned}$$

$\lim_{x \rightarrow \infty} P_X(x) = 1$ Consideremos una sucesión de eventos $\{A_n\} \in \mathcal{S}$ donde $A_n = \{X \leq x\}$. Notemos que $A_{n-1} \subset A_n$ y que $\lim_{x \rightarrow \infty} A_n = \{X \in \mathbb{R}\}$ (ya que X siempre va a ser menor que ∞). Entonces:

$$\begin{aligned} \lim_{x \rightarrow \infty} \mathbb{P}(A_n) &= \lim_{x \rightarrow \infty} \mathbb{P}(X \leq x) \\ &= \lim_{x \rightarrow \infty} P_X(x) \\ &= \mathbb{P}(X \in \mathbb{R}) \\ &= 1 \end{aligned}$$

Antes de pasar a ver distribuciones típicas veamos dos conceptos más que son importantes para el estudio de variables independientes. Definimos la **Esperanza** como la función:

$$E : \mathcal{S} \rightarrow \mathbb{N} / E(X) = \sum_{x \in \mathcal{R}(x)} x \mathbb{P}(X = x)$$

La esperanza representa lo que coloquialmente interpretamos como el “promedio” (aunque no es lo mismo). Nos da un valor (no necesariamente en el rango de X) que representa el “centro de masa” de los posibles valores de X .

Veamos un ejemplo. Si consideramos que se tira una moneda sabemos que los posibles resultados son $X = 1$ (cara) con probabilidad $1/2$ y $X = 0$ (cruz) con probabilidad $1/2$. Por lo tanto, la esperanza de la variable aleatoria X es:

$$\begin{aligned} E(X) &= 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} \\ &= \frac{1}{2} \end{aligned}$$

Notemos que $1/2 \notin \mathcal{R}(X)$ pero nos representa el “centro de masa” de los posibles resultados de X . Ahora, si tenemos una moneda trucha con cara de ambos lados tenemos que la probabilidad de sacar cara es de 1 y la de sacar cruz es de 0, así que la esperanza es:

$$\begin{aligned} E(X) &= 1 \cdot 1 + 0 \cdot 0 \\ &= 1 \end{aligned}$$

Ahora si la moneda está pesada tal que la probabilidad de sacar cara es de $1/3$ y la de sacar cruz es de $2/3$ la esperanza es:

$$\begin{aligned} E(X) &= 1 \cdot \frac{1}{3} + 0 \cdot \frac{2}{3} \\ &= \frac{1}{3} \end{aligned}$$

Notemos que el valor de la esperanza de una variable aleatoria no solo depende de los valores que la variable toma sino que también depende de la probabilidad de obtener cada valor.

Sean X e Y variables aleatorias, consideremos una función $g : \mathcal{R}(X) \rightarrow \mathcal{R}(Y)$ tal que $Y = g(X)$. Entonces:

$$E(Y) = \sum_{x \in \mathcal{R}(X)} g(x) \mathbb{P}(X = x)$$

Demostremoslo. Notemos:

$$\begin{aligned} \{Y = y\} &= \{g(X) = y\} \\ &= \cup_{x \in \mathcal{R}(X) / g(x)=y} \{X = x\} \\ \therefore \mathbb{P}(Y = y) &= \sum_{x \in \mathcal{R}(X) / g(x)=y} \mathbb{P}(X = x) \end{aligned}$$

Por lo tanto:

$$\begin{aligned} E(Y) &= \sum_{y \in \mathcal{R}(Y)} y \mathbb{P}(Y = y) \\ &= \sum_{y \in \mathcal{R}(Y)} y \sum_{x \in \mathcal{R}(X) / g(x)=y} \mathbb{P}(X = x) \\ &= \sum_{y \in \mathcal{R}(Y)} \sum_{x \in \mathcal{R}(X) / g(x)=y} g(x) \mathbb{P}(X = x) \\ &= \sum_{x \in \mathcal{R}(X)} g(x) \mathbb{P}(X = x) \end{aligned}$$

Un corolario de esta propiedad es la linealidad de la esperanza. Sean X e Y variables aleatorias tales que

$Y = aX + b$, $a, b \in \mathbb{R}$ entonces $E(Y) = aE(X) + b$. Notemos:

$$\begin{aligned} E(Y) &= \sum_{y \in \mathcal{R}(Y)} y \mathbb{P}(Y = y) \\ &= \sum_{x \in \mathcal{R}(x)} (ax + b) \mathbb{P}(X = x) \\ &= a \sum_{x \in \mathcal{R}(x)} x \mathbb{P}(X = x) + b \underbrace{\sum_{x \in \mathcal{R}(x)} \mathbb{P}(X = x)}_{=1} \\ &= aE(X) + b \end{aligned}$$

Ahora veamos el concepto de varianza. Definimos la **Varianza** como la función $V(X) = E((X - E(X))^2)$. Además, definimos el **Desvío Estándar** como la función $\sigma(X) = \sqrt{V(X)}$. Así como la varianza describe el “centro de masa” de una variable aleatoria, la varianza describe la desviación de los valores del centro de masa. De cierta forma es una medida de la distancia cuadrática entre los valores de la variable y la esperanza. Por ejemplo, si tenemos una variable aleatoria $X \in \{-1, 1\}$ y otra $Y \in \{-100, 100\}$ (con probabilidades $1/2$ para cada valor) entonces notemos que:

$$E(X) = E(Y) = 0,$$

pero:

$$\begin{aligned} V(X) &= \frac{1}{2} \cdot (-1 - 0)^2 + \frac{1}{2} \cdot (1 - 0)^2 \\ &= 1 \end{aligned}$$

$$\begin{aligned} V(Y) &= \frac{1}{2} \cdot (-100 - 0)^2 + \frac{1}{2} \cdot (100 - 0)^2 \\ &= 10000 \end{aligned}$$

En este caso la varianza de Y es mucho mayor que la de X porque los valores de X (que son -1 y 1) están a menor “distancia” de la esperanza (que es 0) que los valores de Y (que son -100 y 100).

Sea X una variable aleatoria y $m \in \mathbb{R}$ definamos la función $\tilde{V}(X, m) = E((X - m)^2)$. Encontramos el valor $m \in \mathbb{R}$ que minimice esta función:

$$\begin{aligned} \tilde{V}(X, m) &= E((X - m)^2) \\ &= E(X^2 - 2mX + m^2) \\ &= E(X^2) - 2mE(X) + m^2 \end{aligned}$$

$$\begin{aligned} \therefore \frac{d}{dm} \tilde{V}(X, m) &= -2E(X) + 2m \\ &= 0 \end{aligned}$$

$$\therefore m = E(X)$$

Esto resultado tiene valor porque $\tilde{V}(X, m)$ es una función que calcula la distancia cuadrática media entre la variable aleatoria X y un determinado número n . Si buscamos m tal que esta función sea mínima es lo mismo que buscar el punto que menos suele alejarse de todos los valores que puede tomar X ; es el “centro de masa”. Esta función termina siendo la varianza $V(X) = \tilde{V}(X, E(X))$.

Sea X una variable aleatoria entonces vale que:

$$V(X) = E(X^2) - E(X)^2$$

Notemos:

$$\begin{aligned} V(X) &= E\left((X - E(X))^2\right) \\ &= E\left(X^2 - 2XE(X) + E(X)^2\right) \\ \text{Linealidad} \rightarrow &= E(X^2) - 2E(X)E(X) + E(X)^2 \underbrace{E(1)}_{=1} \\ &= E(X^2) - E(X)^2 \end{aligned}$$

Además, vale que:

$$V(aX + b) = a^2V(X)$$

Notemos:

$$\begin{aligned} V(aX + b) &= E\left(\left((aX + b) - E(aX + b)\right)^2\right) \\ &= E\left(\left((aX + b) - (aE(X) + b)\right)^2\right) \\ &= E\left(\left(a(X - E(X))\right)^2\right) \\ &= E\left(a^2(X - E(X))^2\right) \\ &= a^2E\left((X - E(X))^2\right) \\ &= a^2V(X) \end{aligned}$$

Con estas definiciones y propiedades estamos listos para ver algunas distribuciones usuales para distintos tipos de variables discretas.

2.2. Distribución de Bernoulli

La **Distribución de Bernoulli** se utiliza para variables aleatorias que toman dos posibles valores: toma valor $X = 1$ con probabilidad p y $X = 0$ con probabilidad $1 - p$. Se suele usar para experimentos que solo pueden tomar un si o un no como respuesta o famosamente para el lanzamiento de una moneda. Podemos expresar a esta distribución como:

$$\text{Be}(p) = p\delta_{X1} + (1 - p)\delta_{X0}, 0 < p < 1$$

donde δ es la **Delta de Kronecker**:

$$\delta_{mn} = \begin{cases} 1 & m = n \\ 0 & m \neq n \end{cases}$$

Veamos cuál es la esperanza:

$$\begin{aligned} E(X) &= p \cdot 1 + 0 \cdot (1 - p) \\ &= p \end{aligned}$$

Ahora veamos la varianza:

$$\begin{aligned} V(X) &= p \cdot (1 - p)^2 + (1 - p)(0 - p)^2 \\ &= p(1 - p) \underbrace{((1 - p) + p)}_{=1} \\ &= p(1 - p) \end{aligned}$$

2.3. Distribución Geométrica

La **Distribución Geométrica** se utiliza para describir variables aleatorias que representan la cantidad de “ensayos Bernoulli” necesarios hasta obtener un “resultado positivo”. Por ejemplo, vimos que tirar una moneda es una variable Bernoulli, así que un ejemplo de variable con distribución geométrica sería la variable $X \in \mathbb{N}$ que representa la cantidad de veces que se tira la moneda hasta obtener cara. Cada tirada de moneda en este caso es lo que llamamos un “ensayo Bernoulli” y obtener cara es lo que llamamos un “resultado positivo”. Podemos expresar a esta distribución como:

$$\mathcal{G}(p) = p(1-p)^{x-1}, \quad 0 < p < 1$$

Veamos cuál es la esperanza:

$$\begin{aligned} E(X) &= \sum_{x=1}^{\infty} xp(1-p)^{x-1} \\ &= -p \sum_{x=1}^{\infty} \frac{d}{dp} ((1-p)^x) \\ &= -p \frac{d}{dp} \left(\sum_{x=1}^{\infty} (1-p)^x \right) \\ &= -p \frac{d}{dp} \left(\sum_{x=0}^{\infty} (1-p)^x - 1 \right) \\ 0 < p < 1 \rightarrow &= -p \frac{d}{dp} \left(\frac{1}{1-(1-p)} - 1 \right) \\ &= -p \left(-\frac{1}{p^2} \right) \\ &= \frac{1}{p} \end{aligned}$$

Ahora veamos la varianza:

$$\begin{aligned}
V(X) &= E(X^2) - E(X)^2 \\
&= \sum_{x=1}^{\infty} x^2 p (1-p)^{x-1} - \left(\frac{1}{p}\right)^2 \\
&= -p \frac{d}{dp} \left(\sum_{x=0}^{\infty} x (1-p)^x \right) - \frac{1}{p^2} \\
&= p \frac{d}{dp} \left(\sum_{x=0}^{\infty} \left(\frac{d}{dp} (1-p)^{x+1} + (1-p)^x \right) \right) - \frac{1}{p^2} \\
0 < p < 1 \rightarrow &= p \frac{d}{dp} \left(\frac{d}{dp} \left(\sum_{x=1}^{\infty} (1-p)^x \right) + \frac{1}{1-(1-p)} \right) - \frac{1}{p^2} \\
&= p \frac{d}{dp} \left(\frac{d}{dp} \left(\sum_{x=0}^{\infty} (1-p)^x - 1 \right) + \frac{1}{p} \right) - \frac{1}{p^2} \\
0 < p < 1 \rightarrow &= p \frac{d}{dp} \left(\frac{d}{dp} \left(\frac{1}{p} - 1 \right) + \frac{1}{p} \right) - \frac{1}{p^2} \\
&= p \frac{d}{dp} \left(-\frac{1}{p^2} + \frac{1}{p} \right) - \frac{1}{p^2} \\
&= p \left(\frac{2}{p^3} - \frac{1}{p^2} \right) - \frac{1}{p^2} \\
&= \frac{1}{p^2} - \frac{1}{p} \\
&= \frac{1-p}{p^2}
\end{aligned}$$

Una propiedad útil de la distribución geométrica es que “no tiene memoria”. Con esto nos referimos a que si $X \sim \mathcal{G}(p)$ y $x_1, x_2 \in \mathbb{N}$ entonces $\mathbb{P}(X \geq x_1 + x_2 | X \geq x_1) = \mathbb{P}(X \geq x_2)$. Notemos:

$$\begin{aligned}
\mathbb{P}(X \geq x) &= \sum_{n=x}^{\infty} \mathbb{P}(X = n) \\
&= \sum_{n=x}^{\infty} p (1-p)^n \\
&= p \sum_{n=x}^{\infty} (1-p)^n \\
&= p (1-p)^x \sum_{n=x}^{\infty} (1-p)^{n-x} \\
&= p (1-p)^x \sum_{n=0}^{\infty} (1-p)^n \\
0 < 1-p < 1 \rightarrow &= p (1-p)^x \underbrace{\left(\frac{1}{1-(1-p)} \right)}_{=\frac{1}{p}} \\
&= (1-p)^x
\end{aligned}$$

Usemos esta expresión de $\mathbb{P}(X \geq x)$ para demostrar la propiedad que enunciamos:

$$\begin{aligned} \mathbb{P}(X \geq x_1 + x_2 | X \geq x_1) &= \frac{\mathbb{P}((X \geq x_1 + x_2) \cap (X \geq x_1))}{\mathbb{P}(X \geq x_1)} \\ x_2 > 0 \rightarrow &= \frac{\mathbb{P}(X \geq x_1 + x_2)}{\mathbb{P}(X \geq x_1)} \\ &= \frac{(1-p)^{x_1+x_2}}{(1-p)^{x_1}} \\ &= (1-p)^{x_2} \\ &= \mathbb{P}(X \geq x_2) \end{aligned}$$

Notemos que para esta propiedad usamos una versión ligeramente distinta de la distribución geométrica, donde:

$$\mathcal{G}(p) = p(1-p)^x$$

Esta versión tiene un factor de $(1-p)$ que la que definimos inicialmente no tiene. Ambas descripciones son válidas y usuales.

2.4. Distribución Binomial

La **Distribución Binomial** es similar a la geométrica pero con la diferencia que describe variables aleatorias que representan la cantidad de ensayos Bernoulli con resultados positivos de un total de n ensayos. A diferencia de la geométrica, no se fija en cuántos ensayos son necesarios hasta obtener un resultado positivo sino que toma n ensayos independientes y se fija cuántos son positivos. Podemos expresar a esta distribución como:

$$B(n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad 0 < p < 1, \quad n \in \mathbb{N}$$

Veamos cuál es la esperanza:

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} x \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{\infty} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ &= np \sum_{x=0}^{\infty} \frac{(n-1)!}{(x-1)!((n-1)-(x-1))!} p^{x-1} (1-p)^{(n-1)-(x-1)} \\ &= np \sum_{x=0}^{\infty} \binom{n-1}{x-1} p^{x-1} (1-p)^{(n-1)-(x-1)} \\ (a+b)^n &= \sum_{k=0}^n \binom{n}{k} a^k b^{n-k} \rightarrow = np \underbrace{\lim_{n \rightarrow \infty} (p + (1-p))^{n-1}}_{=1} \\ &= np \end{aligned}$$

Otra manera de calcular la esperanza es tomando:

$$X = \sum_{i=1}^n X_i,$$

donde $Y_i \sim \text{Be}(p)$. Por lo tanto:

$$\begin{aligned}
 E(X) &= E\left(\sum_{i=1}^n Y_i\right) \\
 \text{Linealidad} &\rightarrow = \sum_{i=1}^n E(Y_i) \\
 Y_i \sim \text{Be}(p) &\rightarrow = \sum_{i=1}^n p \\
 &= np
 \end{aligned}$$

Ahora veamos la varianza. Al igual que la esperanza, como $Y_i \perp Y_j \forall i \neq j$, la varianza total es igual a la suma de las varianzas:

$$\begin{aligned}
 V(X) &= V\left(\sum_{i=1}^n Y_i\right) \\
 Y_i \perp Y_j \forall i \neq j &\rightarrow = \sum_{i=1}^n V(Y_i) \\
 Y_i \sim \text{Be}(p) &\rightarrow = \sum_{i=1}^n p(1-p) \\
 &= np(1-p)
 \end{aligned}$$

Una propiedad útil de las variables aleatorias con distribución binomial es que la suma de dos variables binomiales independientes con la misma probabilidad es binomial. Es decir, sean X_1 y X_2 dos variables aleatorias binomiales tales que $X_1 \sim B(n_1, p)$, $X_2 \sim B(n_2, p)$ y $X_1 \perp X_2$, y sea $Y = X_1 + X_2$ entonces $Y \sim B(n_1 + n_2, p)$. Notemos:

$$\begin{aligned}
 \mathbb{P}(Y = y) &= \mathbb{P}(X_1 + X_2 = y) \\
 &= \mathbb{P}\left(\bigcup_{x=1}^y (X_1 = x) \cap (X_2 = y - x)\right) \\
 &= \sum_{x=1}^y \mathbb{P}((X_1 = x) \cap (X_2 = y - x)) \\
 X_1 \perp X_2 &\rightarrow = \sum_{x=1}^y \mathbb{P}(X_1 = x) \mathbb{P}(X_2 = y - x) \\
 &= \sum_{x=1}^y \binom{n_1}{x} p^x (1-p)^{n_1-x} \binom{n_2}{y-x} p^{y-x} (1-p)^{n_2-y+x} \\
 &= \sum_{x=1}^y \binom{n_1}{x} \binom{n_2}{y-x} p^y (1-p)^{n_1+n_2-y}
 \end{aligned}$$

La **Identidad de Vandermonde** establece que:

$$\binom{m+n}{r} = \sum_{k=0}^r \binom{m}{k} \binom{n}{r-k}$$

Por lo tanto, usando esta identidad obtenemos que:

$$\begin{aligned}
 \mathbb{P}(Y = y) &= \binom{n_1 + n_2}{y} p^y (1-p)^{n_1+n_2-y} \\
 &= B(n_1 + n_2, p)
 \end{aligned}$$

2.5. Distribución Binomial Negativa

Similarmente a la distribución binomial, la **Distribución Binomial Negativa** se utiliza para variables aleatorias que se fijan en la cantidad de ensayos Bernoulli con resultados negativos antes de obtener una cantidad r de ensayos positivos. Podemos expresar a esta distribución como:

$$\text{NB}(n, p) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad 0 < p < 1, r \in \mathbb{N}$$

Veamos cuál es la esperanza. Sabemos que X representa la cantidad de intentos hasta obtener r aciertos, así que definamos a las variables aleatorias $Y_i \sim \mathcal{G}(p)$, $i \in [1, r] \subset \mathbb{N}$ como la cantidad de fracasos hasta obtener un resultado positivo. Notemos que Y_i representa los ensayos que debemos esperar entre cada victoria. Por lo tanto:

$$X = \sum_{i=1}^r Y_i$$

Entonces:

$$\begin{aligned} E(X) &= E\left(\sum_{i=1}^r Y_i\right) \\ &= \sum_{i=1}^r E(Y_i) \\ Y_i \sim \mathcal{G}(p) &\rightarrow = \sum_{i=1}^r \frac{1}{p} \\ &= \frac{r}{p} \end{aligned}$$

Similarmente, la varianza es:

$$\begin{aligned} V(X) &= V\left(\sum_{i=1}^r Y_i\right) \\ Y_i \perp Y_j \forall i \neq j &\rightarrow = \sum_{i=1}^r V(Y_i) \\ Y_i \sim \mathcal{G}(p) &\rightarrow = \sum_{i=1}^r \frac{1-p}{p^2} \\ &= \frac{r(1-p)}{p^2} \end{aligned}$$

2.6. Distribución de Poisson

La **Distribución de Poisson** se utiliza para variables aleatorias que representan la cantidad de eventos que suceden en un determinado intervalo de tiempo o espacio si los eventos ocurren con una frecuencia promedio conocida λ e independientemente entre sí. Por ejemplo, elementos radioactivos como el bismuto (Bi) emiten radiación γ siguiendo una distribución de Poisson. Cada uno de los fotones emitidos se emite con una frecuencia media λ y cada emisión es independiente de las otras (la emisión de un fotón no hace más o menos probable la emisión del siguiente). Lo mismo sucede con las llamadas telefónicas a un call-center. Cada llamada es independiente y ocurren con una frecuencia media definida. Podemos expresar esta distribución como:

$$\mathcal{P}(\lambda) = \frac{\lambda^x}{x!} \exp(-\lambda), \quad \lambda > 0, \lambda \in \mathbb{R}$$

Veamos cuál es la esperanza:

$$\begin{aligned}
 E(X) &= \sum_{x=1}^{\infty} x \frac{\lambda^x}{x!} \exp(-\lambda) \\
 &= \exp(-\lambda) \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} \\
 &= \exp(-\lambda) \sum_{x=0}^{\infty} \frac{\lambda^{x+1}}{x!} \\
 &= \lambda \exp(-\lambda) \underbrace{\sum_{x=0}^{\infty} \frac{\lambda^x}{x!}}_{=\exp(\lambda)} \\
 &= \lambda
 \end{aligned}$$

Esto es lo que esperábamos ya que mencionamos en la definición que λ representa la frecuencia media. Ahora veamos la varianza:

$$\begin{aligned}
 V(X) &= E(X^2) - E(X)^2 \\
 &= \sum_{x=1}^{\infty} x^2 \frac{\lambda^x}{x!} \exp(-\lambda) - \lambda^2 \\
 &= \exp(-\lambda) \sum_{x=1}^{\infty} x \frac{\lambda^x}{(x-1)!} - \lambda^2 \\
 &= \lambda \exp(-\lambda) \sum_{x=0}^{\infty} (x+1) \frac{\lambda^x}{x!} - \lambda^2 \\
 &= \lambda \exp(-\lambda) \underbrace{\sum_{x=0}^{\infty} x \frac{\lambda^x}{x!}}_{=\lambda \exp(\lambda)} + \lambda \exp(-\lambda) \underbrace{\sum_{x=0}^{\infty} \frac{\lambda^x}{x!}}_{=\exp(\lambda)} - \lambda^2 \\
 &= \lambda^2 + \lambda - \lambda^2 \\
 &= \lambda
 \end{aligned}$$

Notemos que $E(X) = V(X) = \lambda$.

Una propiedad útil de las variables con distribución de Poisson es que la suma de variables Poisson independientes es Poisson. Es decir, sean X_1 y X_2 variables aleatorias de Poisson tales que $X_1 \sim \mathcal{P}(\lambda_1)$, $X_2 \sim \mathcal{P}(\lambda_2)$ y $X_1 \perp X_2$

y sea $Y = X_1 + X_2$ entonces $Y \sim \mathcal{P}(\lambda_1 + \lambda_2)$. Notemos:

$$\begin{aligned}
\mathbb{P}(Y = y) &= \mathbb{P}(X_1 + X_2 = y) \\
&= \mathbb{P}(\cup_{x=0}^y ((X_1 = x) \cap (X_2 = y - x))) \\
&= \sum_{x=0}^y \mathbb{P}((X_1 = x) \cap (X_2 = y - x)) \\
X_1 \perp X_2 \rightarrow &= \sum_{x=0}^y \mathbb{P}(X_1 = x) \mathbb{P}(X_2 = y - x) \\
&= \sum_{x=0}^y \frac{\lambda_1^x}{x!} \exp(-\lambda_1) \frac{\lambda_2^{y-x}}{(y-x)!} \exp(-\lambda_2) \\
&= \exp(-(\lambda_1 + \lambda_2)) \sum_{x=0}^y \frac{\lambda_1^x \lambda_2^{y-x}}{x! (y-x)!} \\
&= \exp(-(\lambda_1 + \lambda_2)) \frac{1}{y!} \sum_{x=0}^y \binom{y}{x} \lambda_1^x \lambda_2^{y-x} \\
&\qquad\qquad\qquad \underbrace{\hspace{10em}}_{=(\lambda_1 + \lambda_2)^y} \\
&= \frac{(\lambda_1 + \lambda_2)^y}{y!} \exp(-(\lambda_1 + \lambda_2)) \\
&= \mathcal{P}(\lambda_1 + \lambda_2)
\end{aligned}$$

Veamos que la distribución de Poisson nace de una binomial. Consideremos:

$$\left\{ I_n = \left[\frac{n-1}{N}, \frac{n}{N} \right] / n \in [1, N] \subset \mathbb{N} \right\}$$

Una sucesión de N intervalos de $[0, 1]$. Consideremos que en cada intervalo puede haber p un punto o ningún punto y que la probabilidad es independiente para cada intervalo. Definamos X_n como la variable aleatoria de Bernoulli que indica si hay un punto en el intervalo I_n y tomemos $X_n \sim \text{Be}(\frac{\lambda}{N})$. Definamos también a la variable aleatoria S_N como la cantidad de puntos en el intervalo $[0, 1]$. O sea:

$$S_N = \sum_{n=1}^N X_n$$

Notemos que como $X_n \sim \text{Be}(\frac{\lambda}{N}) \implies S_N \sim \text{B}(N, \frac{\lambda}{N})$. Notemos:

$$\begin{aligned}
\lim_{N \rightarrow \infty} S_N &\sim \lim_{N \rightarrow \infty} \text{B}\left(N, \frac{\lambda}{N}\right) \\
&= \lim_{N \rightarrow \infty} \binom{N}{x} \left(\frac{\lambda}{N}\right)^x \left(1 - \frac{\lambda}{N}\right)^{N-x} \\
&= \lim_{N \rightarrow \infty} \frac{N!}{x! (N-x)!} \left(\frac{\lambda}{N}\right)^x \left(1 - \frac{\lambda}{N}\right)^{N-x} \\
&= \frac{\lambda^x}{x!} \lim_{N \rightarrow \infty} \underbrace{\frac{N!}{(N-x)! N^x}}_{\rightarrow 1} \underbrace{\left(1 - \frac{\lambda}{N}\right)^N}_{\rightarrow \exp(-\lambda)} \underbrace{\left(1 - \frac{\lambda}{N}\right)^{-x}}_{\rightarrow 1} \\
&= \frac{\lambda^x}{x!} \exp(-\lambda) \\
&= \mathcal{P}(\lambda)
\end{aligned}$$

2.7. Distribución Hipergeométrica

La **Distribución Hipergeométrica** se utiliza para variables aleatorias que representan la cantidad de ensayos positivos de m ensayos totales, sin reemplazos, y de una población finita de N pruebas donde B se consideran como pruebas positivas. Por ejemplo, consideremos que tenemos una bolsa donde hay N bolas rojas y verdes: B son verdes y $N - B$ son rojas. Queremos sacar m bolas de la bolsa (sin volver a meter las bolas que sacamos de la bolsa) y queremos ver cuántas de las bolas que sacamos son verdes. Entonces, la variable aleatoria X que representa la cantidad de bolas verdes que sacamos de la bolsa (sin reemplazo) es una variable aleatoria hipergeométrica. Podemos expresar a esta distribución como:

$$\mathcal{H}(N, B, m) = \frac{\binom{B}{x} \binom{N-B}{m-x}}{\binom{N}{m}}$$

En este caso para esta distribución no vamos a calcular el valor de la esperanza ni el de la varianza porque no tengo la menor idea de cómo se hace. Simplemente voy a escribir lo que dice en Wikipedia:

$$\begin{cases} E(X) = m \frac{B}{N} \\ V(X) = m \frac{B}{N} \frac{(N-B)}{N} \frac{(N-m)}{(N-1)} \end{cases}$$

2.8. Distribución Empírica

La **Distribución Empírica** se utiliza para hacer estadística con variables aleatorias que surgen a partir de experimentos hechos. Es una función simple que toma los valores empíricos obtenidos a medida que se fueron obteniendo. Podemos expresar a esta distribución como:

$$\hat{F}(n) = \frac{1}{n} \delta_{X_i X}$$

donde X_i son los valores empíricos obtenidos de la variable aleatoria X y n es la cantidad de mediciones que se hicieron. Notemos que esta distribución no nos dice nada en particular de la variable, sino que nos sirve para representar la información obtenida de los experimentos hechos. Por eso es que esta distribución se utiliza para hacer estadística.

Veamos cuánto vale la esperanza:

$$\begin{aligned} E(X) &= \sum_{i=1}^n X_i \frac{1}{n} \delta_{X_i X} \\ &= \frac{1}{n} \sum_{i=1}^n X_i \end{aligned}$$

Notemos que la esperanza es igual al promedio. Ahora veamos la varianza:

$$\begin{aligned} V(X) &= \sum_{i=1}^n (X_i - E(X))^2 \frac{1}{n} \delta_{X_i X} \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - E(X))^2 \end{aligned}$$

De la misma forma, la varianza es igual a la varianza muestral.

| Nombre | Distribución | $E(X)$ | $V(X)$ |
|-------------------|-----------------------------------------------------------------------------|--------------------------------|-----------------------------------------------------|
| Bernoulli | $\text{Be}(p) = p\delta_{X1} + (1-p)\delta_{X0}$ | p | $p(1-p)$ |
| Geométrica | $\mathcal{G}(p) = p(1-p)^{x-1}$ | $\frac{1}{p}$ | $\frac{1-p}{p^2}$ |
| Binomial | $B(n, p) = \binom{n}{x} p^x (1-p)^{n-x}$ | np | $np(1-p)$ |
| Binomial Negativa | $\text{NB}(r, p) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$ | $\frac{r}{p}$ | $\frac{r(1-p)}{p^2}$ |
| Poisson | $\mathcal{P}(\lambda) = \frac{\lambda^x}{x!} \exp(-\lambda)$ | λ | λ |
| Hipergeométrica | $\mathcal{H}(N, B, m) = \frac{\binom{B}{x} \binom{N-B}{m-x}}{\binom{N}{m}}$ | $m \frac{B}{N}$ | $m \frac{B}{N} \frac{(N-B)}{N} \frac{(N-m)}{(N-1)}$ |
| Empírica | $\hat{F}(n) = \frac{1}{n} \delta_{X X_i}$ | $\frac{1}{n} \sum_{i=1}^n X_i$ | $\frac{1}{n} \sum_{i=1}^n (X_i - E(X))^2$ |

3. Variables Aleatorias Continuas

3.1. Definiciones

Definimos una **Variable Aleatoria Continua** a una variable aleatoria $X : \mathcal{S} \rightarrow \mathbb{R}$ tal que:

$$\mathbb{P}(X \in A) = \int_A f_X(x) dx$$

Llamamos a f_X la **Distribución de Probabilidad** de X . Esta función debe cumplir que $f_X : \mathbb{R} \rightarrow \mathbb{R}$, $f_X(x) \geq 0 \forall x \in \mathcal{S}$ y cumple que:

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

De la misma forma definimos la **Distribución Acumulada** como:

$$F_X(x) = \int_{-\infty}^x f_X(t) dt$$

Notemos que estas definiciones son el equivalente continuo de los conceptos que definimos para variables aleatorias discretas. También, sea $p \in [0, 1] \subset \mathbb{R}$ llamamos **Percentil** p a $x_p \in \mathcal{S}$ tal que $F_X(x_p) = p$. En particular, llamamos **Mediana** al percentil $1/2$. Por último, definimos el **Promedio** de una función $g : \mathbb{R} \rightarrow \mathbb{R}$ como:

$$\langle g \rangle = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

En particular, la esperanza de una distribución es $E(X) = \langle X \rangle$.

Sea X una variable aleatoria continua que cumple $X \geq 0$ y sea f su distribución de probabilidad entonces:

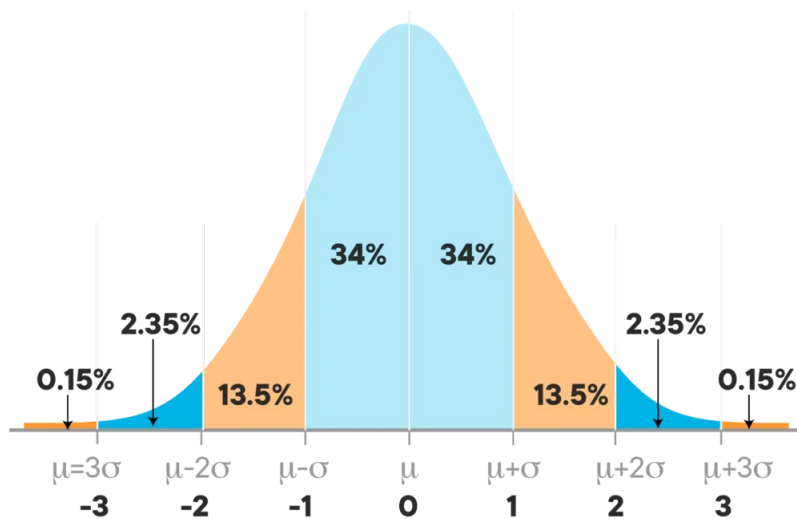
$$E(X) = \int_0^{\infty} (1 - F(x)) dx$$

Notemos:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\ X \geq 0 &\rightarrow = \int_0^{\infty} x f(x) dx \\ &= \int_0^{\infty} \left(\int_0^x dt \right) f(x) dx \\ \text{Fubini} &\rightarrow = \int_0^{\infty} \int_x^{\infty} f(t) dt dx \\ &= \int_0^{\infty} \left(\int_0^{\infty} f(t) dt - \int_0^x f(t) dt \right) dx \\ &= \int_0^{\infty} (1 - F(x)) dx \end{aligned}$$

Con estas definiciones estamos listos para ver algunas distribuciones usuales para distintos tipos de variables continuas.

3.2. Distribución Normal



Una de las distribuciones de probabilidad continua más común es la **Distribución Normal**. Sea X una variable aleatoria con media μ y varianza σ^2 la distribución normal se puede escribir como:

$$\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

Esta distribución se usa para todos los tipos de variables en las ciencias sociales y naturales cuando la “verdadera” distribución no se conoce. La importancia de esta distribución va a ser mejor entendida cuando veamos el Teorema Central de Límite, pero todo a su debido tiempo.

Definimos a la **Distribución Normal Estándar** a la distribución normal con los valores $\mu = 0$ y $\sigma^2 = 1$. En este caso la expresión de la distribución es:

$$\mathcal{N}(0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

Esta versión particular de la distribución es útil ya que sus valores en distintos puntos están bien calculados y se encuentran con facilidad en distintas tablas (lo cual es útil para hacer cálculos de probabilidad, ya que la integral es fea). Además, es útil ya que una propiedad de la distribución normal es que si $X \sim \mathcal{N}(\mu, \sigma^2)$ entonces $Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$. Notemos:

$$\begin{aligned} \Phi_Z(z) & \stackrel{\text{def}}{=} F_Z(z) \\ & = \mathbb{P}(Z \leq z) \\ & = \mathbb{P}\left(\frac{X - \mu}{\sigma} \leq z\right) \\ & = \mathbb{P}(X \leq \sigma z + \mu) \\ & = \Phi_X(\sigma z + \mu) \end{aligned}$$

$$\begin{aligned}
\therefore f_Z(z) &= \frac{d}{dz} \Phi_Z(z) \\
&= \frac{d}{dz} \Phi_X(\sigma z + \mu) \\
&= \sigma f_X(\sigma z + \mu) \\
&= \sigma \left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\sigma z + \mu - \mu)^2}{2\sigma^2}\right) \right) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \\
&= \mathcal{N}(0, 1)
\end{aligned}$$

De la misma forma podemos demostrar que una combinación lineal de variables normales es normal.

Para complementar esta sección y acostumbrarnos a hacer cálculos con la normal demos que la distribución normal es una distribución de probabilidad válida:

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx &= \sqrt{\left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx\right)^2} \\
&= \sqrt{\left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx\right) \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy\right)} \\
&= \sqrt{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right) dx dy} \\
\text{Cambiamos a polares} \rightarrow &= \sqrt{\int_0^{\infty} \int_0^{2\pi} \frac{r}{2\pi} \exp\left(-\frac{r^2}{2}\right) dr d\theta} \\
&= \sqrt{\int_0^{\infty} r \exp\left(-\frac{r^2}{2}\right) dr} \\
&= \sqrt{-\exp\left(-\frac{r^2}{2}\right) \Big|_0^{\infty}} \\
&= \sqrt{-(0 - 1)} \\
&= 1
\end{aligned}$$

3.3. Distribución Uniforme

La **Distribución Uniforme** se utiliza para variables aleatorias cuya probabilidad es uniforme en cierto intervalo. Por ejemplo, podemos pensar en que la probabilidad de que llegue el colectivo a la parada es uniforme en un intervalo de 10 minutos. Dentro de ese intervalo el colectivo puede llegar en cualquier momento y con la misma probabilidad, la probabilidad de que llegue en el primer minuto es la misma de que llegue en el último. Podemos describir a esta distribución como:

$$\mathcal{U}(a, b) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x),$$

donde $[a, b]$ denota el intervalo ($a, b \in \mathbb{R}, b > a$) e \mathbb{I} es la **Función Indicatriz**:

$$\mathbb{I}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

Veamos cuál es la esperanza:

$$\begin{aligned}
 E(X) &= \int_{-\infty}^{\infty} x \frac{1}{b-a} \mathbb{I}_{[a,b]}(x) dx \\
 &= \frac{1}{b-a} \int_a^b x dx \\
 &= \frac{1}{b-a} \left(\frac{x^2}{2} \right) \Big|_a^b \\
 &= \frac{b^2 - a^2}{2(b-a)} \\
 &= \frac{\cancel{(b-a)}(b+a)}{2\cancel{(b-a)}} \\
 &= \frac{a+b}{2}
 \end{aligned}$$

Ahora veamos la varianza:

$$\begin{aligned}
 V(X) &= E(X^2) - E(X)^2 \\
 &= \int_{-\infty}^{\infty} x^2 \frac{1}{b-a} \mathbb{I}_{[a,b]}(x) dx - \left(\frac{a+b}{2} \right)^2 \\
 &= \frac{1}{b-a} \int_a^b x^2 dx - \frac{a^2 + 2ab + b^2}{4} \\
 &= \frac{1}{b-a} \left(\frac{x^3}{3} \right) \Big|_a^b - \frac{a^2 + 2ab + b^2}{4} \\
 &= \frac{b^3 - a^3}{3(b-a)} - \frac{a^2 + 2ab + b^2}{4} \\
 &= \frac{\cancel{(b-a)}(b^2 + ab + a^2)}{3\cancel{(b-a)}} - \frac{a^2 + 2ab + b^2}{4} \\
 &= \frac{\cancel{4}b^2 + \cancel{4}ab + \cancel{4}a^2 - \cancel{3}a^2 - \cancel{6}ab - \cancel{3}b^2}{12} \\
 &= \frac{b^2 - 2ab + a^2}{12} \\
 &= \frac{(b-a)^2}{12}
 \end{aligned}$$

3.4. Distribución Exponencial

La **Distribución Exponencial** se utiliza para procesos continuos que suceden independientemente entre sí y con cierta frecuencia promedio λ . Es la distribución de probabilidad del tiempo entre sucesos de Poisson. Podemos expresar a esta distribución como:

$$\mathcal{E}(\lambda) = \lambda \exp(-\lambda x), \quad x \geq 0$$

Veamos cuál es la esperanza:

$$\begin{aligned}
 E(X) &= \int_0^{\infty} x \lambda \exp(-\lambda x) dx \\
 &= \lambda \int_0^{\infty} x \exp(-\lambda x) dx \\
 &= \lambda \left(\left(-\frac{1}{\lambda} x \exp(-\lambda x) \right) \Big|_0^{\infty} + \frac{1}{\lambda} \int_0^{\infty} \exp(-\lambda x) dx \right) \\
 &= \lambda \left((0 - 0) - \frac{1}{\lambda^2} (\exp(-\lambda x)) \Big|_0^{\infty} \right) \\
 &= -\frac{1}{\lambda} (0 - 1) \\
 &= \frac{1}{\lambda}
 \end{aligned}$$

Ahora veamos la varianza:

$$\begin{aligned}
 V(X) &= E(X^2) - E(X)^2 \\
 &= \int_0^{\infty} x^2 \lambda \exp(-\lambda x) dx - \frac{1}{\lambda^2} \\
 &= \lambda \left(\left(-\frac{1}{\lambda} x^2 \exp(-\lambda x) \right) \Big|_0^{\infty} + \underbrace{\frac{2}{\lambda} \int_0^{\infty} x \exp(-\lambda x) dx}_{=\frac{1}{\lambda^2}} \right) - \frac{1}{\lambda^2} \\
 &= \lambda \left((0 - 0) + \frac{2}{\lambda^3} \right) - \frac{1}{\lambda^2} \\
 &= \frac{1}{\lambda^2}
 \end{aligned}$$

Una propiedad útil de la distribución exponencial es que, al igual que la distribución geométrica, no tiene memoria. Es decir, sea $X \sim \mathcal{E}(\lambda)$ entonces $\mathbb{P}(X \geq x_1 + x_2 | X \geq x_1) = \mathbb{P}(X \geq x_2)$. Notemos:

$$\begin{aligned}
 \mathbb{P}(X \geq x) &= \int_x^{\infty} \lambda \exp(-\lambda t) dt \\
 &= -\frac{\lambda}{\lambda} (\exp(-\lambda t)) \Big|_x^{\infty} \\
 &= -(0 - \exp(-\lambda x)) \\
 &= \exp(-\lambda x)
 \end{aligned}$$

Usemos esta expresión de $\mathbb{P}(X \geq x)$ para demostrar la propiedad que enunciamos:

$$\begin{aligned} \mathbb{P}(X \geq x_1 + x_2 | X \geq x_1) &= \frac{\mathbb{P}((X \geq x_1 + x_2) \cap (X \geq x_1))}{\mathbb{P}(X \geq x_1)} \\ &= \frac{\mathbb{P}(X \geq x_1 + x_2)}{\mathbb{P}(X \geq x_1)} \\ &= \frac{\exp(-\lambda(x_1 + x_2))}{\exp(-\lambda x_1)} \\ &= \frac{\exp(-\lambda x_1) \exp(-\lambda x_2)}{\exp(-\lambda x_1)} \\ &= \exp(-\lambda x_2) \\ &= \mathbb{P}(X \geq x_2) \end{aligned}$$

Otra propiedad útil es que la distribución exponencial es la versión continua de la distribución geométrica. Es por eso que ambas cumplen la propiedad de no-memoria. A lo que nos referimos con “versión continua” es que si $X_n \sim \mathcal{G}\left(\frac{\lambda}{n}\right)$ entonces:

$$\lim_{n \rightarrow \infty} \frac{X_n}{n} \sim \mathcal{E}(\lambda)$$

Notemos:

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{X_n}{n} \geq x\right) &= \lim_{n \rightarrow \infty} \mathbb{P}(X_n \geq nx) \\ &= \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{nx} \\ &= \lim_{n \rightarrow \infty} \left(1 + \frac{1}{(-\frac{n}{\lambda})}\right)^{\left(-\frac{n}{\lambda}\right)(-\lambda x)} \\ &= \exp(-\lambda x) \end{aligned}$$

3.5. Distribución Gamma

La **Distribución Gamma** se suele utilizar en econometría para modelar tiempos de espera y en estadística bayesiana para modelar el parámetro λ de una distribución de Poisson o exponencial. Podemos expresar a esta distribución como:

$$\Gamma(\alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\lambda x), \quad x \geq 0,$$

donde Γ es la **Función Gamma**:

$$\Gamma(z) = \int_0^\infty x^{z-1} \exp(-x) dx, \quad z \in \mathbb{C}$$

Vale que si $z \in \mathbb{N}$ entonces $\Gamma(z) = (z-1)!$. Además, esta función cumple que $\Gamma(z) = (z-1)\Gamma(z) \forall z \in \mathbb{C}$.

No lo vamos a demostrar, pero vale que la esperanza de la distribución gamma es:

$$E(X) = \frac{\alpha}{\lambda}$$

y la varianza es:

$$V(X) = \frac{\alpha}{\lambda^2}$$

Una propiedad útil de la distribución gamma es que si $X_i \sim \mathcal{E}(\lambda)$, $i \in [1, n] \subset \mathbb{N}$, entonces:

$$\sum_{i=1}^n X_i \sim \Gamma(n, \lambda)$$

Tampoco vamos a demostrarla. Otra propiedad útil (que tampoco vamos a demostrar) es que si $X \sim \Gamma(\alpha, \lambda) \implies \mu X \sim \Gamma\left(\alpha, \frac{\lambda}{\mu}\right) \forall \mu \in \mathbb{R}$.

3.6. Distribución Chi Cuadrado

La **Distribución Chi Cuadrado** es un caso particular de la distribución gamma que se utiliza para testeo de hipótesis y modela la suma de cuadrados de n variables aleatorias con distribución estándar normal. Podemos expresar a esta distribución como:

$$\chi_n^2 = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$$

Llamamos a n los **Grados de Libertad** de la distribución.

Notemos que al ser un caso particular de la distribución gamma sabemos que su esperanza es:

$$E(X) = n$$

y su varianza es:

$$V(X) = 2n$$

3.7. Distribución t-Student

La **Distribución t-Student** se utiliza para estimar la media de variables normales cuando la cantidad de muestras n es pequeña ($n < 30$) y la desviación estándar es desconocida. Podemos expresar a esta distribución como:

$$t_n = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

En este caso también llamamos a n los grados de libertad.

No lo vamos a demostrar pero vale que la esperanza de esta distribución es:

$$E(X) = 0, n > 1$$

y la varianza es:

$$V(X) = \frac{n}{n-2}, n > 2$$

| Nombre | Distribución | $E(X)$ | $V(X)$ |
|--------------|--------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|----------------------------|
| Normal | $\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ | μ | σ^2 |
| Uniforme | $\mathcal{U}(a, b) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x)$ | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ |
| Exponencial | $\mathcal{E}(\lambda) = \lambda \exp(-\lambda x)$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |
| Gamma | $\Gamma(\alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\lambda x)$ | $\frac{\alpha}{\lambda}$ | $\frac{\alpha}{\lambda^2}$ |
| Chi Cuadrado | $\chi_n^2 = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ | n | $2n$ |
| t-Student | $t_n = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$ | $0, n > 1$ | $\frac{n}{n-2}, n > 2$ |

4. Vectores Aleatorios

4.1. Distribuciones Marginales

Definimos un **Vector Aleatorio** al vector $\mathbf{X} : \mathcal{S} \rightarrow \mathbb{R}^N$, $N \in \mathbb{N}$ con distribución $f_{\mathbf{X}} : \mathbb{R}^N \rightarrow \mathbb{R}$ tal que:

$$(f_{\mathbf{X}}(\mathbf{x}) \geq 0 \forall \mathbf{x} \in \mathbb{R}^N) \wedge \left(\int_{\mathbb{R}^N} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = 1 \right)$$

Sea $\Xi = \{x_i \in \mathbb{R} / x_i \leq t_i \forall i \in [1, N], \mathbf{t} \in \mathbb{R}^N\}$ definimos la distribución acumulada de \mathbf{X} como:

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{\Xi} f(\mathbf{x}) d\mathbf{x}$$

También, definimos la **Distribución Marginal** de una coordenada x_i de \mathbf{X} como:

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{N-1}} f(\mathbf{x}) d\tilde{\mathbf{x}}_i,$$

donde $\tilde{\mathbf{x}}_i$ es el vector con todas las coordenadas de \mathbf{X} menos x_i .

Veamos un ejemplo. Consideremos la siguiente distribución:

$$f(x, y) = \frac{1}{y} \exp\left(-y - \frac{x}{y}\right), x, y > 0$$

Notemos:

$$f(x, y) > 0 \forall (x, y) \in \mathbb{R}^2$$

$$\begin{aligned} \int_{\mathbb{R}^2} f(x, y) d\mathbf{x} &= \int_0^{\infty} \int_0^{\infty} \frac{1}{y} \exp\left(-y - \frac{x}{y}\right) dx dy \\ &= \int_0^{\infty} \frac{1}{y} \exp(-y) \left(\int_0^{\infty} \exp\left(-\frac{x}{y}\right) dx \right) dy \\ &= \int_0^{\infty} \frac{1}{y} \exp(-y) \left(-y \exp\left(-\frac{x}{y}\right) \right) \Big|_0^{\infty} dy \\ &= \int_0^{\infty} \frac{1}{y} \exp(-y) (0 + y) dy \\ &= \int_0^{\infty} \exp(-y) dy \\ &= (-\exp(-y)) \Big|_0^{\infty} \\ &= -0 + 1 \\ &= 1 \end{aligned}$$

Con esto demostramos que es una distribución válida. Calculemos la densidad marginal de Y :

$$\begin{aligned}
 f_Y(y) &= \int_0^{\infty} f(x, y) dx \\
 &= \int_0^{\infty} \frac{1}{y} \exp\left(-y - \frac{x}{y}\right) dx \\
 &= \frac{1}{y} \exp(-y) \int_0^{\infty} \exp\left(-\frac{x}{y}\right) dx \\
 &= \frac{1}{y} \exp(-y) \left(-y \exp\left(-\frac{x}{y}\right)\right) \Big|_0^{\infty} \\
 &= \frac{1}{y} \exp(-y) (0 + y) \\
 &= \exp(-y)
 \end{aligned}$$

Por lo tanto, vemos que $Y \sim \mathcal{E}(1)$. Notemos que calcular la densidad marginal de X es significativamente más difícil, ya que la integral es más compleja.

Veamos otro ejemplo. Consideremos la siguiente distribución:

$$f(x, y) = \frac{1}{x} \mathbb{I}_{\{0 < y \leq x \leq 1\}}(x, y)$$

Notemos que la función indicatriz logra acotar el dominio de la función. En este caso nos acota el dominio al triángulo rectángulo con puntos $(0, 0)$, $(0, 1)$ y $(1, 1)$. Notemos también que el valor de la función no depende de y . Primero, demostremos que esta función es una distribución válida. Es evidente que $f(x, y) \geq 0 \forall (x, y) \in \mathbb{R}^2$, así que veamos que la integral nos da 1:

$$\begin{aligned}
 \int_{\mathbb{R}^2} f(x, y) d\mathbf{x} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{x} \mathbb{I}_{\{0 < y \leq x \leq 1\}}(x, y) dy dx \\
 &= \int_0^1 \int_0^x \frac{1}{x} dy dx \\
 &= \int_0^1 \frac{x}{x} dx \\
 &= x \Big|_0^1 \\
 &= 1
 \end{aligned}$$

Calculemos la densidad marginal de X :

$$\begin{aligned}
 f_X(x) &= \int_{-\infty}^{\infty} \frac{1}{x} \mathbb{I}_{\{0 < y \leq x \leq 1\}}(x, y) dy \\
 &= \int_0^x \frac{1}{x} \mathbb{I}_{\{0 < x \leq 1\}}(x) dy \\
 &= \mathbb{I}_{\{0 < x \leq 1\}}(x)
 \end{aligned}$$

Ahora calculemos la densidad marginal de Y :

$$\begin{aligned}
 f_Y(y) &= \int_{-\infty}^{\infty} \frac{1}{x} \mathbb{I}_{\{0 < y \leq x \leq 1\}}(x, y) dx \\
 &= \int_y^1 \frac{1}{x} \mathbb{I}_{\{0 < y \leq 1\}}(y) dx \\
 &= \ln(x) \Big|_y^1 \mathbb{I}_{\{0 < y \leq 1\}}(y) \\
 &= -\ln(y) \mathbb{I}_{\{0 < y \leq 1\}}(y)
 \end{aligned}$$

Ahora, veamos la independencia entre X e Y . Sean X e Y variables aleatorias con distribuciones f_X y f_Y respectivamente entonces $X \perp Y \iff f(x, y) = f_X(x) f_Y(y)$. Consideremos el siguiente ejemplo. Supongamos que tiramos una moneda dos veces. Definamos dos variables Bernoulli: $X = 1$ si el número de caras es par e $Y = 1$ si la primera moneda es cara. Entonces, calculemos primero las probabilidades marginales:

$$\begin{aligned}
 \mathbb{P}(X = 1) &= \mathbb{P}(\{\ominus, \ominus\} \cup \{\times, \times\}) \\
 &= \frac{1}{2}
 \end{aligned}$$

$$\begin{aligned}
 \mathbb{P}(Y = 1) &= \mathbb{P}(\{\ominus, \ominus\} \cup \{\ominus, \times\}) \\
 &= \frac{1}{2}
 \end{aligned}$$

Ahora, calculemos la probabilidad conjunta:

$$\begin{aligned}
 \mathbb{P}(X = 1, Y = 1) &= \mathbb{P}(\{\ominus, \ominus\}) \\
 &= \frac{1}{4} \\
 &= \frac{1}{2} \cdot \frac{1}{2} \\
 &= \mathbb{P}(X = 1) \mathbb{P}(Y = 1)
 \end{aligned}$$

Por lo tanto, $X \perp Y$. En este caso es suficiente con mostrar que $\mathbb{P}((X = 1) \cap (Y = 1)) = \mathbb{P}(X = 1) \mathbb{P}(Y = 1)$ porque X e Y solo asumen dos valores.

Sean X e Y variables aleatorias y h y g funciones tales que:

$$\mathbb{P}(X = x, Y = y) = Cg(x)h(y), \quad C \in \mathbb{R},$$

entonces:

$$\begin{cases} X \perp Y \\ \mathbb{P}(X = x) = \frac{g(x)}{\sum_{x'} g(x')} \\ \mathbb{P}(Y = y) = \frac{h(y)}{\sum_{y'} h(y')} \end{cases}$$

Notemos:

$$\begin{aligned}
 \sum_{x'} \sum_{y'} Cg(x)h(y) &= 1 \\
 \therefore C &= \frac{1}{\sum_{x'} \sum_{y'} g(x)h(y)}
 \end{aligned}$$

$$\begin{aligned}
\mathbb{P}(X = x) &= \sum_{y'} Cg(x)h(y) \\
&= \frac{\sum_{y'} g(x)h(y)}{\sum_{x'} \sum_{y'} g(x)h(y)} \\
&= \frac{g(x) \left(\sum_{y'} h(y) \right)}{\left(\sum_{x'} g(x) \right) \left(\sum_{y'} h(y) \right)} \\
&= \frac{g(x)}{\sum_{x'} g(x)}
\end{aligned}$$

Análogamente, podemos demostrar que:

$$\mathbb{P}(Y = y) = \frac{h(y)}{\sum_{y'} h(y')}$$

Por lo tanto:

$$\begin{aligned}
\mathbb{P}(X = x, Y = y) &= Cg(x)h(y) \\
&= \frac{g(x)h(y)}{\sum_{x'} \sum_{y'} g(x)h(y)} \\
&= \mathbb{P}(X = x) \mathbb{P}(Y = y)
\end{aligned}$$

Así que $X \perp Y$.

Antes de pasar al siguiente tema veamos el problema de la Aguja de Buffon. Consideremos que en una mesa tenemos líneas paralelas separadas por una distancia D . Luego, una aguja de largo $L < D$ se lanza sobre la mesa y se considera el evento A como “la aguja interseca una de las líneas”. Veamos cuál es la probabilidad de que esto suceda. Tomemos X como la variable aleatoria que mide la distancia entre el centro de la aguja y la línea paralela más cercana y θ como la variable aleatoria que mide el ángulo entre la aguja y el eje perpendicular a las líneas. Entonces, $X \sim \mathcal{U}\left(0, \frac{D}{2}\right)$ y $\Theta \sim \mathcal{U}\left(0, \frac{\pi}{2}\right)$. Consideremos $X \perp \theta$. Entonces, la condición necesaria para que la aguja interseque una línea es que:

$$X < \frac{L}{2} \cos(\Theta)$$

Entonces:

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}\left(X < \frac{L}{2} \cos(\Theta)\right) \\
&= \int_0^{\frac{\pi}{2}} \int_0^{\frac{D}{2}} \frac{1}{\pi} \frac{1}{D} \mathbb{I}_{\{x < \frac{L}{2} \cos(\theta)\}} dx d\theta \\
&= \frac{4}{\pi D} \int_0^{\frac{\pi}{2}} \int_0^{\frac{L}{2} \cos(\theta)} dx d\theta \\
&= \frac{4}{\pi D} \int_0^{\frac{\pi}{2}} \frac{L}{2} \cos(\theta) d\theta \\
&= \frac{2L}{\pi D}
\end{aligned}$$

Notemos entonces que:

$$\pi = \frac{2L}{D\mathbb{P}(A)}$$

Esto significa que si tiramos muchas veces una aguja (N veces) y contamos la cantidad de veces que la aguja interseca una línea (M veces) entonces podemos aproximar $\mathbb{P}(A) \approx \frac{M}{N}$ y entonces aproximar π como:

$$\pi \approx \frac{2LN}{DM}$$

Resulta que a medida que $N \rightarrow \infty$ entonces la estimación de π cada vez es mejor.

4.2. Esperanza y Covarianza

Al igual que veníamos haciendo, definimos la esperanza de la coordenada X_i de un vector aleatorio \mathbf{X} como:

$$E(X_i) = \int_{\mathbb{R}^N} X_i f(\mathbf{x}) d\mathbf{x}$$

Análogamente, sea $g : \mathbb{R}^N \rightarrow \mathbb{R}$ definimos el promedio de g como:

$$\langle g \rangle = \int_{\mathbb{R}^N} g(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$$

Ahora, un concepto nuevo que vamos a ver es la **Covarianza**. Sean X e Y variables aleatorias definimos la covarianza entre ambas como:

$$\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y)))$$

Notemos que $\text{Cov}(X, X) = V(X)$. Demostremos que $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$:

$$\begin{aligned} \text{Cov}(X, Y) &= E((X - E(X))(Y - E(Y))) \\ &= E(XY - E(Y)X - E(X)Y + E(X)E(Y)) \\ \text{Linealidad} \rightarrow &= E(XY) - E(Y)E(X) - \cancel{E(X)E(Y)} + \cancel{E(X)E(Y)} \\ &= E(XY) - E(Y)E(X) \end{aligned}$$

Notemos que si $X = Y$ entonces nos queda la expresión que conocemos de $V(X)$.

Lo que la covarianza mide es de cierta manera la relación que hay entre ambas variables aleatorias. Es decir, si hay una fuerte relación positiva entre X e Y (X es grande cuando Y es grande y X es chico cuando Y es chico) entonces la covarianza será positiva. Si hay una fuerte relación negativa entre X e Y (X es grande cuando Y es chico y viceversa) entonces la covarianza será negativa. En cambio, si no hay relación entre X e Y la covarianza es nula. En otras palabras, si $X \perp Y \implies \text{Cov}(X, Y) = 0$. Sin embargo, es importante mencionar que no vale la inversa. Es decir, $\text{Cov}(X, Y) \neq 0 \implies X \not\perp Y$. Demostremoslo:

$$\begin{aligned} E(XY) &= \int_{\mathbb{R}^2} xyf(x, y) dx dy \\ X \perp Y \implies f(x, y) = f_X(x)f_Y(y) &\rightarrow \int_{\mathbb{R}^2} xyf_X(x)f_Y(y) dx dy \\ &= \left(\int_{\mathbb{R}} xf_X(x) dx \right) \left(\int_{\mathbb{R}} yf_Y(y) dy \right) \\ &= E(X)E(Y) \end{aligned}$$

$$\begin{aligned} \therefore \text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\ &= 0 \end{aligned}$$

Veamos un ejemplo. Consideremos una distribución:

$$f(x, y) = \frac{6}{5}(x + y^2) \mathbb{I}_{\{(x, y) \in [0, 1] \times [0, 1]\}}(x, y)$$

Entonces:

$$\begin{aligned} E(X) &= \int_{\mathbb{R}^2} x f(x, y) \, dx dy \\ &= \int_0^1 \int_0^1 x \frac{6}{5} (x + y^2) \, dx dy \\ &= \frac{6}{5} \int_0^1 \int_0^1 (x^2 + xy^2) \, dx dy \\ &= \frac{6}{5} \int_0^1 \left(\frac{x^3}{3} + \frac{x^2 y^2}{2} \right) \Big|_{x=0}^{x=1} dy \\ &= \frac{6}{5} \int_0^1 \left(\frac{1}{3} + \frac{y^2}{2} \right) dy \\ &= \frac{6}{5} \left(\frac{y}{3} + \frac{y^3}{6} \right) \Big|_0^1 \\ &= \frac{6}{5} \left(\frac{1}{3} + \frac{1}{6} \right) \\ &= \frac{3}{5} \end{aligned}$$

$$\begin{aligned} E(Y) &= \int_{\mathbb{R}^2} y f(x, y) \, dx dy \\ &= \int_0^1 \int_0^1 y \frac{6}{5} (x + y^2) \, dx dy \\ &= \frac{6}{5} \int_0^1 \int_0^1 (xy + y^3) \, dx dy \\ &= \frac{6}{5} \int_0^1 \left(\frac{x^2}{2} y + xy^3 \right) \Big|_{x=0}^{x=1} dy \\ &= \frac{6}{5} \int_0^1 \left(\frac{y}{2} + y^3 \right) dy \\ &= \frac{6}{5} \left(\frac{y^2}{4} + \frac{y^4}{4} \right) \Big|_0^1 \\ &= \frac{6}{5} \left(\frac{1}{4} + \frac{1}{4} \right) \\ &= \frac{3}{5} \end{aligned}$$

Notemos que $E(X) = E(Y)$. Calculemos la covarianza:

$$\begin{aligned}
 \text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\
 &= \int_{\mathbb{R}^2} xyf(x, y) \, dx dy - \left(\frac{3}{5}\right) \left(\frac{3}{5}\right) \\
 &= \int_0^1 \int_0^1 xy \frac{6}{5} (x + y^2) \, dx dy - \frac{9}{25} \\
 &= \frac{6}{5} \int_0^1 \int_0^1 (x^2 y + xy^3) \, dx dy - \frac{9}{25} \\
 &= \frac{6}{5} \int_0^1 \left(\frac{x^3}{3} y + \frac{x^2}{2} y^3 \right) \Big|_{x=0}^{x=1} dy - \frac{9}{25} \\
 &= \frac{6}{5} \int_0^1 \left(\frac{y}{3} + \frac{y^3}{2} \right) dy - \frac{9}{25} \\
 &= \frac{6}{5} \left(\frac{y^2}{6} + \frac{y^4}{8} \right) \Big|_0^1 - \frac{9}{25} \\
 &= \frac{6}{5} \left(\frac{1}{6} + \frac{1}{8} \right) - \frac{9}{25} \\
 &= \frac{7}{20} - \frac{9}{25} \\
 &= -\frac{1}{100}
 \end{aligned}$$

Como la covarianza es negativa sabemos que X e Y están inversamente relacionadas pero como es un número pequeño ($\frac{1}{100} \ll \frac{3}{5}$) entonces sabemos que las variables están relativamente desvinculadas.

Veamos otro ejemplo. Consideremos ahora la siguiente distribución:

$$f(x, y) = \frac{3}{8} (x^2 + y^2) \mathbb{I}_{\{(x, y) \in [-1, 1] \times [-1, 1]\}}(x, y)$$

Notemos:

$$\begin{aligned}
 \int_{\mathbb{R}^2} f(x, y) &= \int_{-1}^1 \int_{-1}^1 \frac{3}{8} (x^2 + y^2) \, dx dy \\
 &= \frac{3}{8} \int_{-1}^1 \left(\frac{x^3}{3} + xy^2 \right) \Big|_{x=-1}^{x=1} dy \\
 &= \frac{3}{8} \int_{-1}^1 \left(\frac{2}{3} + 2y^2 \right) dy \\
 &= \frac{3}{8} \left(\frac{2}{3} y + \frac{2}{3} y^3 \right) \Big|_{-1}^1 \\
 &= \frac{3}{8} \left(\frac{4}{3} + \frac{4}{3} \right) \\
 &= 1
 \end{aligned}$$

Además, como $f(x, y) \geq 0 \forall (x, y) \in [-1, 1] \times [-1, 1]$ sabemos que f es una distribución válida. Notemos además que $f(x, y) = f(-x, y) = f(x, -y) = f(-x, -y)$. Demostremos que esta propiedad de la función nos asegura que

$$E(X) = E(Y) = E(XY) = 0:$$

$$\begin{aligned}
 E(X) &= \int_{\mathbb{R}^2} x f(x, y) \, dx dy \\
 &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^0 x f(x, y) \, dx + \int_0^{\infty} x f(x, y) \, dx \right) dy \\
 &\stackrel{u \stackrel{\text{def}}{=} -x}{\rightarrow} \int_{-\infty}^{\infty} \left(- \int_0^{\infty} u f(-u, y) \, du + \int_0^{\infty} x f(x, y) \, dx \right) dy \\
 &\stackrel{f(x, y) = f(-x, y)}{\rightarrow} \int_{-\infty}^{\infty} \left(- \int_0^{\infty} u f(u, y) \, du + \int_0^{\infty} x f(x, y) \, dx \right) dy \\
 &= 0 \\
 &\therefore E(X) = 0
 \end{aligned}$$

Análogamente podemos mostrar que como $f(x, y) = f(x, -y)$ entonces $E(Y) = 0$ y que como $f(x, y) = f(-x, -y)$ entonces $E(XY) = 0$. Por lo tanto, la covarianza es:

$$\begin{aligned}
 \text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\
 &= 0
 \end{aligned}$$

Este resultado nos tentaría a decir que $X \perp Y$. Sin embargo, recordemos que $X \perp Y \iff f(x, y) = f_X(x) f_Y(y)$, y en este caso no existen funciones f_X y f_Y tales que $f(x, y) = f_X(x) f_Y(y)$. Por lo tanto, $X \not\perp Y$ pero $\text{Cov}(X, Y) = 0$.

5. Ley de Grandes Números y Teorema Central del Límite

5.1. Definiciones

Finalmente empezamos a adentrarnos a la parte estadística de la materia, que en mi opinión, es la parte más divertida. Sea X una variable aleatoria definimos un **Modelo** de X a un vector \mathbf{X} de n variables aleatorias independientes e idénticamente distribuidas (o *iid*) donde cada componente X_i del modelo representa un ensayo de la variable X . A los valores numéricos \mathbf{x} del modelo de X lo llamamos una **Muestra** de X . Por ejemplo, consideremos a X la variable aleatoria que indica el resultado de tirar una moneda. Entonces, un modelo de X sería el vector:

$$\mathbf{X} = (X_1 \quad X_2 \quad X_3 \quad X_4 \quad X_5)$$

que representa 5 lanzadas de la moneda. Si al lanzar el dado 5 veces obtenemos dos veces cara, luego una vez cruz, luego otra vez cara, y luego otra vez cruz entonces la muestra \mathbf{x} del modelo de X sería:

$$\mathbf{x} = (1 \quad 1 \quad 0 \quad 1 \quad 0)$$

Los modelos y las muestras son el elemento principal y necesario para hacer estadística de una variable aleatoria.

Sea \mathbf{X} modelo de una variable aleatoria X definimos la **Media Muestral** de \mathbf{X} como:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Notemos que vale que $E(\bar{X}_n) = E(X)$ y $V(\bar{X}_n) = \frac{1}{n}V(X)$:

$$E(\bar{X}_n) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right)$$

$$\text{Linealidad} \rightarrow = \frac{1}{n} \sum_{i=1}^n E(X_i)$$

$$X_i \text{ son iid} \rightarrow = E(X)$$

$$V(\bar{X}_n) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right)$$
$$= \frac{1}{n^2} V\left(\sum_{i=1}^n X_i\right)$$

$$X_i \text{ son iid} \rightarrow = \frac{1}{n^2} \sum_{i=1}^n V(X)$$

$$= \frac{1}{n} V(X)$$

Definimos también la **Desviación Estándar** de \mathbf{X} como:

$$\sigma_n(\mathbf{X}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}$$

Además, llamamos **Error Estadístico** al coeficiente:

$$\text{Err}(\mathbf{X}) = \frac{\sigma_n(\mathbf{X})}{\sqrt{n}}$$

5.2. Ley de Grandes Números

Antes de ver la ley de grandes números veamos un par de propiedades necesarias para demostrarla y que nos van a ser útiles a la hora de resolver ejercicios.

Sea $X \geq 0$ una variable aleatoria con esperanza finita y sea $\varepsilon > 0$ entonces:

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}$$

Llamamos a esta propiedad la **Desigualdad de Markov**. Notemos:

$$\begin{aligned} E(X) &= \int_0^{\infty} x\mathbb{P}(X = x) dx \\ &= \int_0^{\varepsilon} x\mathbb{P}(X = x) dx + \int_{\varepsilon}^{\infty} x\mathbb{P}(X = x) dx \\ &\geq \int_{\varepsilon}^{\infty} x\mathbb{P}(X = x) dx \\ &\geq \int_{\varepsilon}^{\infty} \varepsilon\mathbb{P}(X = x) dx \\ &= \varepsilon \int_{\varepsilon}^{\infty} \mathbb{P}(X = x) dx \\ &= \varepsilon\mathbb{P}(X \geq \varepsilon) \\ \therefore \mathbb{P}(X \geq \varepsilon) &\leq \frac{E(X)}{\varepsilon} \end{aligned}$$

Sea X una variable aleatoria con esperanza finita y sea $\varepsilon > 0$ entonces:

$$\mathbb{P}(|X - E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}$$

Llamamos a esta propiedad la **Desigualdad de Chebyshev**. Tomemos $Y = |X - E(X)|$. Entonces:

$$\begin{aligned} \mathbb{P}(Y \geq \varepsilon) &= \mathbb{P}(|X - E(X)| \geq \varepsilon) \\ &= \mathbb{P}\left((X - E(X))^2 \geq \varepsilon^2\right) \\ \text{Desigualdad de Markov} &\rightarrow \leq \frac{E\left((X - E(X))^2\right)}{\varepsilon^2} \\ &= \frac{V(X)}{\varepsilon^2} \end{aligned}$$

Ahora si estamos listos para enunciar y demostrar la **Ley de Grandes Números**. Sea X una variable aleatoria y sea \mathbf{X} un modelo de X entonces:

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(|\bar{X}_n - E(X)| > \varepsilon\right) = 0 \forall \varepsilon > 0$$

Antes de pasar a demostrar esta ley frenémosnos a ver qué significa. Este límite nos dice que a medida que n , la cantidad de ensayos de X en el modelo, crece entonces la probabilidad de que la media muestral difiera de la esperanza tiende a cero. Es necesario remarcar que X y X_n no son la misma cosa: X es la variable aleatoria definida probabilísticamente y X_n es una variable aleatoria que toma un valor puntual al hacer un ensayo de X . Lo que la ley de grandes números nos dice es que si $n \rightarrow \infty$ (o sea, si hacemos infinitos ensayos) entonces $\bar{X}_n \rightarrow E(X)$. Esto

de cierta manera le da validez al proceso de hacer ensayos, ya que de no ser así de nada nos serviría hacer ensayos de X porque no nos diría nada de X . Es por este motivo que al definir a la esperanza la comparamos con la noción típica de promedio (o media), ya que a medida que hacemos muchos ensayos ambos convergen al mismo valor.

La demostración es simple. Notemos:

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - E(X)| > \varepsilon) &= \lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - E(\bar{X}_n)| > \varepsilon) \\ \text{Chebyshev} \rightarrow &\leq \lim_{n \rightarrow \infty} \frac{V(\bar{X}_n)}{\varepsilon^2} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \frac{V(X)}{\varepsilon^2} \\ &= 0 \end{aligned}$$

Veamos un ejemplo. Consideremos un experimento Bernoulli donde $X \sim \text{Be}(p)$. Entonces, veamos cuántas repeticiones del experimento deberían hacerse para que la media muestral difiera de p en menos de 0.01 con probabilidad mayor o igual a 0.95. Es decir, queremos hallar n tal que $\mathbb{P}(|\bar{X}_n - p| \geq 0.01) \leq 0.05$. Notemos:

$$\begin{aligned} \mathbb{P}(|\bar{X}_n - p| \geq 0.01) &= \mathbb{P}(|\bar{X}_n - E(X)| \geq 0.01) \\ &= \mathbb{P}(|\bar{X}_n - E(\bar{X}_n)| \geq 0.01) \\ &\leq \frac{V(\bar{X}_n)}{(0.01)^2} \\ &= \frac{10000}{n} V(X) \\ &= \frac{10000}{n} p(1-p) \\ &\leq 0.05 \end{aligned}$$

$$\therefore n \geq 200000p(1-p)$$

Notemos que $p(1-p) \leq 0.25$, así que $n \geq 50000$.

5.3. Función Generadora de Momentos

Sea X una variable aleatoria definimos el **Momento** de orden k de X como:

$$m_k = E(X^k),$$

siempre y cuando la esperanza exista ($m_k < \infty$). Al primer momento $E(X) = \mu$ lo llamamos *posición*, al segundo momento $E(X^2) = \sigma^2 + \mu^2$ lo llamamos *dispersión*, el tercer momento lo llamamos *asimetría* y el cuarto momento lo llamamos *kurtosis*. Además, definimos la **Función Generadora de Momentos** (FGM) como:

$$M_X(t) = E(\exp(tX)),$$

siempre y cuando exista $t \in (-h, h)$ para algún $h \in \mathbb{R}$. Esta es una condición técnica para que $M_X(t)$ sea diferenciable en $t = 0$.

Veamos algunas propiedades de la FGM:

$$\begin{aligned}
 M_X(t) &= E(\exp(tX)) \\
 &= E\left(\sum_{k=0}^{\infty} \frac{(tX)^k}{k!}\right) \\
 &= E\left(\sum_{k=0}^{\infty} \frac{t^k}{k!} X^k\right) \\
 &= \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X^k) \\
 &= \sum_{k=0}^{\infty} \frac{t^k}{k!} m_k
 \end{aligned}$$

Entonces, notemos que:

$$\begin{aligned}
 \left. \frac{d^n}{dt^n} M_X(t) \right|_{t=0} &= \left. \frac{d^n}{dt^n} \sum_{k=0}^{\infty} \frac{t^k}{k!} m_k \right|_{t=0} \\
 &= \left. \left(\frac{d^n}{dt^n} \sum_{k=0}^{n-1} \frac{t^k}{k!} m_k + \frac{d^n}{dt^n} \sum_{k=n}^{\infty} \frac{t^k}{k!} m_k \right) \right|_{t=0} \\
 &= \left. \left(0 + \sum_{k=n}^{\infty} \frac{\cancel{k!}}{(k-n)!} \frac{t^{k-n}}{\cancel{k!}} m_k \right) \right|_{t=0} \\
 &= \left. \sum_{k=n}^{\infty} \frac{t^{k-n}}{(k-n)!} m_k \right|_{t=0} \\
 &= \left. \sum_{k=0}^{\infty} \frac{t^k}{k!} m_{k+n} \right|_{t=0} \\
 &= m_n + \left. \sum_{k=1}^{\infty} \frac{t^k}{k!} m_{k+n} \right|_{t=0} \\
 &= m_n
 \end{aligned}$$

Es por esto que definimos la FGM de esta manera. Cada una de las derivadas nos devuelve uno de los momentos. Dada una variable aleatoria cualquiera la FGM nos define unívocamente todos los momentos (y viceversa).

Veamos algunos ejemplos de FGM para distribuciones usuales. Empecemos considerando la distribución exponencial. La FGM de la exponencial es:

$$M(t) = \frac{\lambda}{\lambda - t}$$

Notemos:

$$\begin{aligned}
 E(X) &= \left. \frac{d}{dt} M(t) \right|_{t=0} \\
 &= \left. \frac{\lambda}{(\lambda - t)^2} \right|_{t=0} \\
 &= \frac{1}{\lambda} \\
 &= \underbrace{\frac{1}{\lambda}}_{=\mu}
 \end{aligned}$$

$$\begin{aligned}
E(X^2) &= \left. \frac{d^2}{dt^2} M(t) \right|_{t=0} \\
&= \left. \frac{2\lambda}{(\lambda-t)^3} \right|_{t=0} \\
&= \frac{2}{\lambda^2} \\
&= \underbrace{\frac{1}{\lambda^2}}_{=\sigma^2} + \underbrace{\left(\frac{1}{\lambda}\right)^2}_{=\mu^2}
\end{aligned}$$

$$\begin{aligned}
E(X^3) &= \left. \frac{d^3}{dt^3} M(t) \right|_{t=0} \\
&= \left. \frac{6\lambda}{(\lambda-t)^4} \right|_{t=0} \\
&= \frac{6}{\lambda^3}
\end{aligned}$$

$$\begin{aligned}
E(X^4) &= \left. \frac{d^4}{dt^4} M(t) \right|_{t=0} \\
&= \left. \frac{24\lambda}{(\lambda-t)^5} \right|_{t=0} \\
&= \frac{24}{\lambda^4}
\end{aligned}$$

Notemos que en general para la distribución exponencial vale que:

$$m_k = \frac{k!}{\lambda^k}$$

Enunciemos y demostremos una propiedad de la FGM que ya mencioné antes. Dada una distribución de probabilidad la función generadora de momentos es única para esa distribución. Es decir, la FGM tiene la propiedad de unicidad: existe una única FGM para cada distribución y una única distribución para cada FGM (salvo a lo suma en un conjunto de probabilidad 0). Para demostrar esta propiedad consideremos una variable aleatoria X con rango $\mathcal{R}_X = [1, n] \subset \mathbb{N}$. Entonces:

$$\begin{aligned}
M(t) &= E(\exp(Xt)) \\
&= \sum_{i=1}^n \exp(it) \mathbb{P}(X=i) \\
z \stackrel{\text{def}}{=} \exp(t) &\rightarrow = \sum_{i=1}^n \mathbb{P}(X=i) z^i
\end{aligned}$$

Entonces $M(t)$ es un polinomio de $z = \exp(t)$ con coeficientes $a_i = \mathbb{P}(X=i)$. Definamos entonces a este polinomio como:

$$Q(z) = \sum_{i=1}^n a_i z^i$$

Notemos que Q y M contienen la misma información ya que son esencialmente la misma función. Sin embargo,

como Q es un polinomio sabemos que:

$$\begin{aligned}
 \left. \frac{d^k}{dz^k} Q(z) \right|_{z=0} &= \left. \frac{d^k}{dz^k} \sum_{i=1}^n a_i z^i \right|_{z=0} \\
 &= \left. \frac{d^k}{dz^k} \sum_{i=1}^{k-1} a_i z^i + \frac{d^k}{dz^k} \sum_{i=k}^n a_i z^i \right|_{z=0} \\
 &= 0 + \left. \sum_{i=k}^n a_i \frac{i!}{(i-k)!} z^{i-k} \right|_{z=0} \\
 &= k! a_k + \left. \sum_{i=k+1}^n a_i \frac{i!}{(i-k)!} z^{i-k} \right|_{z=0} \\
 &= k! a_k \\
 &= k! \mathbb{P}(X = k)
 \end{aligned}$$

$$\therefore \mathbb{P}(X = k) = \frac{1}{k!} \left. \frac{d^k}{dz^k} Q(z) \right|_{z=0} \quad \forall k \in \mathcal{R}_X$$

Por lo tanto, dada la FGM la distribución de probabilidad queda unívocamente definida. La inversa se demuestra fácilmente por definición.

Veamos algunas propiedades más de la FGM. Sean X_i variables aleatorias independientes entonces:

$$M_{\sum_i X_i}(t) = \prod_i M_{X_i}(t)$$

Notemos:

$$\begin{aligned}
 M_{\sum_i X_i}(t) &= E \left(\exp \left(t \sum_i X_i \right) \right) \\
 &= E \left(\prod_i \exp(t X_i) \right) \\
 X_i \perp X_j \forall i \neq j &\rightarrow \prod_i E(\exp(t X_i)) \\
 &= \prod_i M_{X_i}(t)
 \end{aligned}$$

Otras propiedades son:

- $M_{aX+b}(t) = \exp(tb) M_X(at)$.
- $Z \sim \mathcal{N}(0, 1) \implies M_Z(t) = \exp\left(\frac{t^2}{2}\right)$.
- $X \sim \mathcal{N}(\mu, \sigma^2) \implies M_X(t) = \exp(t\mu) \exp\left(\frac{\sigma^2 t^2}{2}\right)$.
- Si X_i son variables aleatorias iid con media μ y varianza σ^2 entonces si $S_n = \sum_i X_i \implies M_{S_n}(t) = (M_X(t))^n$.
- $T_n = \frac{1}{\sqrt{n}} S_n \implies M_{T_n}(t) = \left(M_X\left(\frac{t}{\sqrt{n}}\right) \right)^n$.

5.4. Teorema Central del Límite

Antes de enunciar el teorema central del límite veamos el concepto de convergencia en distribución. Sea $\{X_i\}$ una sucesión de variables aleatorias decimos que **Converge en Distribución** a F_Y si:

$$\lim_{n \rightarrow \infty} F_{Y_n}(y) = F_Y(y)$$

Esto ya lo vimos cuando enunciamos la propiedad de la distribución exponencial que es la “versión continua” de la distribución geométrica. Vimos que si definimos la sucesión de variables aleatorias $\{\frac{X_n}{n}\}$ con $X_n \sim \mathcal{G}\left(\frac{\lambda}{n}\right)$ entonces $\{\frac{X_n}{n}\}$ converge en distribución a $\mathcal{E}(\lambda)$:

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{X_n}{n} \geq x\right) &= \lim_{n \rightarrow \infty} \mathbb{P}(X_n \geq nx) \\ &= \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{nx} \\ &= \lim_{n \rightarrow \infty} \left(1 + \frac{1}{(-\frac{n}{\lambda})}\right)^{\left(-\frac{n}{\lambda}\right)(-\lambda x)} \\ &= \exp(-\lambda x) \end{aligned}$$

Ahora si podemos enunciar el teorema. Sean $X_i, i \in [1, n] \subset \mathbb{N}$ variables aleatorias iid con media μ y varianza σ^2 y sea $S_n = \sum_i X_i$ entonces definiendo $Z_n \stackrel{\text{def}}{=} \frac{S_n - n\mu}{\sigma\sqrt{n}}$ resulta que Z_n converge en distribución a una normal estándar $\mathcal{N}(0, 1)$. Antes de demostrar este teorema veamos qué significa. Este teorema es importante para hacer estadística porque nos dice que si armamos un modelo \mathbf{X} de una variable aleatoria X cualquiera (cuya distribución es desconocida) entonces siempre y cuando tengamos suficientes muestras ($n \rightarrow \infty$) entonces podemos armarnos una variable aleatoria normal estándar como:

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

que nos va a ser útil para hacer cálculos de probabilidad incluso cuando desconocemos la distribución de X . Si queremos hacer cálculos de probabilidades en X sin saber su distribución simplemente hacemos cálculos sobre Z_n y de allí sacamos conclusiones sobre X .

Demostremos este teorema. Consideremos que existe la FGM de X_i y notémosla como M . Tomemos también $\mu = 0$ y $\sigma = 1$ para simplificar (después vamos a ver cómo se extiende). Entonces, calculemos la FGM de Z_n en función de M :

$$\begin{aligned} M_{Z_n}(t) &= E(\exp(tZ_n)) \\ &= E\left(\exp\left(\frac{t}{\sqrt{n}} \sum_{i=1}^n X_i\right)\right) \\ &= E\left(\prod_{i=1}^n \exp\left(\frac{t}{\sqrt{n}} X_i\right)\right) \\ X \text{ iid} \rightarrow &= \prod_{i=1}^n E\left(\exp\left(\frac{t}{\sqrt{n}} X_i\right)\right) \\ M(t) = E(\exp(tX)) \rightarrow &= M^n\left(\frac{t}{\sqrt{n}}\right) \end{aligned}$$

Entonces, veamos cómo es la FGM de Z_n a medida que aumenta la cantidad de muestras. Es decir, veamos cuál es

el límite de $M_{Z_n}(t)$ cuando $n \rightarrow \infty$:

$$\begin{aligned}
 \lim_{n \rightarrow \infty} M_{Z_n}(t) &= \lim_{n \rightarrow \infty} M^n \left(\frac{t}{\sqrt{n}} \right) \\
 &= \exp \left(\lim_{n \rightarrow \infty} \log \left(M^n \left(\frac{t}{\sqrt{n}} \right) \right) \right) \\
 &= \exp \left(\lim_{n \rightarrow \infty} n \log \left(M \left(\frac{t}{\sqrt{n}} \right) \right) \right) \\
 M(0) = 1 \rightarrow &= \exp \left(\lim_{n \rightarrow \infty} \frac{\overbrace{\log \left(M \left(\frac{t}{\sqrt{n}} \right) \right)}^{\rightarrow 0}}{\underbrace{\frac{1}{n}}_{\rightarrow 0}} \right) \\
 \text{L'Hôpital} \rightarrow &= \exp \left(\lim_{n \rightarrow \infty} \frac{\frac{t}{2n^{\frac{3}{2}}} M' \left(\frac{t}{\sqrt{n}} \right)}{\frac{1}{n^2} M \left(\frac{t}{\sqrt{n}} \right)} \right) \\
 M'(0) = \mu = 0 \rightarrow &= \exp \left(\lim_{n \rightarrow \infty} \frac{t}{2} \frac{\overbrace{M' \left(\frac{t}{\sqrt{n}} \right)}^{\rightarrow 0}}{\underbrace{\frac{1}{\sqrt{n}} M \left(\frac{t}{\sqrt{n}} \right)}_{\rightarrow 0}} \right) \\
 \text{L'Hôpital} \rightarrow &= \exp \left(\lim_{n \rightarrow \infty} \frac{t}{2} \frac{\frac{t}{2n^{\frac{3}{2}}} M'' \left(\frac{t}{\sqrt{n}} \right)}{\frac{1}{2n^{\frac{3}{2}}} M \left(\frac{t}{\sqrt{n}} \right) + \frac{t}{2n^{\frac{3}{2}}} M' \left(\frac{t}{\sqrt{n}} \right)} \right) \\
 &= \exp \left(\lim_{n \rightarrow \infty} \frac{t^2}{2} \frac{M'' \left(\frac{t}{\sqrt{n}} \right)}{M \left(\frac{t}{\sqrt{n}} \right) + t M' \left(\frac{t}{\sqrt{n}} \right)} \right) \\
 M''(0) = \sigma^2 + \mu^2 = 1 \rightarrow &= \exp \left(\frac{t^2}{2} \right)
 \end{aligned}$$

O sea, que la FGM de Z_n tiende a $M_{Z_n} = \exp \left(\frac{t^2}{2} \right)$. Esta es la FGM de una distribución normal, y como la FGM determina la distribución eso significa que Z_n es una variable normal estándar. Esto es lo que el teorema central del límite establece. Notemos que si $\mu \neq 0$ y/o $\sigma \neq 1$ entonces lo único que hay que hacer es estandarizar la variable aleatoria tomando:

$$Y_i = \frac{X_i - \mu}{\sigma}$$

Consideremos un ejemplo. Supongamos que sumamos números en una calculadora y que aproxima a cada número por el entero más próximo. Por lo tanto, asumamos que los errores de aproximación son independientes y con distribución $U \left(-\frac{1}{2}, \frac{1}{2} \right)$. Entonces, si sumamos 1500 números, ¿cuál es la probabilidad de que el valor absoluto del error exceda 15? Tomemos X_i como el error correspondiente al i -ésimo sumando y tomemos T_n como el error total de una suma de n sumandos. Vale que:

$$T_n = \sum_{i=1}^n X_i$$

Queremos calcular $\mathbb{P}(|T_{1500}| > 15)$. Notemos:

$$\begin{aligned}
 \mu &= \frac{-\frac{1}{2} + \frac{1}{2}}{2} \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}\sigma^2 &= \frac{\left(\frac{1}{2} - \left(-\frac{1}{2}\right)\right)^2}{12} \\ &= \frac{1}{12}\end{aligned}$$

Entonces:

$$\begin{cases} E(T_n) = 0 \\ V(T_n) = \frac{n}{12} \end{cases}$$

Por lo tanto, si tomamos $Z_n = \frac{T_n}{\sqrt{\frac{n}{12}}}$ tenemos que:

$$\begin{aligned}\mathbb{P}(|T_{1500}| > 15) &= \mathbb{P}\left(\left|\sqrt{\frac{1500}{12}}Z_{1500}\right| > 15\right) \\ &= \mathbb{P}\left(\left(\sqrt{125}Z_{1500} < -15\right) \cup \left(\sqrt{125}Z_{1500} > 15\right)\right) \\ &= \mathbb{P}(Z_{1500} < -1.34) + \mathbb{P}(Z_{1500} > 1.34) \\ \text{Mirando la tabla} \rightarrow &= 2(1 - 0.9099) \\ &= 0.1802\end{aligned}$$

Ahora veamos, ¿cuántos números n pueden sumarse a fin de que $\mathbb{P}(|T_n| \leq 10) \geq 0.9$? Notemos:

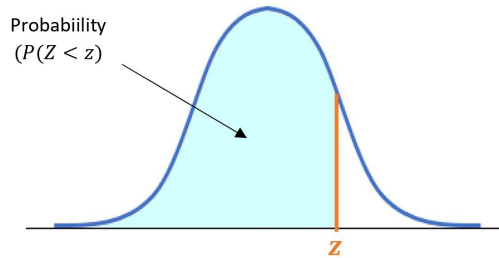
$$\begin{aligned}\mathbb{P}(|T_n| \leq 10) &= \mathbb{P}(-10 \leq T_n \leq 10) \\ &= \mathbb{P}\left(-10 \leq \sqrt{\frac{n}{12}}Z_n \leq 10\right) \\ &= \mathbb{P}\left(-20\sqrt{\frac{3}{n}} \leq Z_n \leq 20\sqrt{\frac{3}{n}}\right) \\ &\geq 0.9\end{aligned}$$

Entonces, buscamos en la tabla qué valor máximo de z_p encontramos cuya integral valga menos que 0.95 (ya que en estamos acotando por ambos extremos). Resulta que $z_{0.9495} = 1.64$, así que:

$$\begin{aligned}1.64 &\leq 20\sqrt{\frac{3}{n}} \\ n &\leq \left(\frac{20\sqrt{3}}{1.64}\right)^2 \\ n &\leq 446\end{aligned}$$

Parte II

Estadística



| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5754 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7258 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7518 | 0.7549 |
| 0.7 | 0.7580 | 0.7612 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7996 | 0.8023 | 0.8051 | 0.8079 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9430 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9485 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| 1.8 | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9700 | 0.9706 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9762 | 0.9767 |
| 2.0 | 0.9773 | 0.9778 | 0.9783 | 0.9788 | 0.9793 | 0.9798 | 0.9803 | 0.9808 | 0.9812 | 0.9817 |
| 2.1 | 0.9821 | 0.9826 | 0.9830 | 0.9834 | 0.9838 | 0.9842 | 0.9846 | 0.9850 | 0.9854 | 0.9857 |
| 2.2 | 0.9861 | 0.9865 | 0.9868 | 0.9871 | 0.9875 | 0.9878 | 0.9881 | 0.9884 | 0.9887 | 0.9890 |
| 2.3 | 0.9893 | 0.9896 | 0.9898 | 0.9901 | 0.9904 | 0.9906 | 0.9909 | 0.9911 | 0.9913 | 0.9916 |
| 2.4 | 0.9918 | 0.9920 | 0.9922 | 0.9925 | 0.9927 | 0.9929 | 0.9931 | 0.9932 | 0.9934 | 0.9936 |
| 2.5 | 0.9938 | 0.9940 | 0.9941 | 0.9943 | 0.9945 | 0.9946 | 0.9948 | 0.9949 | 0.9951 | 0.9952 |
| 2.6 | 0.9953 | 0.9955 | 0.9956 | 0.9957 | 0.9959 | 0.9960 | 0.9961 | 0.9962 | 0.9963 | 0.9964 |
| 2.7 | 0.9965 | 0.9966 | 0.9967 | 0.9968 | 0.9969 | 0.9970 | 0.9971 | 0.9972 | 0.9973 | 0.9974 |
| 2.8 | 0.9974 | 0.9975 | 0.9976 | 0.9977 | 0.9977 | 0.9978 | 0.9979 | 0.9980 | 0.9980 | 0.9981 |
| 2.9 | 0.9981 | 0.9982 | 0.9983 | 0.9983 | 0.9984 | 0.9984 | 0.9985 | 0.9985 | 0.9986 | 0.9986 |
| 3.0 | 0.9987 | 0.9987 | 0.9987 | 0.9988 | 0.9988 | 0.9989 | 0.9989 | 0.9989 | 0.9990 | 0.9990 |
| 3.1 | 0.9990 | 0.9991 | 0.9991 | 0.9991 | 0.9992 | 0.9992 | 0.9992 | 0.9992 | 0.9993 | 0.9993 |
| 3.2 | 0.9993 | 0.9993 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9995 | 0.9995 | 0.9995 |
| 3.3 | 0.9995 | 0.9995 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9997 |
| 3.4 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9998 | 0.9998 |

6. Estadística Paramétrica

6.1. Estimación Puntual

La **Estadística Paramétrica** consiste en fijar un modelo \mathbf{X} de una variable aleatoria X y buscar los parámetros de la distribución. Para eso, vamos a definir al **Estimador Puntual** $\hat{\theta}_n$ como una variable aleatoria función de \mathbf{X} que determina los parámetros de la distribución. Si al estimador lo aplicamos sobre una muestra el resultado es la **Estimación**. Existen distintos métodos para armar estimadores puntuales que cumplan las condiciones que esperamos de un estimador.

6.1.1. Método de Momentos

Consideremos la variable aleatoria $X \sim \text{Be}(p)$. Entonces, definamos al estimador \hat{p}_n como una variable aleatoria que vamos a usar para estimar p . Tenemos n muestras que forman un modelo \mathbf{X} , así que podemos definir a \hat{p}_n como:

$$\begin{aligned}\hat{p}_n &= \frac{1}{n} \sum_{i=1}^n X_i \\ &= \bar{X}_n\end{aligned}$$

Esto tiene sentido ya que:

$$\begin{aligned}\lim_{n \rightarrow \infty} \bar{X}_n &= E(X) \\ &= p\end{aligned}$$

Llamamos a un estimador $\hat{\theta}_n$ **Consistente** si:

$$\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta,$$

donde θ es el parámetro que queremos estimar. Definimos también el **Sesgo** entre un estimador y el parámetro como:

$$b(\hat{\theta}_n) = E(\hat{\theta}_n) - \theta$$

Con esta definición del sesgo podemos definir a un estimador como **Insesgado** si $b(\hat{\theta}_n) = 0$ o **Asintóticamente Insesgado** si $b(\hat{\theta}_n) \rightarrow 0$.

Veamos un ejemplo. Consideremos la variable aleatoria $X \sim \mathcal{E}(\lambda)$ y estimemos λ . Definamos:

$$\hat{\lambda}_n = \frac{1}{\bar{X}_n}$$

Notemos:

$$\begin{aligned}b(\hat{\lambda}_n) &= E(\hat{\lambda}_n) - \lambda \\ &= E\left(\frac{1}{\bar{X}_n}\right) - \lambda \\ &= \frac{1}{E(X)} - \lambda \\ &= \lambda - \lambda \\ &= 0\end{aligned}$$

Así que sabemos que $\hat{\lambda}_n$ es un estimador insesgado. A su vez, es claro que $\hat{\lambda}_n \rightarrow \lambda$. Esta forma de construir estimadores en base a la media muestral y desviación estándar se llama **Método de Momentos**, ya que utilizamos los momentos muestrales de X para armarnos de un estimador.

Veamos ahora otro ejemplo. Consideremos que $X \sim \mathcal{U}(0, \theta)$ y estimemos θ . Como $E(X) = \frac{a+b}{2}$ podemos usar el método de momentos y definir $\hat{\theta}_n = 2\bar{X}_n$. Notemos:

$$\begin{aligned} b(\hat{\theta}_n) &= E(\hat{\theta}_n) - \theta \\ &= E(2\bar{X}_n) - \theta \\ &= 2E(X) - \theta \\ &= 2 \cdot \frac{0 + \theta}{2} - \theta \\ &= 0 \end{aligned}$$

Este estimador sería insesgado. Sin embargo, también podemos definir al estimador como:

$$\hat{\theta}_n = \text{máx}(\{X_i\})$$

Supongamos que $\text{máx}(\mathbf{X}) = X_j$. Entonces:

$$\begin{aligned} E(\hat{\theta}_n) &= E(\text{máx}(\{X_i\})) \\ &= \int_0^\theta x \mathbb{P}(x = \text{máx}(\{x\})) dx \\ \delta_D \text{ es la Delta de Dirac} \rightarrow &= \int_{-\infty}^\infty x \delta_D(x - \theta) dx \\ &= \theta \\ \therefore b(\hat{\theta}_n) &= 0 \end{aligned}$$

Este estimador también es insesgado y también cumple que $\hat{\theta}_n \rightarrow \theta$.

Veamos un ejemplo más. Ahora consideremos que $X \sim \mathcal{U}(-\theta, \theta)$ y estimemos θ . Notemos que ahora no podemos usar el primer momento para definir al estimador, ya que $E(X) = 0$. Entonces, usemos el segundo momento. Sabemos que $V(X) = \frac{\theta^2}{3}$ así que:

$$\hat{\theta}_n = \sqrt{\frac{3}{n} \sum_{i=1}^n X_i^2}$$

Es claro que este estimador es insesgado y que tiende al parámetro.

6.1.2. Método de Máxima Verosimilitud

El **Método de Máxima Verosimilitud** es un método alternativo al método de momentos para construir estimadores puntuales. Fue creado por Fisher en los 1920s. Este método consiste en usar el modelo que tenemos de la variable aleatoria para buscar la forma del estimador que vuelva más verosímil al modelo obtenido. Es decir, queremos que el estimador tenga la forma que haga que la probabilidad de haber obtenido el modelo que obtuvimos sea máxima. Para mostrar mejor cómo funciona este método veamos algunos ejemplos.

Supongamos que le preguntamos a n personas si votarían a A o B candidato en las próximas elecciones presidenciales. Queremos estimar la probabilidad p de que voten al candidato A . Entonces, tenemos una variable aleatoria $X \sim \text{Be}(p)$ y tenemos un modelo de X , \mathbf{X} . Tenemos que la probabilidad de haber obtenido el modelo \mathbf{X} es:

$$\begin{aligned} \mathbb{P}(\mathbf{X} = \mathbf{x}) &= \mathbb{P}((X_1 = x_1) \cap (X_2 = x_2) \cap \dots \cap (X_n = x_n)) \\ X \text{ es i.i.d.} \rightarrow &= \prod_{i=1}^n \mathbb{P}(X_i = x_i) \\ x_i \in \{0, 1\} \rightarrow &= \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} \end{aligned}$$

Entonces, busquemos p tal que $\mathbb{P}(\mathbf{X} = \mathbf{x})$ sea máximo. Es decir, busquemos el valor de p que hace que nuestro modelo sea el más verosímil (el modelo “más esperable” de obtener). Para eso, definamos al estimador \hat{p}_n tal que $\frac{\partial \mathbb{P}(\mathbf{X} = \mathbf{x})}{\partial p} = 0$. Para simplificar esta cuenta tomemos el logaritmo de $\mathbb{P}(\mathbf{X} = \mathbf{x})$. Definamos $\mathcal{L} = \ln(\mathbb{P}(\mathbf{X} = \mathbf{x}))$:

$$\begin{aligned}
 \frac{\partial \mathcal{L}}{\partial p} &= \frac{\partial}{\partial p} \ln \left(\prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} \right) \\
 &= \frac{\partial}{\partial p} \sum_{i=1}^n \ln \left(p^{x_i} (1-p)^{1-x_i} \right) \\
 &= \frac{\partial}{\partial p} \sum_{i=1}^n (x_i \ln(p) + (1-x_i) \ln(1-p)) \\
 &= \sum_{i=1}^n \left(\frac{x_i}{p} - \frac{1-x_i}{1-p} \right) \\
 &= \frac{1}{p} \sum_{i=1}^n x_i + \frac{1}{1-p} \sum_{i=1}^n x_i - \frac{n}{1-p} \\
 &= \frac{n}{p} \bar{X}_n + \frac{n}{1-p} \bar{X}_n - \frac{n}{1-p} \\
 &= \frac{n(1-p) + np}{p(1-p)} \bar{X}_n - \frac{n}{1-p} \\
 &= \frac{n}{p(1-p)} \bar{X}_n - \frac{n}{1-p} \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 \therefore \frac{n}{p(1-p)} \bar{X}_n &= \frac{n}{1-p} \\
 \cancel{(1-p)n} \bar{X}_n &= \cancel{np(1-p)} \\
 \bar{X}_n &= p
 \end{aligned}$$

Por lo tanto, definimos a nuestro estimador como:

$$\hat{p}_n = \bar{X}_n$$

En este caso el resultado de este método es el mismo que con el método de momentos. Sin embargo, en otros casos este método puede devolvernos un estimador distinto.

Ahora consideremos que $X \sim \mathcal{U}(0, \theta)$ y busquemos estimar θ . Este problema ya lo resolvimos antes usando el método de momentos. Notemos:

$$\begin{aligned}
 \mathbb{P}(X = x) &= \frac{1}{\theta} \mathbb{I}_{(0, \theta)}(x) \\
 &= \begin{cases} \frac{1}{\theta} & x \in [0, \theta] \\ 0 & x \notin [0, \theta] \end{cases}
 \end{aligned}$$

$$\begin{aligned}
 \therefore \mathcal{L} &= \ln \left(\prod_{i=1}^n \frac{1}{\theta} \right) \mathbb{I}_{\{X_i \leq \theta \forall i\}}(X_i) \\
 &= -n \ln(\theta) \mathbb{I}_{\{\max(\{X_i\}) \leq \theta\}}(X)
 \end{aligned}$$

Notemos que \mathcal{L} es un función decreciente de θ . Por lo tanto, si queremos maximizar θ entonces vamos a tomar el menor valor posible de θ . Ya que $\max(\{X_i\}) \leq \theta$ tomemos al estimador de máxima verosimilitud como $\hat{\theta}_n = \max(\{X_i\})$. Este estimador ya lo habíamos propuesto anteriormente, solo que sin un fundamento riguroso como ahora.

Sea $\hat{\theta}_n$ un estimador de θ llamamos **Error** a la diferencia entre la estimación y el parámetro:

$$\text{Err}(\hat{\theta}_n) = \hat{\theta}_n - \theta$$

Definimos el **Error Estándar** como la desviación estándar de la estimación $(\sigma(\hat{\theta}_n))$ y definimos el **Error Cuadrático Medio** como:

$$\text{ECM}(\hat{\theta}_n) = E\left((\hat{\theta}_n - \theta)^2\right)$$

Notemos:

$$\begin{aligned} \text{ECM}(\hat{\theta}_n) &= E\left((\hat{\theta}_n - \theta)^2\right) \\ &= E\left((\hat{\theta}_n - E(\hat{\theta}_n) + E(\hat{\theta}_n) - \theta)^2\right) \\ &= E\left(\left((\hat{\theta}_n - E(\hat{\theta}_n)) + (E(\hat{\theta}_n) - \theta)\right)^2\right) \\ &= E\left((\hat{\theta}_n - E(\hat{\theta}_n))^2 + 2(\hat{\theta}_n - E(\hat{\theta}_n))(E(\hat{\theta}_n) - \theta) + (E(\hat{\theta}_n) - \theta)^2\right) \\ &= E\left((\hat{\theta}_n - E(\hat{\theta}_n))^2\right) + 2E\left((\hat{\theta}_n - E(\hat{\theta}_n))(E(\hat{\theta}_n) - \theta)\right) + E\left((E(\hat{\theta}_n) - \theta)^2\right) \\ &= \sigma^2(\hat{\theta}_n) + 2E\left(\hat{\theta}_n E(\hat{\theta}_n) - \hat{\theta}_n \theta - (E(\hat{\theta}_n))^2 + E(\hat{\theta}_n)\theta\right) + b^2(\hat{\theta}_n) \\ &= \sigma^2(\hat{\theta}_n) + 2\left(E(\hat{\theta}_n)E(\hat{\theta}_n) - E(\hat{\theta}_n)\theta - E(\hat{\theta}_n)E(\hat{\theta}_n) + E(\hat{\theta}_n)\theta\right) + b^2(\hat{\theta}_n) \\ &= \sigma^2(\hat{\theta}_n) + 2E(\hat{\theta}_n) \underbrace{(E(\hat{\theta}_n) - \theta - E(\hat{\theta}_n) + \theta)}_{=0} + b^2(\hat{\theta}_n) \\ &= \sigma^2(\hat{\theta}_n) + b^2(\hat{\theta}_n) \end{aligned}$$

Por último, definamos al **Estimador de Varianza** o **Varianza Muestral**:

$$\hat{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Notemos que $\hat{S}_n^2 = \frac{n}{n-1}\sigma^2$. Esta diferencia menor se debe a lo siguiente:

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

$$\begin{aligned}
E(\hat{S}_n^2) &= \frac{1}{n-1} \sum_{i=1}^n E((X_i - \bar{X}_n)^2) \\
&= \frac{1}{n-1} \sum_{i=1}^n E(X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2) \\
&= \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n E(X_i^2) - \frac{2}{n} \sum_{i=1}^n E(X_i\bar{X}_n) + E(\bar{X}_n^2) \right) \\
\text{Ojo con los índices repetidos} \rightarrow &= \frac{n}{n-1} \left(E(X^2) - \frac{2}{n^2} \sum_{i=1}^n \sum_{j=1}^n E(X_i X_j) \left(\frac{1}{2} + \frac{1}{2} \delta_{ij} \right) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n E(X_i X_j) \right) \\
&= \frac{n}{n-1} \left(\sigma^2 - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n E(X_i X_j) - \frac{1}{n^2} \sum_{i=1}^n E(X_i^2) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n E(X_i X_j) \right) \\
&= \frac{n}{n-1} \left(\sigma^2 - \frac{1}{n} \sigma^2 \right) \\
&= \sigma^2
\end{aligned}$$

$$\begin{aligned}
\therefore b(\hat{S}_n^2) &= E(\hat{S}_n^2) - \sigma^2 \\
&= \sigma^2 - \sigma^2 \\
&= 0
\end{aligned}$$

O sea que el estimador de varianza definido de esta manera rara tiene la propiedad de ser insesgado. Además:

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

Notemos:

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n X_i^2 &\rightarrow \mu^2 + \sigma^2, \quad \bar{X}_n^2 \rightarrow \mu^2, \quad \frac{n}{n-1} \rightarrow 1 \\
\therefore \hat{S}_n^2 &\rightarrow 1 \cdot (\mu^2 + \sigma^2 - \mu^2) \\
&= \sigma^2
\end{aligned}$$

O sea que el estimador de varianza además es consistente.

6.2. Intervalo de Confianza

Vimos que una forma de determinar los parámetros de una distribución de probabilidad es usando un estimador puntual. Otro modo es reemplazar la estimación puntual del parámetro por un intervalo de valores posibles. Llamamos a este intervalo de posibles valores del parámetro un **Intervalo de Confianza**. Todo intervalo de confianza tiene cierto nivel de “confianza”. Llamamos confianza a la probabilidad de que el parámetro esté dentro de ese

intervalo. Es decir, si θ es el parámetro que queremos estimar entonces el intervalo de confianza p de θ es I si $\mathbb{P}(\theta \in I) = p$. Por lo general, se define α como la **Probabilidad de Error** y entonces se define al intervalo como $\mathbb{P}(\theta \in I_{1-\alpha}) = 1 - \alpha$. Veamos un ejemplo.

Consideremos una variable aleatoria $X \sim \mathcal{N}(\mu, \sigma^2)$ con parámetro μ desconocido. Sabemos que si tenemos un modelo \mathbf{X} de X podemos definir:

$$Z = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$$

Usando la tabla normal podemos ver que $\mathbb{P}(-1.96 \leq Z \leq 1.96) = 0.95$. Notemos:

$$\begin{aligned} \mathbb{P}(-1.96 \leq Z \leq 1.96) &= \mathbb{P}\left(-1.96 \leq \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96\right) \\ &= \mathbb{P}\left(-1.96 \frac{\sigma}{\sqrt{n}} \leq \bar{X}_n - \mu \leq 1.96 \frac{\sigma}{\sqrt{n}}\right) \\ &= \mathbb{P}\left(\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}}\right) \\ &= 0.95 \end{aligned}$$

Notemos entonces que esto significa que la probabilidad de que μ esté en el intervalo $\left[\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}}\right]$ es de 0.95. Definimos entonces el intervalo:

$$I = \left[\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}}\right]$$

como el intervalo de confianza para μ de confianza 0.95 (o con probabilidad de error $\alpha = 0.05$).

En general, para armar intervalos de confianza se utiliza un método conocido como el **Método del Pivote**. Sea θ el parámetro que queremos encerrar en un intervalo y sea \mathbf{X} un modelo de la variable aleatoria X , el método consiste en encontrar una función $T(\mathbf{X}, \theta)$ tal que la distribución de T no dependa de θ . Entonces, podemos construir al intervalo como $\mathbb{P}(T(\mathbf{X}, \theta) \in I) = 1 - \alpha$. Veamos un ejemplo.

Consideremos la variable aleatoria $X \sim \mathcal{E}(\lambda)$. Vimos que una propiedad de la distribución gamma es que la suma de variables exponenciales sigue una distribución gamma. Es decir:

$$\sum_{i=1}^n X_i \sim \Gamma(n, \lambda)$$

Además, vimos que la distribución gamma cumple que:

$$X \sim \Gamma(n, \lambda) \implies \mu X \sim \Gamma\left(n, \frac{\lambda}{\mu}\right)$$

Entonces, podemos definir:

$$\begin{aligned} T(\mathbf{X}, \lambda) &= \lambda \sum_{i=1}^n X_i \\ &= \lambda n \bar{X}_n \\ &\sim \Gamma(n, 1) \end{aligned}$$

Notemos que esta función pivote cumple que es función de \mathbf{X} y de λ y además cumple que su distribución no depende de λ . Por lo tanto, la podemos usar como función pivote:

$$\begin{aligned} \mathbb{P}(T(\mathbf{X}, \lambda) \in [z_{\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}]) &= \mathbb{P}(\lambda n \bar{X}_n \in [z_{\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}]) \\ &= \mathbb{P}\left(\lambda \in \left[\frac{z_{\frac{\alpha}{2}}}{n \bar{X}_n}, \frac{z_{1-\frac{\alpha}{2}}}{n \bar{X}_n}\right]\right) \\ &= 1 - \alpha \end{aligned}$$

$$\therefore I = \left[\frac{z_{\frac{\alpha}{2}}}{n\bar{X}_n}, \frac{z_{1-\frac{\alpha}{2}}}{n\bar{X}_n} \right]$$

Notemos que en este caso como estamos usando la distribución gamma $z_{\frac{\alpha}{2}}$ es el percentil $\frac{\alpha}{2}$ de $\Gamma(n, 1)$ y $z_{1-\frac{\alpha}{2}}$ es el percentil $1 - \frac{\alpha}{2}$ de $\Gamma(n, 1)$.

7. Tests de Hipótesis

7.1. Tests Paramétricos

Antes de adentrarnos a describir los tests de hipótesis veamos de qué tratan con un ejemplo. Consideremos que estamos decidiendo irnos de viaje a una isla donde habitan dos distintas tribus de aborígenes: una tribu pacífica y amigable donde la altura promedio de la población es de 170 cm y otra tribu hostil de caníbales donde la altura promedio de la población es de 150 cm. Un explorador acaba de llegar a la isla y nos cuenta que vio a 9 aborígenes y anotó sus alturas. Tomemos X como la altura de uno de los aborígenes y \mathbf{X} como el modelo de $n = 9$ aborígenes que observó el explorador. Queremos estimar μ , la altura promedio de los aborígenes, para ver si la tribu es amigable o no y en base a eso decidir si vamos a viajar a esta isla. Para eso, vamos a definir una altura media crítica H donde si $\bar{X}_9 < H$ no viajamos y si $\bar{X}_9 \geq H$ viajamos. Como estamos trabajando con una muestra de 9 aborígenes no podemos saber con seguridad si la tribu es amigable o no. Por eso, debemos definir una probabilidad α que vamos a interpretar como las chances de que la tribu sea de caníbales pero que toleramos como para viajar de todas formas. Es decir, si las probabilidades que la tribu sea de caníbales es de α o menos entonces vamos a viajar, estamos cómodos con esas chances (α en principio debería ser un valor pequeño de probabilidad si valoramos nuestras vidas). Entonces, vamos a buscar un intervalo de confianza para μ con probabilidad de error α . Definamos la *Hipótesis Nula* H_0 como “la tribu es hostil”. Es decir:

$$H_0 : \mu = 150 \text{ cm}$$

Queremos ver si podemos rechazar esta hipótesis y viajar a la isla. En este caso, α es la probabilidad de rechazar la hipótesis nula cuando es verdadera (lo cual significa nuestra muerte). Definamos entonces la *Hipótesis Alternativa* H_1 como un posible resultado del experimento si no se cumple la hipótesis nula. Es decir:

$$H_1 : \mu \neq 150 \text{ cm}$$

Entonces, tomemos $\alpha = 0.05$ tal que las probabilidades que decidamos viajar y que seamos comidos por caníbales sea igual o menor a 5%. Si asumimos que $X \sim \mathcal{N}(\mu, \sigma^2)$ entonces calculemos la altura media crítica de H donde $\mathbb{P}(\bar{X}_n \geq H | H_0) = \alpha$. Sabemos que $\mathbb{P}(Z \geq z_{1-\alpha}) = \alpha$ así que si definimos:

$$Z = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

y si asumimos que H_0 es verdadera ($\mu = 150$ cm) entonces:

$$z_{1-\alpha} = \frac{H - \overbrace{150 \text{ cm}}^{=\mu_0}}{\frac{\sigma}{\sqrt{n}}}$$
$$\therefore H = \mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$$

Esta altura cumple que $\mathbb{P}(\bar{X}_n \geq H | H_0) = \alpha$. Entonces, nuestro test va a consistir en calcular \bar{X}_n y si $\bar{X}_n \geq \mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$ entonces vamos a viajar y si no no vamos a viajar. Supongamos que el explorador nos dijo que para $n = 9$ obtuvo que $\bar{X}_9 = 158$ cm y que sabemos que la desviación estándar de las alturas es de $\sigma = 10$ cm. Entonces:

$$H = \mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$$
$$= 150 \text{ cm} + 1.64 \cdot \frac{10 \text{ cm}}{\sqrt{9}}$$
$$= 155.47 \text{ cm}$$

Como $\bar{X}_9 = 158 \text{ cm} \geq H = 155.47 \text{ cm}$ sabemos que la probabilidad que la tribu sea hostil es menor que del 5% y, por lo tanto, nos vamos a ir de viaje.

Un **Test de Hipótesis** es un método de decisión basada en un estadístico (o variable estadística) para ver si ciertos datos apoyan cierta hipótesis. El método consiste en definir dos hipótesis: la **Hipótesis Nula** (H_0) que es

la hipótesis que queremos testear y que nos da información precisa sobre un estadístico ($\mu = 150$ cm en el ejemplo anterior) y la **Hipótesis Alternativa** (H_1) que es la hipótesis que vamos a tomar si es que rechazamos la hipótesis nula (no necesariamente es el complemento de la hipótesis nula). Una vez definidas las hipótesis se define una **Región de Rechazo** que es una región o intervalo donde si el estadístico está dentro se rechaza la hipótesis nula ($\bar{X}_n \geq H$ en el ejemplo anterior). Para definir esta región será necesario definir dos tipos de errores. Llamamos **Error de Tipo 1** (α) a la probabilidad de rechazar la hipótesis nula cuando es verdadera (en el ejemplo anterior es la probabilidad de viajar cuando los aborígenes son hostiles). Llamamos **Error de Tipo 2** (β) a la probabilidad de no rechazar la hipótesis nula cuando es falsa (en el ejemplo anterior es la probabilidad de no viajar cuando los aborígenes son amigables). Una vez definidas las hipótesis y los errores se puede armar una región de rechazo y entonces el test consiste en medir el estadístico en cuestión y fijarse si cae dentro de la zona de rechazo. Si es así entonces se rechaza la hipótesis nula y se toma la hipótesis alternativa, y si **no se puede rechazar la hipótesis nula**. Esto no es lo mismo que aceptar la hipótesis nula, si no que significa que los datos adquiridos no pueden rechazar la hipótesis nula. Esto puede ser porque la hipótesis nula sea verdadera o porque no tenemos suficientes datos. El test no sirve para afirmar la veracidad de una hipótesis, sino para averiguar si se puede rechazar una hipótesis.

Habiendo definido como funciona un test de hipótesis falta ver cómo podemos armar la región de rechazo \mathcal{R} . Por lo general, vamos a definir la región de rechazo fijando al error de tipo 1 α a un valor determinado. En el ejemplo anterior fijamos $\alpha = 0.05$ y buscamos la zona de rechazo para el estadístico \bar{X}_n que cumpla que $\mathbb{P}(\bar{X}_n \in \mathcal{R}) = \alpha$. De esta forma, hallamos que la zona de rechazo era $\mathcal{R} = [\mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}, \infty)$. Otra forma de hacer el test es en vez de definir una región de rechazo calcular un estadístico llamado **p-Valor**. El p-valor es la probabilidad de que el estadístico sea más extremo que el observado con los parámetros de la hipótesis nula. Con el p-valor calculado lo que debemos hacer es compararlo con α y si $p < \alpha$ se rechaza la hipótesis nula. Veamos un ejemplo.

Consideremos un mago que saca 1500 cartas del mazo y logra adivinar el color de 795 de ellas. Queremos hacer un test para ver si el mago es verdaderamente mago o si simplemente tuvo suerte. Definamos la variable de Bernoulli $X \sim \text{Be}(p)$ para expresar si el mago adivina o no una carta. Entonces, las hipótesis del test son:

$$\begin{cases} H_0 : & p = \frac{1}{2} \\ H_1 : & p > \frac{1}{2} \end{cases}$$

Usemos:

$$\begin{aligned} Z &= \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \\ &= \frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \\ &\sim \mathcal{N}(0, 1) \end{aligned}$$

$$\begin{aligned} \therefore \mathbb{P}(Z > z_{1-\alpha} | H_0) &= \mathbb{P}\left(\frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} > z_{1-\alpha} | H_0\right) \\ &= \mathbb{P}\left(\bar{X}_n > p + \sqrt{\frac{p(1-p)}{n}} z_{1-\alpha} | H_0\right) \\ &= \mathbb{P}\left(\bar{X}_n > \frac{1}{2} \left(1 + \frac{z_{1-\alpha}}{\sqrt{n}}\right)\right) \\ &= \alpha \end{aligned}$$

Por lo tanto, la región de rechazo para este problema es $\mathcal{R} = \left(\frac{1}{2} \left(1 + \frac{z_{1-\alpha}}{\sqrt{n}}\right), \infty\right)$. Ahora, si queremos calcular el p-valor debemos ver cuáles son las probabilidades de que el estadístico (\bar{X}_n en este caso) sea más extremo que el

calculado con los parámetro de la hipótesis nula ($p = 1/2$). Es decir:

$$\begin{aligned} p &= \mathbb{P}(X > \bar{X}_n | H_0) \\ &= \mathbb{P}\left(Z > \frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} | H_0\right) \\ &= \mathbb{P}(Z > \sqrt{n}(2\bar{X}_n - 1)) \end{aligned}$$

Ahora usemos los datos del problema para sacar conclusiones del test. Sabemos que $n = 1500$ y sabemos que acertó 795 cartas, así que:

$$\begin{aligned} \bar{X}_n &= \frac{795}{1500} \\ &= 0.530 \end{aligned}$$

Tomemos $\alpha = 0.005$, tal que $z_{1-\alpha} = 2.57$ y entonces la región de rechazo nos quede $\mathcal{R} \approx (0.533, \infty)$. Como $0.530 \leq 0.533$ entonces $\bar{X}_n \notin \mathcal{R}$ y no se puede rechazar la hipótesis nula. Es decir, es posible que el mago esté adivinando (las probabilidades de que el mago esté adivinando son mayores al 0.5%). Ahora, veamos si usando el p-valor llegamos a la misma conclusión:

$$\begin{aligned} p &= \mathbb{P}(Z > \sqrt{n}(2\bar{X}_n - 1)) \\ &\approx \mathbb{P}(Z > 2.32) \\ &\approx 1 - 0.9898 \\ &\approx 0.010 \end{aligned}$$

Como $0.010 \geq 0.005$ entonces $p \geq \alpha$ y no se puede rechazar la hipótesis nula. Notemos que ambos métodos nos dan el mismo resultado para el test. En este caso el p-valor nos dice que la probabilidad que el mago haya adivinado 795 cartas de 1500 es del 1%. Como nosotros definimos al error del tipo 1 como de 0.5% entonces no podemos rechazar la hipótesis nula. Es decir, no podemos estar lo suficientemente seguros de que el mago no haya adivinado.

Hasta ahora vimos casos de test donde σ es conocido o n es lo suficientemente grande como para que podamos estimar σ con \hat{S}_n . En estos casos, tomamos al estadístico como $Z = \frac{\bar{X}_n - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}}$. Ahora, ¿qué pasa si μ es desconocido y queremos hacer un test para σ ? En estos casos, podemos usar el siguiente estadístico:

$$T_n = (n-1) \frac{\hat{S}_n^2}{\sigma^2} \sim \chi_{n-1}^2$$

Consideremos un ejemplo. Supongamos que tenemos 25 mediciones de la altura de los alumnos del curso y queremos estimar su desviación. Sabemos que $\hat{S}_5 = 4$ cm y supongamos que tomamos $\alpha = 0.1$. Tomemos las siguientes hipótesis:

$$\begin{cases} H_0 : \sigma = 5.5 \text{ cm} \\ H_1 : \sigma < 5.5 \text{ cm} \end{cases}$$

Entonces:

$$\begin{aligned} \therefore \mathbb{P}(T_n < \chi_{25-1, \alpha}^2 | H_0) &= \mathbb{P}\left((n-1) \frac{\hat{S}_n^2}{\sigma^2} < 0.2971 | H_0\right) \\ &= \mathbb{P}\left(\hat{S}_n < \sqrt{\frac{0.2971}{(25-1)} \cdot (5.5)^2}\right) \\ &= \mathbb{P}(\hat{S}_n < 3.69) \\ &= 0.1 \end{aligned}$$

Por lo tanto, la zona de rechazo es $\mathcal{R} = (-\infty, 3.69)$. Como $\hat{S}_n \notin \mathcal{R}$ entonces no podemos rechazar la hipótesis nula.

7.2. Tests No Paramétricos

7.2.1. Test de Adherencia

El **Test de Adherencia** consiste en testear si un modelo probabilístico es adecuado para un dado conjunto de datos observados. Antes de describir la metodología de este test veamos un ejemplo.

Supongamos que somos biólogos estudiando genética. Tenemos a un padre con genotipo Aa y una madre con genotipo Aa . El modelo teórico dice que las probabilidades de los genotipos de los hijos son:

| | | | |
|--------------|---------------|---------------|---------------|
| Tipo | AA | Aa | aa |
| Probabilidad | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ |

Entonces, se estudian 100 descendientes de una pareja con esos genotipos y la distribución que se observa es la siguiente:

| | | | |
|------------|------|------|------|
| Tipo | AA | Aa | aa |
| Frecuencia | 26 | 45 | 29 |

Queremos ver si los datos encontrados se corresponden al modelo teórico. Empecemos por armar una tabla que compare las frecuencias observadas con las esperadas según el modelo:

| | | | |
|----------------------|------|------|------|
| Tipo | AA | Aa | aa |
| Frecuencia Observada | 26 | 45 | 29 |
| Frecuencia Esperada | 25 | 50 | 25 |

A ojo podemos ver que los valores son parecidos. Definamos las siguientes hipótesis:

$$\begin{cases} H_0 : (\mathbb{P}(AA) = \frac{1}{4}) \wedge (\mathbb{P}(Aa) = \frac{1}{2}) \wedge (\mathbb{P}(aa) = \frac{1}{4}) \\ H_1 : \neg H_0 \end{cases}$$

Tomemos $O_i, i \in [1, 3]$, como las frecuencias observadas (O_1 para la frecuencia de AA , O_2 para la de Aa y O_3 para la de aa) y E_i como las frecuencias esperadas. Entonces, definamos el siguiente estadístico:

$$\chi^2 = \sum_{i=1}^3 \frac{(O_i - E_i)^2}{E_i}$$

Si H_0 es verdadera entonces el estadístico tiene distribución asintótica chi-cuadrado con $3 - 1$ grados de libertad. En este caso tenemos que:

$$\begin{aligned} \chi^2 &= \frac{(26 - 25)^2}{25} + \frac{(45 - 50)^2}{50} + \frac{(29 - 25)^2}{25} \\ &= 0.04 + 0.5 + 0.64 \\ &= 1.18 \end{aligned}$$

Si tomamos $\alpha = 0.1$ entonces viendo la tabla vemos que $\chi_{3-1, \alpha}^2 = 9.210$. Como $\chi^2 \leq \chi_{3-1, \alpha}^2$ entonces no podemos rechazar la hipótesis nula.

En general, para hacer un test de adherencia de esta forma vamos a pedir que n sea grande y que $E_i \geq 5 \forall i$. Vamos a tener k categorías y el estadístico que vamos a usar que debería seguir una distribución chi-cuadrado si se cumple H_0 es:

$$X = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

7.2.2. Test de Independencia

El **Test de Independencia** consiste en testear si hay independencia entre dos variables aleatorias. Para describir la metodología de este test veamos un ejemplo.

Supongamos que queremos verificar si hay dependencia entre renta y número de hijos en las familias de una ciudad. Se eligen al azar 250 familias y se obtiene la siguiente información:

| Renta \ # de Hijos | 0 | 1 | 2 | 3+ | Total |
|--------------------|----|----|----|----|-------|
| Menos de \$2000 | 15 | 27 | 50 | 43 | 135 |
| \$2000 - \$5000 | 25 | 30 | 12 | 8 | 75 |
| Más de \$5000 | 8 | 13 | 9 | 10 | 40 |
| Total | 48 | 70 | 71 | 61 | 250 |

Entonces, tomemos X como la variable aleatoria que nos dice el precio de la renta y Y como la variables aleatoria que nos dice la cantidad de hijos y definamos las siguientes hipótesis:

$$\begin{cases} H_0 : X \perp Y \\ H_1 : X \not\perp Y \end{cases}$$

Si H_0 es verdadera entonces $\mathbb{P}((X = x) \cap (Y = y)) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$. Por lo tanto, definamos $\eta(x, y)$ como las observaciones de $X = x$ e $Y = y$ con la siguiente notación:

$$\begin{cases} \eta(x, \cdot) : & \text{Cantidad de observaciones de } X = x \\ \eta(\cdot, y) : & \text{Cantidad de observaciones de } Y = y \\ \eta(x, y) : & \text{Cantidad de observaciones de } X = x \text{ e } Y = y \end{cases}$$

Por lo tanto, bajo la hipótesis de independencia el número esperado de observaciones en cada celda es:

$$\begin{aligned} E_{xy} &= \frac{\eta(x, y)}{n} \\ &= \frac{\eta(x, \cdot)\eta(\cdot, y)}{n} \end{aligned}$$

Definamos el siguiente estadístico:

$$\chi^2 = \sum_{x,y} \frac{(E_{xy} - O_{xy})^2}{E_{xy}},$$

donde O_{xy} es el número total de observaciones en cada celda. Si H_0 es verdadera entonces $\chi^2 \sim \chi_{k,\alpha}^2$, donde $k = (f - 1)(c - 1)$, f es el número de filas y c el número de columnas. Entonces, para calcular este estadístico armemos una tabla con los valores de E_{xy} para cada celda:

| Renta \ # de Hijos | 0 | 1 | 2 | 3+ | Total |
|--------------------|-------|-------|-------|-------|-------|
| Menos de \$2000 | 25.92 | 37.80 | 38.34 | 32.94 | 135 |
| \$2000 - \$5000 | 14.40 | 21.00 | 21.30 | 18.30 | 75 |
| Más de \$5000 | 7.68 | 11.20 | 11.36 | 9.76 | 40 |
| Total | 48 | 70 | 71 | 61 | 250 |

$$\therefore \chi^2 \approx 36.62$$

Usando que $k = 6$, $\alpha = 0.05$ y la tabla de chi-cuadrado vemos que $\chi_{k,\alpha}^2 = 18.55$. Como $\chi^2 > \chi_{k,\alpha}^2$ entonces rechazamos la hipótesis nula y concluimos que el precio de renta y el tamaño de la familia no son variables independientes. Esto es razonable, ya que familias más grandes requieren hogares más grandes y, por lo tanto, el precio de renta es mayor.

7.2.3. Test de Bondad de Ajuste

El **Test de Bondad de Ajuste** consiste en testear si ciertos datos observados se ajustan con una distribución dada. Supongamos que tenemos una variable aleatoria X con distribución F desconocida y supongamos que queremos testear si la distribución de X es F_0 , una distribución que si conocemos. Entonces, definimos las siguientes hipótesis:

$$\begin{cases} H_0 : F = F_0 \\ H_1 : F \neq F_0 \end{cases}$$

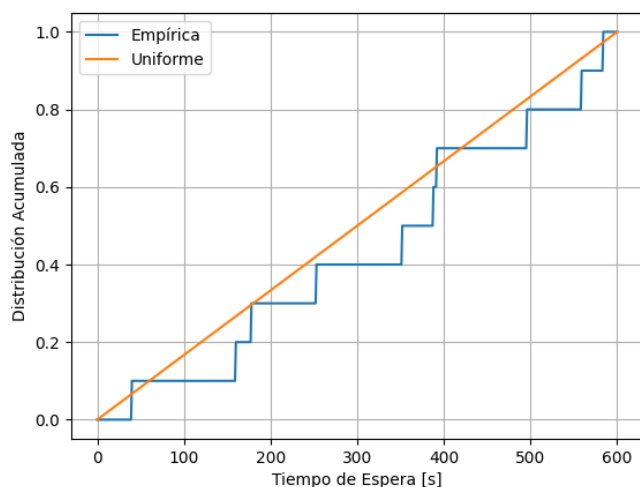


Figura 7.1: Gráficos de la distribución empírica y de la distribución uniforme para $x \in [0, 600]$.

Para testear estas hipótesis vamos a usar un test de bondad de ajuste llamado **Test de Kolmogorov-Smirnov**. Dado un modelo \mathbf{X} de X vamos a usar la distribución acumulada de la distribución empírica:

$$\hat{F}_{\mathbf{X}}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i \leq x\}}(x)$$

Entonces, definamos las siguientes funciones:

$$\begin{cases} D_n^+ = \sup \left(\hat{F}_{\mathbf{X}}(x) - F(x) \right) \\ D_n^- = \sup \left(F(x) - \hat{F}_{\mathbf{X}}(x) \right) \\ D_n = \sup \left(\left| \hat{F}_{\mathbf{X}}(x) - F(x) \right| \right) = \max(D_n^+, D_n^-) \end{cases}$$

No lo vamos a demostrar pero resulta que $4nD_n^2 \sim \chi_2^2$ para n grande. Por lo tanto, podemos usar a esta cantidad como estadístico para hacer el test. También, se puede plantear la hipótesis alternativa como $H_1 : F > F_0$ o $H_1 : F < F_0$. En estos casos, vamos a usar como estadístico a $4n(D_n^+)^2$ y $4n(D_n^-)^2$ respectivamente. Veamos algún ejemplo.

Supongamos que tomamos una medición del tiempo (en segundos) que esperamos al colectivo de una dada línea en una determinada parada y obtenemos los siguientes datos:

$$\mathbf{X} = \{160, 559, 392, 584, 178, 496, 253, 388, 352, 40\}$$

Queremos testear si $X \sim \mathcal{U}(0, 600)$, así que definimos las siguientes hipótesis:

$$\begin{cases} H_0 : F = \frac{x}{600} \mathbb{I}_{[0,600]}(x) \\ H_1 : F \neq \frac{x}{600} \mathbb{I}_{[0,600]}(x) \end{cases}$$

Entonces, empecemos por ordenar los datos:

$$\mathbf{X} = \{40, 160, 178, 253, 352, 388, 392, 496, 559, 584\}$$

Si graficamos la distribución empírica y la uniforme para $x \in [0, 600]$ obtenemos los gráficos de la **Figura 7.1**.

Ahora, armemos la siguiente tabla:

| Intervalo | $\hat{F}_{\mathbf{X}}(x)$ | $F(x)$ | $\hat{F}_{\mathbf{X}}(x) - F(x)$ | $\sup \left(\left \hat{F}_{\mathbf{X}}(x) - F(x) \right \right)$ |
|--------------------|---------------------------|-----------------|----------------------------------|---------------------------------------------------------------------|
| $0 \leq x < 40$ | 0 | $\frac{x}{600}$ | $\frac{x}{600}$ | $\frac{1}{15}$ |
| $40 \leq x < 160$ | $\frac{1}{10}$ | $\frac{x}{600}$ | $\frac{1}{10} - \frac{x}{600}$ | $\frac{1}{6}$ |
| $160 \leq x < 178$ | $\frac{2}{10}$ | $\frac{x}{600}$ | $\frac{2}{10} - \frac{x}{600}$ | $\frac{29}{300}$ |
| $178 \leq x < 253$ | $\frac{3}{10}$ | $\frac{x}{600}$ | $\frac{3}{10} - \frac{x}{600}$ | $\frac{73}{600}$ |
| $253 \leq x < 352$ | $\frac{4}{10}$ | $\frac{x}{600}$ | $\frac{4}{10} - \frac{x}{600}$ | $\frac{14}{75}$ |
| $352 \leq x < 388$ | $\frac{5}{10}$ | $\frac{x}{600}$ | $\frac{5}{10} - \frac{x}{600}$ | $\frac{11}{75}$ |
| $388 \leq x < 392$ | $\frac{6}{10}$ | $\frac{x}{600}$ | $\frac{6}{10} - \frac{x}{600}$ | $\frac{4}{75}$ |
| $392 \leq x < 496$ | $\frac{7}{10}$ | $\frac{x}{600}$ | $\frac{7}{10} - \frac{x}{600}$ | $\frac{19}{150}$ |
| $496 \leq x < 559$ | $\frac{8}{10}$ | $\frac{x}{600}$ | $\frac{8}{10} - \frac{x}{600}$ | $\frac{79}{600}$ |
| $559 \leq x < 584$ | $\frac{9}{10}$ | $\frac{x}{600}$ | $\frac{9}{10} - \frac{x}{600}$ | $\frac{11}{150}$ |
| $584 \leq x < 600$ | 1 | $\frac{x}{600}$ | $1 - \frac{x}{600}$ | $\frac{2}{75}$ |
| $600 \leq x$ | 1 | 1 | 0 | 0 |

Notemos que de la columna de supremos tenemos que el máximo es el del intervalo $253 \leq x < 352$, o sea que $D_n = \frac{14}{75}$. Por lo tanto, nuestro estadístico es:

$$\begin{aligned} \chi^2 &= 4nD_n^2 \\ &= 4 \cdot 10 \cdot \left(\frac{14}{75} \right)^2 \\ &\approx 1.394 \end{aligned}$$

Si usamos $\alpha = 0.05$ entonces $\chi_{2,\alpha}^2 = 10.597$. Como $\chi^2 < \chi_{2,\alpha}^2$ entonces no rechazamos la hipótesis nula. Esto es razonable, ya que se puede ver en el gráfico que la curva de la uniforme ajusta bien a la distribución empírica.

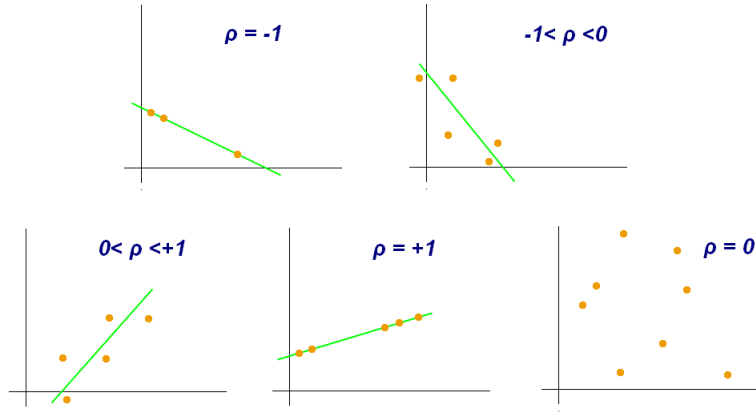


Figura 8.1: Valores del coeficiente de Pearson para distintos gráficos.

8. Correlación y Regresión Lineal

Supongamos que tenemos dos variables aleatorias y que queremos expresar la intensidad de la relación entre dos variables. Para esto, empecemos por definir el **Coefficiente de Correlación** (o simplemente la **Correlación**) entre dos variables aleatorias X e Y como:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

Este coeficiente mide no solo la intensidad sino que también la dirección entre las variables. Mientras mayor sea la correlación en módulo mayor es la intensidad de la relación ($|\rho| \leq 1$). Además, al igual que la covarianza, la correlación nos dice la dirección a través del signo: si $\rho > 0$ entonces la relación es positiva, si $\rho < 0$ la relación es negativa, y si $\rho = 0$ entonces no hay relación entre las variables.

El coeficiente de correlación muestral se conoce como el **Coefficiente de Pearson**, y es:

$$r = \frac{\text{CovMuestral}(X, Y)}{\sqrt{S_X^2 S_Y^2}},$$

donde $\text{CovMuestral}(X, Y)$ es la covarianza muestral entre X e Y :

$$\text{CovMuestral}(X, Y) = \frac{1}{n} \sum_i (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)$$

Este coeficiente se suele usar a la hora de hacer **Regresión Lineal** sobre un modelo de ambas variables. El objetivo de la regresión lineal es predecir o estimar los resultados de una variable basados en la otra (u otras). Al graficar los datos y hacer regresión lineal se obtiene una función lineal que busca ajustar los datos graficados. El coeficiente de Pearson nos da un valor que nos indica qué tan bueno es el ajuste sobre los datos. Mientras mejor sea el ajuste, mayor es el coeficiente en módulo. En la **Figura 8.1** se encuentran distintos gráficos con sus respectivos valores del coeficiente de Pearson.

Índice alfabético

- Coefficiente de Correlación, 60
- Coefficiente de Pearson, 60
- Convergencia en Distribución, 41
- Covarianza, 32

- Delta de Kronecker, 12
- Desigualdad de Chebyshev, 37
- Desigualdad de Markov, 37
- Desvío Estándar, 11
- Desviación Estándar, 36
- Distribución Binomial, 15
- Distribución Binomial Negativa, 17
- Distribución Chi Cuadrado, 27
- Distribución de Bernoulli, 12
- Distribución de Poisson, 17
- Distribución de Probabilidad, 21
- Distribución de Probabilidad Acumulada, 21
- Distribución de Probabilidad Puntual, 8
- Distribución de Probabilidad Puntual Acumulada, 8
- Distribución Empírica, 20
- Distribución Exponencial, 24
- Distribución Gamma, 26
- Distribución Geométrica, 13
- Distribución Hipergeométrica, 20
- Distribución Marginal, 28
- Distribución Normal, 22
- Distribución Normal Estándar, 22
- Distribución t-Student, 27
- Distribución Uniforme, 23

- Error, 49
- Error Cuadrático Medio, 49
- Error de Tipo 1, 54
- Error de Tipo 2, 54
- Error Estándar, 49
- Error Estadístico, 36
- Espacio de Eventos, 4
- Espacio Muestral, 4
- Esperanza, 9
- Estadística Paramétrica, 46
- Estimación, 46
- Estimador Asintóticamente Insesgado, 46
- Estimador Consistente, 46
- Estimador de Varianza, 49
- Estimador Insesgado, 46
- Estimador Puntual, 46
- Evento, 4
- Eventos Independientes, 5

- Fórmula de Bayes, 6
- Fórmula de Probabilidad Total, 5

- Familia de Eventos Independientes, 7
- Función Gamma, 26
- Función Generadora de Momentos, 38
- Función Indicatriz, 23

- Grados de Libertad, 27

- Hipótesis Alternativa, 54
- Hipótesis Nula, 53

- Intervalo de Confianza, 50

- Ley de Grandes Números, 37

- Método de Máxima Verosimilitud, 47
- Método de Momentos, 46
- Método del Pivote, 51
- Media Muestral, 36
- Mediana, 21
- Modelo, 36
- Momento, 38
- Muestra, 36

- p-Valor, 54
- Percentil, 21
- Probabilidad Condicional, 5
- Probabilidad de Error, 51
- Promedio, 21

- Región de Rechazo, 54
- Regresión Lineal, 60

- Sesgo, 46

- Test de Adherencia, 56
- Test de Bondad de Ajuste, 57
- Test de Hipótesis, 53
- Test de Independencia, 56
- Test de Kolmogorov-Smirnov, 58

- Variable Aleatoria, 8
- Variable Aleatoria Continua, 21
- Varianza, 11
- Varianza Muestral, 49
- Vector Aleatorio, 28