INGENIERIA INDUSTRIAL
UNIVERSIDAD DE CHILE

Facultad de Ciencias Físicas y Matemáticas
Universidad de Chile

Probabilidades

Apunte de Cátedra

Iván Vidal R. & Charles Thraves

Marzo, 2025

Aclaraciones

Este apunte fue elaborado por Iván Vidal Romero, Profesor Auxiliar del curso IN3141 - Probabilidades, en colaboración y supervisión de Charles Thraves, Profesor de Cátedra del Departamento de Ingeniería Industrial. Su contenido se basa en las clases y materiales del mismo, complementado con diversos textos que estudian la Teoría de la Probabilidad.

El objetivo principal de este material es apoyar el aprendizaje y desarrollo de habilidades para enfrentar el curso, sirviendo como complemento a las clases, pero no como un reemplazo de estas. Confío en que este recurso será útil para los estudiantes y futuros ingenieros del departamento de Ingeniería Industrial, así como para quienes deseen profundizar en el estudio de las probabilidades. Espero que este material junto a las guías de estudio aporten un granito de arena a su proceso de aprendizaje y haya sido de utilidad a la hora de su estudio.

Quiero enfatizar que los errores contenidos en este material son exclusivamente míos y de nadie más. Si encuentran algún error o les gustaría sugerir algún cambio, les agradeceré que lo informen a Ivan.vidal@ug.uchile.cl o a Charles Thraves. Por favor, traten este material con cariño, espero que el curso despierte su curiosidad por el mundo de las probabilidades, el cual es mucho más amplio de lo que podrían imaginar. Si desean profundizar en este campo, pueden considerar cursar IN79O - Modelos Estocásticos. Un curso que aborda de manera más teórica algunos de los temas aquí presentados y explora otras áreas igualmente interesantes. ¡Disfruten el proceso de aprendizaje y su etapa en la universidad!

Índice general

Introducción

La aleatoriedad y la incertidumbre son conceptos fundamentales que se manifiestan en nuestra vida cotidiana y en prácticamente todas las disciplinas de la ciencia, la ingeniería y la tecnología. La teoría de la probabilidad proporciona un marco matemático para describir y analizar fenómenos aleatorios, es decir, aquellos eventos cuyos resultados no pueden predecirse con certeza. Este marco es esencial para modelar situaciones donde la información es incompleta o los resultados dependen de variables no controlables.

Un ejemplo sencillo que ilustra este concepto es el lanzamiento de una moneda balanceada. Aunque factores físicos como la fuerza del lanzamiento o la orientación inicial podrían, en teoría, determinar el resultado, en la práctica, la falta de conocimiento detallado convierte al evento en aleatorio. Así, al hablar de "aleatoriedad" nos referimos a nuestra incapacidad de conocer todos los elementos necesarios para predecir el resultado. En este caso, la probabilidad de que salga cara se expresa como 50 %, lo cual puede interpretarse desde dos perspectivas: como la frecuencia relativa en una gran cantidad de lanzamientos o como una medida subjetiva del grado de certeza que asignamos al evento.

La teoría de la probabilidad no solo es flexible al acomodar diferentes interpretaciones ya sea como frecuencia relativa o como creencia subjetiva, sino que también ofrece herramientas versátiles para abordar problemas de incertidumbre. Este enfoque ha demostrado ser invaluable en contextos que van desde la toma de decisiones en ingeniería y negocios, hasta la formulación de modelos científicos y el análisis de datos en otras áreas.

Capítulo 1

Teoría Básica de Conjuntos

La teoría de probabilidad se construye sobre el lenguaje de los conjuntos. Estos permiten modelar los resultados posibles de obtener de un experimento o una colección de resultados que sean de interés. A continuación, revisaremos los conceptos básicos de teoría de conjuntos, incluyendo definiciones, notaciones, operaciones y propiedades fundamentales. También introducimos la distinción entre conjuntos contables y no contables, y establecemos algunas herramientas visuales, como los diagramas de Venn, para facilitar la comprensión.

1.1 Conjuntos

Un *conjunto* es una colección no ordenada de elementos sin repetición. Los conjuntos generalmente los denotaremos con letras mayúsculas, como por ejemplo A, B, C; mientras que los elementos individuales se denotarán con letras minúsculas, como por ejemplo a, b, c. Si un elemento pertenece a un conjunto, se denotará como $a \in A$; si no pertenece, escribiremos $a \notin A$.

Ejemplo:

Sea el conjunto $A = \{1, 2, 3\}$, decimos entonces que $2 \in A$, y que $4 \notin A$. Además, el orden de los elementos en un conjunto no importa, y un elemento no se repite, por lo que $A = \{1, 2, 3\}$ es igual a $B = \{3, 2, 1\}$.

1.2 Cardinalidad de un Conjunto

La *cardinalidad* de un conjunto A, denotada como $|A|$, es el número de elementos en A. Según las definiciones formales introducidas por Georg Cantor, un conjunto puede ser clasificado en dos tipos según la cardinalidad de este:

- **Conjuntos Numerables o Contables:** Son aquellos conjuntos cuyos elementos pueden ser enumerados (finita o infinitamente), es decir, conjuntos que poseen una cardinalidad menor o igual que los números naturales ($\mathbb{N} = \{1, 2, 3, \dots\}$), como por ejemplo, el conjunto de los números enteros $\mathbb{Z}$.
- **Conjuntos no Numerables o no Contables:** Son aquellos conjuntos que no pueden ponerse en correspondencia uno a uno con los números naturales, es decir, son conjuntos con tantos elementos que no pueden ser enumerados, como los números reales en el intervalo $[0, 1]$ o el conjunto de los números reales $\mathbb{R}$.

1.3 Tipos de Conjuntos

Se describen a continuación definiciones básicas sobre conjuntos.

- **Subconjuntos:** Un conjunto A es un *subconjunto* de otro conjunto B si todos los elementos de A están también en B. Esto se denota como $A \subseteq B$, y formalmente, se expresa como:

$$A \subseteq B \iff (x \in A \implies x \in B).$$

- **Conjunto Universo:** El *conjunto universo*, Ω, contiene todos los posibles elementos en un contexto dado, y cualquier conjunto es un subconjunto de Ω. En probabilidades, se utiliza para modelar el conjunto de todos los posibles resultados de un experimento.

- **Conjunto Vacío:** El *conjunto vacío*, denotado como $\emptyset$, es el único conjunto que no contiene elementos, es decir, el conjunto vacío es tal que $|\emptyset| = 0$. Por definición, $\emptyset \subseteq A$ para cualquier conjunto A.

- **Conjunto Potencia:** El *conjunto potencia* de un conjunto A, denotado como $\mathcal{P}(A)$, es el conjunto de todos los subconjuntos posibles de A. Su cardinalidad es $|\mathcal{P}(A)| = 2^{|A|}$. Por ejemplo, si $A = \{1, 2\}$, entonces:

$$\mathcal{P}(A) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}.$$

- **Conjuntos Disjuntos:** Dos conjuntos A y B se dicen *disjuntos* si no tienen elementos en común, formalmente:

$$A \cap B = \emptyset.$$

- **Partición:** Una *partición* de un conjunto A es una colección de subconjuntos de A, tal que estos conjuntos son disjuntos, y su unión es A. Formalmente:

$$A = \bigcup_{i=1}^{n} A_i, \quad A_i \cap A_j = \emptyset \text{ para } i \neq j.$$

  Por ejemplo, si $A = \{1, 2, 3, 4\}$, una partición posible de A sería $\{A_1, A_2\}$, donde $A_1 = \{1, 2\}$ y $A_2 = \{3, 4\}$.

1.4 Operaciones con Conjuntos

Además de relacionar los conjuntos, también podemos crear unos nuevos a través de las operaciones entre conjuntos. A continuación, se presentan las operaciones básicas entre conjuntos.

- **Unión:** La *unión* de dos conjuntos A y B, denotada $A \cup B$, es el conjunto que contiene todos los elementos que están en A, en B, o en ambos:

$$A \cup B = \{x \mid x \in A \text{ o } x \in B\}.$$

Por ejemplo, sean $A = \{1, 2\}$ y $B = \{2, 3\}$. Entonces:

$$A \cup B = \{1, 2, 3\}.$$

Así, se define también la unión de una *familia arbitraria de los conjuntos* $\{A_i\}_{i \in I}$, denotada $\bigcup_{i \in I} A_i$, la cúal es el conjunto de todos los elementos que pertenecen a al menos uno de los conjuntos A_i:

$$\bigcup_{i \in I} A_i = \{x \mid \exists i \in I \text{ tal que } x \in A_i\}.$$

- **Intersección:** La *intersección* de dos conjuntos A y B, denotada $A \cap B$, es el conjunto de todos los elementos que están tanto en A como en B:

$$A \cap B = \{x \mid x \in A \text{ y } x \in B\}.$$

Por ejemplo, sean $A = \{1, 2\}$ y $B = \{2, 3\}$. Entonces:

$$A \cap B = \{2\}.$$

Así, se define la intersección de una *familia arbitraria de los conjuntos* $\{A_i\}_{i \in I}$, denotada $\bigcap_{i \in I} A_i$, tal que, es el conjunto de todos los elementos que pertenecen a todos los conjuntos A_i:

$$\bigcap_{i \in I} A_i = \{x \mid x \in A_i \quad \forall i \in I\}.$$

- **Complemento:** El *complemento* de un conjunto A, denotado A^c, es el conjunto de todos los elementos que están en el conjunto universo Ω, pero no en A:

$$A^c = \{x \mid x \in \Omega \text{ y } x \notin A\}.$$

Por ejemplo, sea $\Omega = \{1, 2, 3, 4\}$ el conjunto universo y $A = \{1, 2\}$. Entonces:

$$A^c = \{3, 4\}.$$

- **Diferencia:** La *diferencia* entre dos conjuntos A y B, denotada $A - B$ o $A \setminus B$, es el conjunto de elementos que están en A pero no en B:

$$A \setminus B = \{x \mid x \in A \text{ y } x \notin B\}.$$

Por ejemplo, sean $A = \{1, 2, 3\}$ y $B = \{2, 4\}$. Entonces:

$$A \setminus B = \{1, 3\}.$$

- **Producto Cartesiano:** El *producto cartesiano* de dos conjuntos A y B, denotado como $A \times B$, es el conjunto de todos los pares ordenados (a, b) donde $a \in A$ y $b \in B$. Formalmente, se define como:

$$A \times B = \{(a, b) \in A \times B \mid a \in A \wedge b \in B\}.$$

Por ejemplo, sean $A = \{1, 2\}$ y $B = \{x, y\}$. Entonces:

$$A \times B = \{(1, x), (1, y), (2, x), (2, y)\}.$$

Propiedades:

- Si uno de los conjuntos es vacío, entonces el producto cartesiano también es vacío:
$$A \times \emptyset = \emptyset \times B = \emptyset.$$
- El producto cartesiano no es conmutativo en general, es decir, $A \times B \neq B \times A$, ya que los pares (a, b) y (b, a) son diferentes.
- El producto cartesiano no es asociativo en términos de estructura, ya que el orden de agrupación de los pares influye en la forma de los elementos:
$$(A \times B) \times C \neq A \times (B \times C).$$
En $(A \times B) \times C$, cada elemento tiene la forma $((a, b), c)$, mientras que en $A \times (B \times C)$, cada elemento tiene la forma $(a, (b, c))$. Aunque las agrupaciones difieren, ambos conjuntos contienen las mismas combinaciones de elementos y, por lo tanto, son *isomorfos.* [1]
- Si A y B son finitos, entonces el número de elementos en $A \times B$ es el producto del número de elementos en A y B:
$$|A \times B| = |A| \cdot |B|.$$

1.5 Propiedades de Operaciones de los Conjuntos

A continuación, se enuncian propiedades que se desprenden de las operaciones entre conjuntos. Estas propiedades permiten simplificar y reestructurar expresiones matemáticas de manera más eficiente.

- **Conmutatividad:** El orden en que se consideran los conjuntos no altera el resultado de la unión ni de la intersección:
$$A \cup B = B \cup A, \quad A \cap B = B \cap A.$$
Esto asegura que las operaciones son simétricas.
- **Asociatividad:** Al combinar varios conjuntos mediante unión o intersección, el agrupamiento de los conjuntos no afecta el resultado:
$$(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cap (B \cap C).$$
De forma que es posible omitir paréntesis cuando evaluamos expresiones de unión o intersección.
- **Leyes de De Morgan:** Estas leyes establecen relaciones entre las operaciones de unión, intersección y complemento. Para dos conjuntos:
$$(A \cup B)^c = A^c \cap B^c, \quad (A \cap B)^c = A^c \cup B^c.$$

[1] El término *isomorfo* se refiere a una correspondencia biunívoca (bijección) entre dos estructuras que preserva las propiedades relevantes. En este caso, aunque las agrupaciones de los pares difieren, ambas estructuras representan las mismas combinaciones de elementos.

Para una *familia arbitraria de conjuntos* $\{A_i\}_{i\in I}$, se generalizan como:

$$\left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c, \quad \left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c.$$

- **Distributividad:** La intersección distribuye sobre la unión, y la unión distribuye sobre la intersección:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

De forma de que se puede reorganizar operaciones entre conjuntos.

1.6 Diagramas de Venn

Las tipos de conjuntos u operaciones anteriores se pueden visualizar mediante *diagramas de Venn*. Estos diagramas representan conjuntos como regiones en un espacio cerrado, y las operaciones como relaciones geométricas entre esas regiones, son útiles para desarrollar una intuición visual sobre los conceptos de conjuntos y las relaciones entre ellos. No obstante, su empleo es limitado a fines ilustrativos, ya que no son adecuados para demostrar propiedades formales ni para extraer conclusiones generales aplicables a cualquier conjunto.

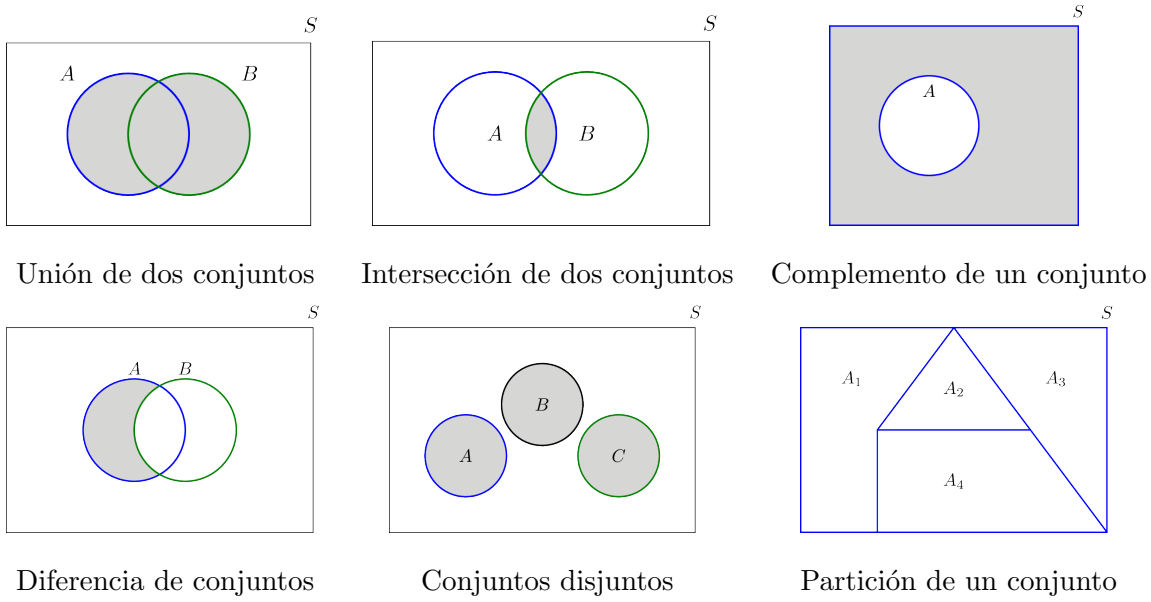

Figura 1.1: Diagramas de Venn.

Capítulo 2

Espacios de Probabilidad

La probabilidad es una disciplina matemática que nos permite modelar y comprender fenómenos aleatorios e inciertos. A continuación, se presentarán los conceptos fundamentales de esta teoría, como los experimentos aleatorios, los espacios muestrales, los axiomas y sus diferentes propiedades. Estos principios forman la base sobre la cual se construye toda la teoría de la probabilidad y sus numerosas aplicaciones.

2.1 Espacio Muestral

Un *experimento aleatorio* es un proceso cuyo resultado no puede predecirse con certeza antes de realizarlo. Por ejemplo, antes de lanzar un dado, no sabemos cuál será el número que aparecerá. Después de realizar el experimento, el resultado del experimento aleatorio se conoce y se le llama *outcome*.

Definición 2.1 (Espacio Muestral) El conjunto de todos los posibles resultados de un experimento aleatorio se denomina *espacio muestral* y se denota por Ω. Es el conjunto universo en el contexto de la probabilidad.

Ejemplo:

A continuación se muestran diferentes experimentos aleatorios y sus espacios muestrales.

- Lanzar una moneda: $\Omega = \{C, S\}$, donde C es cara y S es sello.
- Lanzar un dado: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Número de iPhones vendidos en un día: $\Omega = \{0, 1, 2, 3, \dots\}$.
- Número de goles en un partido de fútbol: $\Omega = \{0, 1, 2, 3, \dots\}$.
- Tiempo de espera para un autobús: $\Omega = [0, \infty)$.
- Posición de una partícula en un intervalo: $\Omega = [0, 1]$.

Cuando un experimento se realiza varias veces, cada repetición se denomina un *ensayo*. Por ejemplo, en el caso de lanzar una moneda, cada ensayo resulta en cara o sello. Es fundamental notar que el espacio muestral depende de cómo se defina el experimento aleatorio y los resultados de interés. Por ejemplo, supongamos que lanzamos dos monedas y estamos interesados en distintos tipos de eventos:

- Si nos interesa el resultado exacto de cada moneda, el espacio muestral será:

$$\Omega_1 = \{(C, C), (C, S), (S, C), (S, S)\}.$$

- Si, en cambio, solo nos importa el número total de caras obtenidas, el espacio muestral será:

$$\Omega_2 = \{0, 1, 2\}.$$

Es decir, el espacio muestral cambia dependiendo de qué aspecto del experimento se considere relevante, observando así cómo la teoría de la probabilidad permite modelar diferentes situaciones de manera adecuada.

2.1.1 Eventos

Un *evento* es cualquier subconjunto del espacio muestral Ω. Puede ser un único resultado o una colección de resultados. Por ejemplo, sí $\Omega = \{1, 2, 3, 4, 5, 6\}$, entonces el evento $A = \{2, 4, 6\}$ representa obtener un número par al lanzar un dado.

Ejemplo:

Si lanzamos un dado, algunos eventos posibles son:

- Obtener un número impar: $A = \{1, 3, 5\}$.
- Obtener un número mayor a 4: $B = \{5, 6\}$.
- Obtener un número divisible por 2: $C = \{2, 4, 6\}$.

2.2 Medida de Probabilidad

En matemáticas, el concepto de *medida* surge como una forma de generalizar la noción de tamaño o cantidad, que puede aplicarse a objetos geométricos como longitudes, áreas y volúmenes, y también a conjuntos abstractos. A lo largo de la historia, este concepto ha evolucionado desde ideas intuitivas hasta definiciones rigurosas.

Por ejemplo, los trabajos de Arquímedes sentaron las bases al vincular el área de figuras geométricas con proporciones conocidas, mientras que las ideas de Cantor y Borel introdujeron las primeras nociones formales de medida aplicadas a conjuntos en la recta real. Finalmente, el desarrollo de la teoría de la medida, impulsado por Lebesgue, permitió extender estas ideas a conjuntos más complejos y a contextos infinitos.
Para cuantificar la incertidumbre en experimentos aleatorios, utilizamos una función $\mathbb{P}(\,\cdot\,)$ llamada *medida de probabilidad*, que asigna a cada evento un valor real en el intervalo $[0, 1]$. Este valor representa la probabilidad de que ocurra dicho evento, si es cercano a 0, el evento es poco probable; si es cercano a 1, es muy probable. La medida de probabilidad debe cumplir ciertos principios fundamentales (conocidos como axiomas de probabilidad), que garantizan que las asignaciones de probabilidad sean consistentes matemáticamente. Estas propiedades permiten describir de manera metódica la ocurrencia de eventos y son la base para calcular probabilidades en espacios muestrales discretos y continuos.

2.3 Axiomas de Probabilidad

La medida de probabilidad $\mathbb{P}(\cdot)$ obedece tres principios fundamentales, conocidos como los Axiomas de Probabilidad:

Definición 2.2 (Axiomas de Probabilidad)

1. **No-negatividad:** Para cualquier evento $A \subseteq \Omega$,
$$\mathbb{P}(A) \geq 0.$$
2. **Normalización:** La probabilidad del espacio muestral es 1,
$$\mathbb{P}(\Omega) = 1.$$
3. **Aditividad:** Si $\{A_i\}_{i=1}^{\infty}$ es una colección de eventos disjuntos (es decir, $A_i \cap A_j = \emptyset$ para $i \neq j$), entonces:
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Estos axiomas forman la base teórica para trabajar con probabilidades de eventos tanto simples como compuestos, y nos permiten calcular probabilidades de manera consistente en experimentos aleatorios.

2.4 Espacios de Probabilidad

En teoría de la probabilidad, trabajamos con espacios muestrales que pueden ser finitos o infinitos. En el caso de espacios infinitos, no siempre es posible asignar probabilidades a todos los subconjuntos de Ω sin generar inconsistencias. Por ejemplo, intentar asignar probabilidades directamente a cada subconjunto podría contradecir ciertos principios fundamentales de la probabilidad, los cuales se describirán más adelante.

La *σ-álgebra* resuelve este problema al definir una clase (es decir, un conjunto de conjuntos) que incluye únicamente los eventos sobre los cuales se pueden realizar operaciones entre los conjuntos como la unión, intersección y complemento de manera consistente. Esta clase garantiza que la medida de probabilidad esté bien definida y que las asignaciones de probabilidad respeten los axiomas fundamentales, permitiendo así construir un espacio de probabilidad coherente.

Intuitivamente, bajo el contexto de la teoría de la probabilidad una σ-álgebra es una colección de todos los eventos relevantes para el experimento a los que nos interesa asignarle probabilidad. Incluye el conjunto vacío y está cerrada bajo operaciones como la unión, intersección y complemento, lo que asegura que cualquiera de estas operaciones entre eventos dentro de la σ-álgebra produce un nuevo evento que también pertenece a ella, garantizando así la consistencia matemática.

En base a lo argumentado anteriormente, es posible construir un espacio de probabilidad coherente, que proporciona la estructura necesaria para trabajar con eventos y asignar probabilidades de forma consistente, así, se presenta la definición

formal de un espacio de probabilidad.

Definición 2.3 (Espacio de Probabilidad) Un *espacio de probabilidad* es la tripleta $(\Omega, \mathcal{F}, \mathbb{P})$, donde:

- Ω es el conjunto de todos los posibles resultados (espacio muestral).
- $\mathcal{F}$ es una σ-álgebra, que contiene los eventos definidos sobre Ω.
- $\mathbb{P}$ es una medida de probabilidad que asigna a cada evento de $\mathcal{F}$ un valor en $[0, 1]$.

Estos axiomas aseguran que las probabilidades asignadas a los eventos en la σ-álgebra sean consistentes con las operaciones entre conjuntos y respeten las reglas fundamentales de la teoría de la probabilidad.

2.5 Propiedades de la Probabilidad

A partir de los axiomas y la definición de espacio de probabilidad, se derivan varias propiedades útiles que nos permiten analizar relaciones entre eventos y simplificar el cálculo de las probabilidades:

- **Probabilidad del Complemento:** La probabilidad de que un evento no ocurra (A^c) es igual a uno menos la probabilidad de que ocurra (A). Intuitivamente, dado que A y su complemento A^c son mutuamente excluyentes (disjuntos) y juntos abarcan todo el espacio muestral, sus probabilidades deben sumar 1:

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

- **Probabilidad del Vacío:** El vacío $\emptyset$ representa la imposibilidad absoluta, es decir, no contiene ningún resultado del espacio muestral. Por lo tanto, la probabilidad de que no ocurra algún elemento del espacio muestral es cero:

$$\mathbb{P}(\emptyset) = 0.$$

- **Monotonía:** Si un evento A está contenido dentro de otro evento B, significa que siempre que ocurra A, también ocurrirá B. Por esta razón, la probabilidad de A no puede ser mayor que la de B:

$$A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B).$$

- **Probabilidad de la Diferencia:** La diferencia entre dos eventos, denotada como $A \setminus B$, representa los resultados que están en A pero no en B. Su probabilidad se calcula restando la probabilidad de la intersección de A y B de la probabilidad de A:

$$\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B).$$

  Intuitivamente, esto refleja que al eliminar de A los resultados que comparte con B, obtenemos solo los resultados exclusivos de A.

- **Principio de Inclusión-Exclusión:** Este principio permite calcular la probabilidad de la unión de dos eventos A y B, teniendo en cuenta la intersección entre ellos. La fórmula para dos eventos es:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Para una *familia arbitraria* de eventos $\{A_i\}_{i \in I}$, el principio de inclusión-exclusión se generaliza de la siguiente manera:

$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{\emptyset \neq J \subseteq I} (-1)^{|J|+1} \mathbb{P}\left(\bigcap_{j \in J} A_j\right).$$

Esto es equivalente a, para n eventos $A_1, A_2, \ldots, A_n$, se tiene que:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n-1} \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right).$$

2.6 Ley Discreta Uniforme

El cálculo de probabilidades se simplifica considerablemente cuando los eventos de un experimento aleatorio son equiprobables. En estos casos, la probabilidad de un evento se obtiene simplemente contando los casos favorables y dividiéndolos entre el número total de casos posibles.

Definición 2.4 (Ley Discreta Uniforme) Si un espacio muestral finito Ω contiene $N = |\Omega|$ resultados igualmente probables, la probabilidad de un evento $A \subseteq \Omega$ se calcula como:

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

La *ley discreta uniforme* se aplica a espacios muestrales finitos, donde el cálculo de probabilidades se reduce a contar elementos, y todos los resultados tienen la misma probabilidad. Por lo tanto, es crucial contar correctamente tanto los elementos de A como los de Ω.

2.6.1 Espacios Infinitos Contables

En algunos casos, el espacio muestral puede ser infinito y aun así conservar la propiedad de equiprobabilidad. Por ejemplo, consideremos el conjunto de números enteros $\mathbb{Z}$. Si queremos calcular la probabilidad de que un número tomado al azar sea par, podemos definir el subconjunto de números pares en $\{-N, \ldots, N\}$ como:

$$A = \{k \in \{-N, \ldots, N\} \,|\, k = 2m,\ m \in \mathbb{Z}\}.$$

La probabilidad de seleccionar un número par se calcula como el límite de la proporción de números pares sobre el total de elementos en el espacio muestral:

$$\mathbb{P}(\text{número par}) = \lim_{N \to \infty} \frac{|A|}{|\{-N, \ldots, N\}|}.$$

El número de números pares en $\{-N, \ldots, N\}$ es $\lfloor N/2 \rfloor + 1$, ya que cada número par en este rango puede representarse como $2m$ con $m \in \{-\lfloor N/2 \rfloor, \ldots, \lfloor N/2 \rfloor\}$, Sin embargo, para valores grandes de N, $\lfloor N/2 \rfloor$ se aproxima a $N/2$. Además, el tamaño total del espacio muestral es $2N + 1$. Por lo tanto:

$$\mathbb{P}(\text{número par}) = \lim_{N \to \infty} \frac{N+1}{2N+1}.$$

Simplificando y dividiendo numerador y denominador por N, obtenemos:

$$\mathbb{P}(\text{número par}) = \lim_{N \to \infty} \frac{1 + \frac{1}{N}}{2 + \frac{1}{N}} = \frac{1}{2}.$$

Sin embargo, si el subconjunto A que nos interesa tiene cardinalidad finita mientras que $|\Omega| = \infty$, entonces la probabilidad de A será:

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \lim_{N \to \infty} \frac{|A|}{2N+1} = 0.$$

Esto ocurre porque, en un espacio infinito contable, cualquier subconjunto finito o con una cardinalidad menor que la de Ω tendrá probabilidad cero.

2.6.2 Espacios Infinitos No Contables

Cuando el espacio muestral es infinito y no contable, como el intervalo $[0, 1]$, el concepto de equiprobabilidad cambia. Aquí, las probabilidades no se basan en el conteo, sino en la *medida de Lebesgue*, que asigna longitudes, áreas o volúmenes a los subconjuntos del espacio.

Por ejemplo, si queremos calcular la probabilidad de seleccionar un punto específico, como $x = 0{,}5$, aunque un punto tiene longitud nula, podemos aproximar su probabilidad considerando un intervalo pequeño alrededor de 0,5, como $[0{,}5 - \epsilon, 0{,}5 + \epsilon]$. La probabilidad de este intervalo es proporcional a su longitud:

$$\mathbb{P}([0{,}5 - \epsilon, 0{,}5 + \epsilon]) = 2\epsilon.$$

Al tomar el límite cuando $\epsilon \to 0$, obtenemos:

$$\mathbb{P}(\{0{,}5\}) = \lim_{\epsilon \to 0} 2\epsilon = 0.$$

Esto muestra que, en un espacio continuo, los puntos individuales siempre tienen probabilidad cero, ya que su "tamaño" según la medida de Lebesgue es nulo.

Por otro lado, la probabilidad de un segmento dentro del intervalo $[0, 1]$, como $[a, b]$ con $0 \leq a < b \leq 1$, se calcula directamente a partir de su longitud:

$$\mathbb{P}([a, b]) = b - a.$$

Esto evidencia que, aunque los puntos tienen probabilidad cero, los subconjuntos con extensión positiva (como segmentos) pueden tener probabilidades no nulas.

Capítulo 3

Probabilidad Condicional e Independencia

¿Qué sucede cuando obtenemos nueva información sobre un experimento aleatorio? ¿Cómo cambia la probabilidad de un evento si sabemos que otro ya ha ocurrido? ¿Y qué pasa si los eventos no se afectan entre sí? Estas preguntas nos llevan al estudio de la probabilidad condicional, la cuál permite "reformular" las probabilidades de un evento en función de información adicional. A lo largo de este capítulo, exploraremos su definición, sus propiedades y los teoremas fundamentales que nos ayudarán a comprender las relaciones entre eventos en distintos contextos.

3.1 Probabilidad Condicional

Definición 3.1 (Probabilidad Condicional) Sean A y B eventos del mismo espacio muestral. *La probabilidad condicional* de A dado B, denotada $\mathbb{P}(A|B)$, se define como:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{cuando } \mathbb{P}(B) > 0.$$

La intuición detrás de la probabilidad condicional radica en que, al saber que el evento B ha ocurrido, debemos restringir nuestra atención al conjunto B, ya que todos los resultados fuera de él son irrelevantes. En este nuevo espacio muestral reducido, para que ocurra el evento A, el resultado debe pertenecer a $A \cap B$.

Sin embargo, al trabajar en un espacio muestral reducido, necesitamos redefinir las probabilidades de manera que cumplan con los principios fundamentales, en particular que la probabilidad total del espacio reducido sea 1. Esto se logra mediante un proceso de "normalización", en donde dividimos $\mathbb{P}(A \cap B)$ entre $\mathbb{P}(B)$, asegurando que la probabilidad de B en este nuevo espacio sea $\mathbb{P}(B|B) = 1$, garantizando así que este nuevo espacio de probabilidad reducido cumpla con los axiomas de probabilidad.

Es importante señalar que $\mathbb{P}(A|B)$ no está definida cuando $\mathbb{P}(B) = 0$, ya que no tendría sentido hablar de la probabilidad de A condicionado a un evento que nunca ocurre.

Ejemplo:

Supongamos que lanzamos un dado justo. El espacio muestral es:

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Sea el evento A: "Obtener un número par", y B: "Obtener un número mayor que 3", definidos como:

$$A = \{2, 4, 6\}, \quad B = \{4, 5, 6\}.$$

La intersección de ambos eventos es:

$$A \cap B = \{4, 6\}.$$

Dado que el dado es justo y cada resultado tiene la misma probabilidad bajo la ley discreta uniforme, calculamos:

$$\mathbb{P}(B) = \frac{3}{6} = \frac{1}{2}, \quad \mathbb{P}(A \cap B) = \frac{2}{6} = \frac{1}{3}.$$

La probabilidad condicional es entonces:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{3}.$$

Por lo tanto, dado que sabemos que el número es mayor que 3, la probabilidad de que sea par es $\frac{2}{3}$.

3.1.1 Propiedades

La probabilidad condicional cumple las mismas propiedades enunciadas en el capítulo anterior, con la diferencia de que las probabilidades están restringidas al evento ocurrido. A continuación, se enuncian algunas de las principales propiedades de la probabilidad condicional para tres eventos A, B y C, con $\mathbb{P}(C) > 0$:

- **Complemento:** La probabilidad de que no ocurra A, dado que C ocurre, es:

$$\mathbb{P}(A^c|C) = 1 - \mathbb{P}(A|C).$$

- **Vacío:** La probabilidad del vacío, dado que C ocurre, es siempre cero:

$$\mathbb{P}(\emptyset|C) = 0.$$

- **Cota superior:** La probabilidad condicional de cualquier evento está acotada por 1:

$$\mathbb{P}(A|C) \leq 1.$$

- **Diferencia de conjuntos:** La probabilidad condicional de A excluyendo B, dado C, es:

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

- **Unión de eventos:** La probabilidad condicional de la unión de dos eventos A y B, dado C, es:

$$\mathbb{P}(A \cup B|C) = \mathbb{P}(A|C) + \mathbb{P}(B|C) - \mathbb{P}(A \cap B|C).$$

- **Monotonía:** Si $A \subseteq B$, entonces la probabilidad condicional de A, dado C, es menor o igual que la de B:

$$\mathbb{P}(A|C) \leq \mathbb{P}(B|C).$$

3.2 Independencia de Eventos

Definición 3.2 (Independencia) Dos eventos A y B son *independientes* si y solo si:

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

La intuición detrás de la independencia de eventos radica en determinar si el conocimiento de que el evento B ha ocurrido influye en la probabilidad del evento A. Dos eventos A y B se consideran independientes cuando la ocurrencia de B, es decir, disponer de esta información adicional, no altera la probabilidad de A. En términos de probabilidad condicional, esto se puede escribir de la siguiente forma:

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

Al reorganizar esta expresión utilizando la definición de probabilidad condicional, obtenemos la condición de independencia:

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

Esta igualdad nos dice que, para eventos independientes, la probabilidad de que ambos ocurran simultáneamente se calcula simplemente multiplicando sus probabilidades individuales. Es importante destacar que, si $\mathbb{P}(B) = 0$, no es posible evaluar la independencia, ya que la probabilidad condicional $\mathbb{P}(A|B)$ no está definida en este caso. Intentar analizar la independencia bajo esta condición conduciría nuevamente a una contradicción, pues estaríamos considerando un evento que nunca ocurre.

Ejemplo:

Consideremos el experimento de lanzar dos monedas. El espacio muestral es:

$$\Omega = \{(C, C), (C, S), (S, C), (S, S)\},$$

donde C representa cara y S sello.

Sea el evento A: "La primera moneda muestra cara", y B: "La segunda moneda muestra cara". Estos eventos se definen como:

$$A = \{(C, C), (C, S)\}, \quad B = \{(C, C), (S, C)\}.$$

Bajo la ley discreta uniforme, cada resultado en Ω tiene probabilidad $\frac{1}{4}$. Calculamos las probabilidades:

$$\mathbb{P}(A) = \frac{2}{4} = \frac{1}{2}, \quad \mathbb{P}(B) = \frac{2}{4} = \frac{1}{2}, \quad \mathbb{P}(A \cap B) = \frac{1}{4}.$$

Verificamos la independencia de A y B:

$$\mathbb{P}(A \cap B) = \frac{1}{4}, \quad \mathbb{P}(A) \cdot \mathbb{P}(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Como $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$, concluimos que A y B son independientes.

Independencia de una Cantidad Arbitraria de Eventos

La definición de independencia puede extenderse a una cantidad arbitraria de eventos. Para un conjunto de n eventos $\{A_1, A_2, \ldots, A_n\}$, decimos que son independientes si se cumplen *todas* las siguientes condiciones:

- Para cada par de eventos A_i, A_j, con $i \neq j$:

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j).$$

- Para cada tripleta de eventos A_i, A_j, A_k, con $i \neq j \neq k$:

$$\mathbb{P}(A_i \cap A_j \cap A_k) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j) \cdot \mathbb{P}(A_k).$$

$$\vdots$$

- Para el conjunto completo de n eventos:

$$\mathbb{P}(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} \mathbb{P}(A_i).$$

Por ejemplo, consideremos tres eventos A, B y C. Decimos que estos eventos son independientes si se verifican las siguientes cuatro condiciones:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B), \quad \mathbb{P}(A \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(C), \quad \mathbb{P}(B \cap C) = \mathbb{P}(B) \cdot \mathbb{P}(C),$$

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

Es fundamental señalar que no basta con que se cumplan algunas de estas condiciones. Por ejemplo, podrían darse casos en los que A, B y C sean independientes a pares, es decir:

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

pero no se cumpla la condición conjunta:

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

En este caso, los eventos A, B y C no son independientes.

De igual forma, no es suficiente que solo se cumpla la última condición (la independencia conjunta) si las demás no se verifican. La independencia requiere que todas las condiciones se cumplan simultáneamente.

3.2.1 Propiedades

Si los eventos A y B son independientes, entonces también se cumple que:

- A y B^c son independientes,
- A^c y B son independientes,
- A^c y B^c son independientes.

Observación. Si A y B son disjuntos, si y solo si no son independientes.

Para ver esto consideremos dos eventos A y B, con $\mathbb{P}(A) \neq 0$ y $\mathbb{P}(B) \neq 0$. Si A y B son disjuntos, se cumple que:

$$A \cap B = \emptyset.$$

Por lo tanto,

$$\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0.$$

Sin embargo, si A y B fueran independientes, debería cumplirse que:

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

Como $\mathbb{P}(A) \neq 0$ y $\mathbb{P}(B) \neq 0$, tenemos que:

$$\mathbb{P}(A)\mathbb{P}(B) > 0.$$

Esto contradice el hecho de que $\mathbb{P}(A \cap B) = 0$. Por lo tanto, se concluye que A y B no son independientes si son disjuntos.

A partir de las definiciones de probabilidad condicional e independencia, se derivan diversos teoremas y propiedades que permiten analizar eventos complejos y sus interacciones. A continuación, se examinarán los resultados que surgen directamente de la probabilidad condicional y la independencia.

Proposición 3.1

Si $A_1, A_2, \ldots, A_n$ son eventos independientes, entonces la probabilidad de la unión de estos eventos se calcula como:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \prod_{i=1}^{n} \mathbb{P}(A_i^c) = 1 - \prod_{i=1}^{n} \left(1 - \mathbb{P}(A_i)\right).$$

Demostración: Considerando la independencia de los eventos $A_1, A_2, \ldots, A_n$. Según las leyes de De Morgan, la probabilidad de que ninguno de estos eventos ocurra es equivalente a la probabilidad del complemento de su unión:

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i^c\right) = \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right)^c.$$

Dado que los eventos $A_1, A_2, \ldots, A_n$ son independientes, sus complementos $A_1^c, A_2^c, \ldots, A_n^c$ también lo son. Por lo tanto, la probabilidad de la intersección de sus complementos se puede descomponer como el producto de las probabilidades individuales:

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i^c\right) = \prod_{i=1}^{n} \mathbb{P}(A_i^c).$$

Recordemos que la probabilidad del complemento de un evento se relaciona con la probabilidad del evento como:

$$\mathbb{P}(A_i^c) = 1 - \mathbb{P}(A_i).$$

Sustituyendo esta relación en la ecuación anterior, obtenemos:

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i^c\right) = \prod_{i=1}^{n} \left(1 - \mathbb{P}(A_i)\right).$$

Finalmente, la probabilidad de la unión de los eventos se calcula como el complemento de la probabilidad de que ninguno ocurra:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \mathbb{P}\left(\bigcap_{i=1}^{n} A_i^c\right).$$

Sustituyendo el resultado anterior:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \prod_{i=1}^{n} \left(1 - \mathbb{P}(A_i)\right).$$

3.3 Regla de la Cadena

En muchos casos, la probabilidad de la intersección de varios eventos puede resultar difícil de calcular directamente. Sin embargo, es posible descomponer dicha intersección mediante una serie de probabilidades condicionales, lo que facilita su evaluación.

Teorema 3.1 Regla de la Cadena

La probabilidad conjunta de varios eventos puede descomponerse en términos de probabilidades condicionales:

$$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdot \ldots \cdot \mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1}).$$

Demostración: Utilizaremos la definición de probabilidad condicional y el hecho de que la probabilidad conjunta de dos eventos arbitrarios A_1 y A_2 se expresa como:

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1).$$

Ahora extendemos esto al caso de n eventos. Por asociatividad tenemos:

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}((A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \cap A_n).$$

Aplicando la probabilidad condicional, esto se puede escribir como:

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \cdot \mathbb{P}(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

Repetimos este proceso recursivamente para descomponer $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_{n-1})$:

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_{n-1}|A_1 \cap \cdots \cap A_{n-2}).$$

Sustituyendo esta expresión en la anterior, obtenemos finalmente:

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1}).$$

3.4 Ley de Probabilidades Totales

La Ley de Probabilidades Totales nos permite calcular la probabilidad de un evento descomponiendo el espacio muestral en diferentes subconjuntos que forman una partición. En esencia, esta ley se basa en dividir un evento B en su intersección con cada subconjunto de la partición y luego sumar las probabilidades correspondientes. Es especialmente útil cuando conocemos cómo se comporta B bajo diferentes condiciones o escenarios.

Teorema 3.2 Ley de Probabilidades Totales

Sea $\{A_1, A_2, \ldots, A_n\}$ una *partición* del espacio muestral Ω. Para cualquier evento B:

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i) = \sum_{i=1}^{n} \mathbb{P}(B|A_i) \cdot \mathbb{P}(A_i).$$

Demostración: Consideremos la partición de Ω, $\{A_1, A_2, \ldots, A_n\}$ y que cualquier evento $B \subseteq \Omega$ puede expresarse como la unión disjunta de sus intersecciones con los A_i:

$$B = (B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n).$$

Por el axioma de aditividad de la probabilidad, se tiene que:

$$\mathbb{P}(B) = \mathbb{P}\left(\bigcup_{i=1}^{n}(B \cap A_i)\right) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i).$$

Finalmente, utilizando la definición de probabilidad condicional, sabemos que:

$$\mathbb{P}(B \cap A_i) = \mathbb{P}(B|A_i) \cdot \mathbb{P}(A_i).$$

Sustituyendo esto en la suma anterior, obtenemos:

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i) \cdot \mathbb{P}(A_i).$$

Ejemplo:

Supongamos que una persona toma un tren desde tres posibles estaciones (A_1, A_2, A_3), con las siguientes probabilidades:

$$\mathbb{P}(A_1) = 0{,}5, \quad \mathbb{P}(A_2) = 0{,}3, \quad \mathbb{P}(A_3) = 0{,}2.$$

Además, la probabilidad de que llegue a tiempo (B) depende de la estación desde donde parte:

$$\mathbb{P}(B|A_1) = 0{,}8, \quad \mathbb{P}(B|A_2) = 0{,}6, \quad \mathbb{P}(B|A_3) = 0{,}9.$$

Aplicando la Ley de Probabilidades Totales, calculamos:

$$\mathbb{P}(B) = \mathbb{P}(B|A_1) \cdot \mathbb{P}(A_1) + \mathbb{P}(B|A_2) \cdot \mathbb{P}(A_2) + \mathbb{P}(B|A_3) \cdot \mathbb{P}(A_3).$$

Sustituyendo los valores:

$$\mathbb{P}(B) = 0{,}8 \cdot 0{,}5 + 0{,}6 \cdot 0{,}3 + 0{,}9 \cdot 0{,}2 = 0{,}76.$$

3.5 Teorema de Bayes

El Teorema de Bayes nos permite recalcular la probabilidad de un evento después de obtener nueva información. Comenzamos con una probabilidad inicial (probabilidad *a priori*) y, al incorporar datos adicionales que afectan el evento, ajustamos esa probabilidad para obtener una estimación más precisa. Es especialmente útil en situaciones donde disponemos de información parcial o condicional sobre el evento en cuestión.

Teorema 3.3 Teorema de Bayes

La probabilidad condicional de un evento A dado B es tal que:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B)}, \quad \text{con } \mathbb{P}(B) > 0.$$

Donde:

- $\mathbb{P}(A)$ Probabilidad a priori (*prior*). Es la probabilidad inicial del evento A antes de considerar la nueva información B.
- $\mathbb{P}(B|A)$ Verosimilitud (*likelihood*). Es la probabilidad de observar B bajo la suposición de que A ocurre.
- $\mathbb{P}(B)$ Evidencia (*evidence*). Es la probabilidad total de B, considerando todos los escenarios posibles.
- $\mathbb{P}(A|B)$: Probabilidad a posteriori (*posteriori*). Es la probabilidad "actualizada"de A tras incorporar la nueva información B.

Demostración: Por definición, la probabilidad de A dado B es:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{con } \mathbb{P}(B) > 0.$$

De manera similar, podemos expresar la probabilidad conjunta $\mathbb{P}(A \cap B)$ en términos de B dado A:

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

Sustituyendo esta última ecuación en la definición de $\mathbb{P}(A|B)$, obtenemos:

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

Teorema de Bayes para una Partición

Si el evento B puede ocurrir bajo distintas condiciones representadas por una partición del espacio muestral Ω, denotada como $\{H_1, H_2, \ldots, H_n\}$, entonces, según la Ley de Probabilidades Totales, se tiene:

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|H_i) \cdot \mathbb{P}(H_i).$$

Sustituyendo esta expresión en la fórmula del Teorema de Bayes, obtenemos:

$$\mathbb{P}(H_k|B) = \frac{\mathbb{P}(B|H_k) \cdot \mathbb{P}(H_k)}{\sum_{i=1}^{n} \mathbb{P}(B|H_i) \cdot \mathbb{P}(H_i)},$$

donde H_k es uno de los eventos de la partición.

Ejemplo:

Supongamos que el 1 % de la población tiene una enfermedad (A), y una prueba médica tiene un 99 % de precisión, es decir:

$$\mathbb{P}(B|A) = 0{,}99, \quad \mathbb{P}(A) = 0{,}01.$$

Por otro lado, la probabilidad de obtener un falso positivo ($B|A^c$) también es del 1 %:

$$\mathbb{P}(B|A^c) = 0{,}01.$$

Queremos calcular la probabilidad de que una persona esté enferma dado que la prueba resultó positiva ($A|B$). Aplicamos el Teorema de Bayes:

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

Primero calculamos $\mathbb{P}(B)$ usando la Ley de Probabilidades Totales:

$$\mathbb{P}(B) = \mathbb{P}(B|A) \cdot \mathbb{P}(A) + \mathbb{P}(B|A^c) \cdot \mathbb{P}(A^c).$$

Sustituyendo los valores:

$$\mathbb{P}(B) = 0{,}99 \cdot 0{,}01 + 0{,}01 \cdot 0{,}99 = 0{,}0198.$$

Ahora sustituimos todo en el Teorema de Bayes:

$$\mathbb{P}(A|B) = \frac{0{,}99 \cdot 0{,}01}{0{,}0198} \approx 0{,}5.$$

Esto indica que, aunque la prueba es altamente precisa, la probabilidad de que una persona esté enferma, dado que obtuvo un resultado positivo, es del 50 %.

3.6 Independencia Condicional

Dos eventos A y B son independientes si la ocurrencia de uno no afecta la probabilidad de que ocurra el otro. La independencia condicional por su parte evalúa la relación de independencia entre dos eventos A y B bajo la condición de que un tercer evento C haya ocurrido. En este contexto, decimos que A y B son independientes condicionalmente dado C si el conocimiento de A no altera la probabilidad de B dentro del marco en que C es cierto, y viceversa.

Definición 3.3 (Independencia Condicional) Dos eventos A y B son *independientes condicionalmente* dado C si:

$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C) \cdot \mathbb{P}(B|C), \quad \text{con } \mathbb{P}(C) > 0.$$

La definición anterior se puede deducir a través de la definición de probabilidad condicional e independencia. Si A es condicionalmente independiente de B dado el evento C, la probabilidad de A dado que ocurrió B y C es:

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

Luego, por definición de probabilidad condicional:

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

Así, se obtiene finalmente que:

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

3.6.1 Propiedades

Si A y B son condicionalmente independientes dado C, entonces:

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

Observación. A partir de esta última expresión, y recordando la definición de independencia, es importante tener en cuenta lo siguiente:

- La independencia condicional entre A y B dado C implica que, dentro del contexto en el que C ocurre, los eventos A y B no se afectan mutuamente. Sin embargo, esto no significa que A y B sean independientes sin considerar la condición de que no haya ocurrido el evento C.
- De manera similar, el hecho de que A y B sean independientes sin la condición de algún evento previo, no garantiza que lo sean cuando se condicionan al evento C.

Capítulo 4

Principio de Conteo

El cálculo de probabilidades en espacios muestrales finitos está profundamente ligado a los principios de conteo, que nos permiten contar de manera metódica los posibles resultados de un experimento. El principio de conteo se basa en un enfoque de "divide y vencerás", en el que el conteo se divide en etapas mediante el uso de un árbol. Dependiendo de la naturaleza del problema, el conteo puede variar según diversos factores, como si el orden de los elementos es relevante o si se permite la repetición.

Estos principios nos permiten aplicar la ley discreta uniforme de manera correcta, ya que el cálculo de probabilidades a menudo se reduce a determinar el tamaño del evento de interés y el del espacio muestral. A lo largo de este capítulo, exploraremos diversas formas de conteo, como las permutaciones y combinaciones, y su papel en el calculo de las probabilidades.

En problemas de conteo, es fundamental tener una terminología precisa para evitar ambigüedades. A continuación, presentamos conceptos que se utilizarán más adelante:

- **Muestreo:** El muestreo consiste en seleccionar elementos de un conjunto. Al realizar un muestreo al azar, todos los elementos del conjunto tienen la misma probabilidad de ser seleccionados.
- **Con o sin reemplazo:** Cuando realizamos múltiples extracciones de un conjunto:
  - Si cada elemento extraído es devuelto al conjunto antes de la siguiente extracción, hablamos de *muestreo con reemplazo*. En este caso, es posible que un mismo elemento sea seleccionado varias veces. Por ejemplo, si $A = \{a_1, a_2, a_3, a_4\}$ y seleccionamos tres elementos con reemplazo, un posible resultado es (a_3, a_1, a_3).
  - Si los elementos seleccionados no se devuelven al conjunto, hablamos de *muestreo sin reemplazo*. En este caso, cada elemento solo puede ser seleccionado una vez.
- **Ordenado o no ordenado:** Si el orden en el que se seleccionan los elementos es relevante (es decir, $(a_1, a_2, a_3) \neq (a_2, a_3, a_1)$), se trata de *muestreo ordenado*. Por otro lado, si el orden no importa, se denomina *muestreo no ordenado*.

De acuerdo a lo anterior, al hablar de muestreo podemos clasificarlo en las siguientes cuatro combinaciones:

- Muestreo ordenado con reemplazo.
- Muestreo ordenado sin reemplazo.

- Muestreo no ordenado con reemplazo.
- Muestreo no ordenado sin reemplazo.

A continuación exploraremos más sobre cada uno de estos conceptos y su aplicación.

4.1 Principios Combinatoriales

Los principios combinatorios son conceptos en el conteo de estructuras discretas. Nos permiten determinar de manera eficiente el número total de configuraciones posibles en un problema, evitando la necesidad de contar caso por caso. Entre estos principios se encuentran el *principio de adición*, el *principio del palomar* y el *principio multiplicativo*, cada uno aplicable en diferentes contextos.

El *principio de adición* establece que si una tarea puede realizarse de m maneras y otra tarea, mutuamente excluyente, puede realizarse de n maneras, entonces el número total de formas de realizar una de ellas es $m + n$. Este principio es útil cuando existen opciones alternativas y el resultado final proviene de elegir una entre varias opciones disjuntas. Por ejemplo, si tenemos 3 maneras de elegir un libro de matemáticas y 5 maneras de elegir un libro de física, y solo podemos elegir uno de los dos, entonces tenemos un total de $3 + 5 = 8$ opciones.

Por otro lado, el *principio del palomar* (también llamado principio de Dirichlet) establece que si se distribuyen más objetos que compartimentos, entonces al menos un compartimento debe contener más de un objeto. Un ejemplo clásico es el de los casilleros: si tenemos 10 estudiantes y solo 9 casilleros disponibles, al menos un casillero debe ser compartido por dos estudiantes. Este principio es especialmente útil en demostraciones de existencia y en problemas de teoría de números y combinatoria avanzada.

En este capítulo, nos enfocaremos en el *principio multiplicativo*, ya que es un método fundamental en el conteo y en la determinación de espacios muestrales.

4.1.1 Principio Multiplicativo

El Principio Multiplicativo nos permite determinar el número total de resultados posibles en experimentos compuestos, donde cada experimento tiene un conjunto finito de resultados. Formalmente, se define como:

Definición 4.1 (Principio Multiplicativo) El número total de resultados posibles de r experimentos sucesivos se calcula multiplicando el número de resultados posibles de cada experimento:

$$n_1 \cdot n_2 \cdot \ldots \cdot n_r.$$

Este se aplica cuando realizamos varios experimentos sucesivos, cada uno con un conjunto de posibles resultados. Proporciona una manera directa de contar el número total de combinaciones posibles multiplicando el número de resultados de cada experimento.

El producto cartesiano nos permite describir explícitamente estas combinaciones en términos de pares, tríos o r-tuplas de resultados. Formalmente, dado que el primer

experimento tiene un conjunto de resultados S_1 y el segundo tiene S_2, el producto cartesiano se define como:

$$S_1 \times S_2 = \{(s_1, s_2) \mid s_1 \in S_1,\ s_2 \in S_2\}.$$

Para r experimentos, el producto cartesiano generaliza esta idea al generar todas las r-tuplas posibles:

$$S_1 \times S_2 \times \cdots \times S_r = \{(s_1, s_2, \ldots, s_r) \mid s_1 \in S_1,\ s_2 \in S_2,\ \ldots,\ s_r \in S_r\}.$$

La cardinalidad del producto cartesiano, es decir, el número de r-tuplas generadas, coincide con el número total de resultados que se obtienen bajo el Principio Multiplicativo:

$$|S_1 \times S_2 \times \cdots \times S_r| = |S_1| \cdot |S_2| \cdot \cdots \cdot |S_r|.$$

Ejemplo:

Supongamos que lanzamos dos dados. El conjunto de posibles resultados para cada dado es $S_1 = S_2 = \{1, 2, 3, 4, 5, 6\}$. El producto cartesiano genera el conjunto de todos los pares ordenados:

$$S_1 \times S_2 = \{(1, 1), (1, 2), \ldots, (6, 6)\}.$$

La cardinalidad de este conjunto es:

$$|S_1 \times S_2| = 6 \cdot 6 = 36.$$

Es decir, el Principio Multiplicativo nos da el número total de combinaciones posibles, mientras que el producto cartesiano nos muestra todas esas combinaciones de manera explícita.

4.2 Muestreo Ordenado con Reemplazo

Cuando aplicamos el Principio Multiplicativo en situaciones donde se permiten repeticiones, la tarea de contar se simplifica considerablemente.

Definición 4.2 (Muestreo Ordenado con Reemplazo) Si de un conjunto de tamaño n seleccionamos k elementos, permitiendo repeticiones y considerando el orden, el número de secuencias posibles es:

$$n \cdot n \cdot \ldots \cdot n = n^k.$$

El muestreo ordenado con reemplazo permite seleccionar elementos de un conjunto, con la posibilidad de repetir un mismo elemento en diferentes posiciones y donde el orden en que se seleccionan los elementos importa. Esto se refleja en que, para cada posición de la secuencia, cualquier elemento del conjunto puede ser elegido nuevamente.

La expresión n^k se deriva de la aplicación del Principio Multiplicativo. Para construir una secuencia de k elementos:

- Hay n opciones posibles para elegir el primer elemento.

- Después de cada elección, debido a que se permite el reemplazo, también hay n opciones para el segundo elemento.
- Este proceso se repite k veces, ya que seleccionamos un total de k elementos.

Por lo tanto, el número total de secuencias posibles es el producto de n opciones repetidas k veces:

$$n \cdot n \cdot \ldots \cdot n = n^k.$$

Ejemplo:

Supongamos que tenemos un conjunto $S = \{a, b, c\}$ de tamaño $n = 3$, y queremos formar secuencias de longitud $k = 2$.

- Para el primer elemento, podemos elegir entre a, b, c.
- Para el segundo elemento, nuevamente podemos elegir entre a, b, c, ya que se permite el reemplazo.

Las secuencias posibles son:

$$\{(a,a),(a,b),(a,c),(b,a),(b,b),(b,c),(c,a),(c,b),(c,c)\}.$$

En total, hay $3 \cdot 3 = 3^2 = 9$ secuencias posibles.

4.3 Muestreo Ordenado sin Reemplazo

Al aplicar el Principio Multiplicativo en situaciones donde no se permiten repeticiones, el conteo se ajusta para reflejar que cada elección reduce las opciones disponibles para las siguientes.

Definición 4.3 (Muestreo Ordenado sin Reemplazo) Si de un conjunto de tamaño n seleccionamos k elementos sin permitir repeticiones y considerando el orden, el número de secuencias posibles es:

$$n \cdot (n-1) \cdot \cdots \cdot (n-k+1) = \frac{n!}{(n-k)!} = P_k^n.$$

En el muestreo ordenado sin reemplazo, seleccionamos elementos de un conjunto sin repetir ninguno, y el orden de selección importa. Esto significa que, a medida que seleccionamos cada elemento, las opciones disponibles disminuyen porque no se permite volver a elegir el mismo elemento.

La expresión $n \cdot (n-1) \cdot \cdots \cdot (n-k+1)$ puede entenderse como:

- Para el primer elemento, hay n opciones posibles.
- Para el segundo elemento, quedan $n-1$ opciones, ya que el primer elemento no puede volver a ser seleccionado.
- Para el tercer elemento, hay $n-2$ opciones, y así sucesivamente.
- Esto continúa hasta que se han seleccionado k elementos.

Por lo tanto, el número total de secuencias es:

$$n \cdot (n-1) \cdot \ldots \cdot (n-k+1).$$

Esta expresión puede escribirse de forma compacta utilizando factoriales:

$$\frac{n!}{(n-k)!}.$$

es decir el número de *permutaciones* de k elementos seleccionados de un conjunto de n elementos.

Ejemplo:

Supongamos que tenemos un conjunto $S = \{a, b, c, d\}$ de tamaño $n = 4$, y queremos formar secuencias de longitud $k = 2$:

- Para el primer elemento, hay 4 opciones: a, b, c, d.
- Para el segundo elemento, quedan 3 opciones, ya que no se permite repetir el primer elemento.

Las secuencias posibles son:

$$\{(a,b),(a,c),(a,d),(b,a),(b,c),(b,d),(c,a),(c,b),(c,d),(d,a),(d,b),(d,c)\}.$$

En total, hay $4 \cdot 3 = 12$ secuencias posibles, que se calcula como:

$$\frac{4!}{(4-2)!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} = 12.$$

4.4 Muestreo no Ordenado sin Reemplazo

Cuando el orden de selección no importa y no se permiten repeticiones, no podemos simplemente aplicar el Principio Multiplicativo. En este caso, necesitamos contar las selecciones únicas, ignorando las permutaciones de los elementos seleccionados.

Definición 4.4 (Muestreo no Ordenado sin Reemplazo) Si de un conjunto de n elementos seleccionamos k elementos sin considerar el orden y sin permitir repeticiones, el número de combinaciones posibles es:

$$\binom{n}{k} = \frac{n!}{(n-k)!\,k!}.$$

El muestreo no ordenado sin reemplazo, también conocido como *combinatoria*, tiene como objetivo seleccionar subconjuntos de tamaño k de un conjunto de n elementos, donde el orden no importa y no se permiten repeticiones.

La expresión $\binom{n}{k}$, conocida como *coeficiente binomial*, cuenta el número de formas en que se pueden elegir k elementos de n posibles. Para entender la intuición de dicha formula consideremos lo siguiente:

- Si el orden importara, el número de formas de elegir k elementos sería $n \cdot (n-1) \cdots (n-k+1)$, lo que equivale a $\frac{n!}{(n-k)!}$.

- Sin embargo, cuando el orden no importa, cada subconjunto de k elementos puede aparecer en $k!$ disposiciones diferentes, ya que hay $k!$ formas de permutar los k elementos. Por ejemplo, el subconjunto $\{a, b, c\}$ puede representarse de las siguientes maneras si el orden importara: (a, b, c), (a, c, b), (b, a, c), (b, c, a), (c, a, b), y (c, b, a), lo cual da un total de $3! = 6$ disposiciones.

  Para evitar esta redundancia en el conteo, dividimos entre $k!$, que es el número de permutaciones posibles de los k elementos seleccionados. Así, eliminamos las repeticiones y contamos cada subconjunto una sola vez, obteniendo la expresión:

$$\binom{n}{k} = \frac{n!}{(n-k)!\,k!}.$$

Ejemplo:

Supongamos que tenemos un conjunto $S = \{a, b, c, d\}$ con $n = 4$ elementos y deseamos elegir $k = 2$ elementos. Las combinaciones posibles son:

$$\{a, b\}, \{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}, \{c, d\}.$$

El número total de combinaciones es:

$$\binom{4}{2} = \frac{4!}{(4-2)!\,2!} = \frac{4 \cdot 3}{2 \cdot 1} = 6.$$

4.4.1 Propiedades

- **Simetría:** Elegir k elementos de un conjunto de n es equivalente a no elegir los otros $n-k$ elementos. Es decir, las combinaciones que incluyen k elementos son exactamente las mismas que las que excluyen $n - k$:

$$\binom{n}{k} = \binom{n}{n-k}.$$

- **Suma de combinaciones:** La suma de todas las combinaciones posibles de un conjunto de tamaño n, considerando todos los tamaños posibles de subconjuntos ($k = 0, 1, \ldots, n$), equivale al número total de subconjuntos de n. Esto corresponde al cardinal del conjunto potencia de n, que es 2^n:

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n.$$

  El resultado también puede ser deducido del desarrollo del Teorema del Binomio de Newton.

- **Relación de recurrencia:** Si agregamos un nuevo elemento a un conjunto de tamaño n, las combinaciones de tamaño $k+1$ pueden formarse de dos maneras:

  - Incluyendo el nuevo elemento ($k + 1$ elementos provienen del conjunto previo de tamaño k).
  - Excluyendo el nuevo elemento ($k+1$ elementos provienen de un conjunto previo de tamaño $k + 1$).

Esto se expresa como:

$$\binom{n+1}{k+1} = \binom{n}{k+1} + \binom{n}{k}.$$

- **Identidad de Vandermonde:** Si combinamos un conjunto de tamaño m y otro de tamaño n, las combinaciones de tamaño k en el conjunto total pueden formarse seleccionando i elementos del primero y $k-i$ del segundo, con i variando desde 0 hasta k. Esto se escribe como:

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

4.4.2 Coeficiente Multinomial

Cuando dividimos un conjunto de n elementos en varios grupos de tamaños específicos, necesitamos calcular de cuántas maneras podemos distribuir los elementos sin que se repitan. Este problema generaliza el concepto de combinaciones.

Definición 4.5 (Coeficiente Multinomial) Si de un conjunto de n elementos formamos r grupos de tamaños $n_1, n_2, \ldots, n_r$ sin repeticiones tal que se cumple que $\sum_{i \in R} n_i = n$, entonces el número de maneras posibles es:

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1!\, n_2! \cdots n_r!}.$$

El coeficiente multinomial es una generalización del coeficiente binomial y cuenta de cuántas maneras se pueden distribuir n elementos en r grupos de tamaños $n_1, n_2, \ldots, n_r$, respetando que el orden dentro de cada grupo no importa. La expresión es:

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1!\, n_2! \cdots n_r!},$$

donde se cumple que:

$$n_1 + n_2 + \cdots + n_r = n.$$

Para deducir dicha expresión, supongamos que tenemos n elementos distintos y queremos dividirlos en r grupos de tamaños $n_1, n_2, \ldots, n_r$. Para entender cómo se calcula el número de distribuciones únicas, analicemos paso a paso:

- Inicialmente, podemos permutar los n elementos de todas las formas posibles. Esto da lugar a $n!$ maneras, ya que consideramos todas las posibles ordenaciones de los elementos.

- Sin embargo, dentro de cada grupo, el orden de los elementos no importa. Por ejemplo, si un grupo contiene los elementos $\{A, B\}$, las permutaciones $\{A, B\}$ y $\{B, A\}$ se cuentan como la misma distribución. Por lo tanto, debemos dividir $n!$ entre el número de permutaciones posibles dentro de cada grupo:

$$n_i! \quad \text{para cada grupo } i.$$

- En total, debemos dividir $n!$ entre el producto de los factoriales de los tamaños de los grupos:
$$n_1! \cdot n_2! \cdots n_r!.$$
Esto elimina las repeticiones dentro de los grupos y asegura que cada distribución única se cuente solo una vez.

- Por lo tanto, el número de distribuciones únicas es:
$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1!\, n_2! \cdots n_r!}.$$

Es decir, primero consideramos todas las posibles formas de ordenar los n elementos. Luego, eliminamos las permutaciones redundantes dentro de cada grupo, lo que nos lleva a contar únicamente las distribuciones distintas.

Ejemplo:

Si $n = 5$ elementos deben ser distribuidos en tres grupos de tamaños $n_1 = 2$, $n_2 = 2$ y $n_3 = 1$, la cantidad de distribuciones posibles es:

$$\binom{5}{2, 2, 1} = \frac{5!}{2!\, 2!\, 1!}.$$

Calculando:

$$5! = 120, \quad 2! = 2, \quad 1! = 1.$$

Sustituyendo:

$$\binom{5}{2, 2, 1} = \frac{120}{2 \cdot 2 \cdot 1} = \frac{120}{4} = 30.$$

Esto indica que hay 30 formas de distribuir los 5 elementos en tres grupos de tamaños 2, 2 y 1.

4.5 Muestreo no Ordenado con Reemplazo

Cuando seleccionamos elementos de un conjunto, permitiendo repeticiones y sin importar el orden, el problema se traduce en contar cuántas formas distintas podemos repartir K elementos entre n categorías. Este tipo de conteo se resuelve utilizando un elemento combinatorio conocida como el *principio de estrellas y barras.*

Definición 4.6 (Muestreo no Ordenado con Reemplazo) El número de formas de seleccionar K elementos de un conjunto de n categorías, permitiendo repeticiones y sin considerar el orden, o bien, La cantidad de posibles soluciones al problema

$$\begin{cases} x_1 + x_2 + \cdots + x_n = K \\ x_i \geq 0 \ \forall i \in \{1, \cdots n\} \end{cases}$$

está dado por:

$$\binom{n + K - 1}{K} = \binom{n + K - 1}{n - 1}.$$

Este tipo de muestreo se puede interpretar como un problema de *distribuciones equivalentes*, donde buscamos distribuir K objetos indistinguibles en n categorías distintas.

Para visualizar esta idea, podemos imaginar que cada categoría está separada por una barra vertical. Por ejemplo, si $n = 3$ y $K = 4$, una posible distribución de los objetos se representa así:

$$\text{Distribución: } \bullet\,\bullet\,|\,\bullet\,|\,\bullet$$

Esto indica que la primera categoría tiene 2 elementos, la segunda tiene 1, y la tercera también tiene 1. Cada distribución de este tipo corresponde a una forma única de asignar los K elementos entre las n categorías.

$$\underbrace{\bullet\bullet}_{x_1=2} \,|\, \underbrace{\bullet}_{x_2=1} \,|\, \underbrace{\bullet}_{x_3=1}$$

En este contexto, el problema se reduce a organizar K elementos ($\bullet$) y $n-1$ separadores ($|$), lo que da un total de $K + n - 1$ objetos, es decir, $K + n - 1$ posibles posiciones para organizar los elementos. Elegir K posiciones para los elementos indistinguibles equivale a determinar las posiciones de las barras o separadores. Por lo tanto, el número total de configuraciones se calcula como:

$$\binom{n + K - 1}{K}.$$

Esto se relaciona directamente con el coeficiente binomial, ya que estamos contando las formas de elegir K posiciones para los elementos ($\bullet$) dentro de un conjunto de $K + n - 1$ posibles posiciones. Las posiciones restantes estarán necesariamente ocupadas por los separadores ($|$), los cuales dividen los elementos en categorías y determinan cuántos elementos pertenecen a cada una.

Ejemplo:

Supongamos que queremos formar grupos de 4 personas seleccionadas de 3 categorías posibles, permitiendo que una categoría tenga más de una persona o incluso ninguna. El número de configuraciones posibles es:

$$\binom{3 + 4 - 1}{4} = \binom{6}{4}.$$

Calculando:

$$\binom{6}{4} = \frac{6!}{4!\,(6-4)!} = \frac{6 \cdot 5}{2 \cdot 1} = 15.$$

Capítulo 5

Variables Aleatorias Discretas

A menudo es necesario asignar valores numéricos a los resultados de un experimento aleatorio para facilitar su análisis y comprensión. Por ejemplo, consideremos una carrera de Fórmula 1 como un experimento aleatorio. Podríamos estar interesados en diversas métricas numéricas relacionadas con la carrera, como el tiempo total de carrera de cada piloto, el número de vueltas completadas, la posición final, la velocidad máxima alcanzada, o incluso la cantidad de paradas en boxes realizadas. Cada una de estas medidas numéricas proporciona información específica sobre el resultado del experimento aleatorio, permitiendo un estudio más detallado de los datos generados por la carrera. Estos valores se definen mediante variables aleatorias, las cuales representan matemáticamente los resultados del experimento.

En este capítulo, nos centraremos en las variables aleatorias discretas, que toman un conjunto finito o contable de valores. Este tipo de variables es útil para modelar situaciones en las que los resultados posibles pueden enumerarse, como el número de caras obtenidas al lanzar una moneda varias veces o la cantidad de clientes que llegan a una tienda en una hora.

Exploraremos su definición formal, sus propiedades y las medidas asociadas, como la esperanza y la varianza, además de analizar distribuciones discretas comunes. A través de este estudio, podremos describir y predecir comportamientos en sistemas con resultados discretos.

5.1 Variable Aleatoria

Las variables aleatorias permiten representar de forma cuantitativa resultados abstractos o categóricos asociados a experimentos aleatorios, facilitando su análisis y manipulación matemática. Estas se asocian a valores numéricos, ya que esto permite utilizar algunas medidas que se verán a continuación como la función de masa de probabilidad (PMF), la función de distribución acumulada (CDF) y otras.

Definición 5.1 (Variable Aleatoria) Una *variable aleatoria* es una función que asigna un valor numérico a cada resultado posible de un experimento aleatorio. Formalmente:

$$X : \Omega \to \mathbb{R},$$

donde Ω es el espacio muestral del experimento.

En este texto, adoptaremos la convención de utilizar letras mayúsculas para denotar las variables aleatorias (X, Y, Z) y letras minúsculas para los valores específicos que estas pueden tomar (x, y, z).

5.2 Rango

El rango de una variable aleatoria describe todos los valores que dicha variable puede asumir con probabilidad no nula. Este concepto es fundamental para entender la distribución de una variable, ya que delimita el conjunto de posibles resultados numéricos que pueden observarse en un experimento aleatorio.

Definición 5.2 (Rango) El *rango* de una variable aleatoria X es el conjunto de valores posibles que puede tomar y se denota como:

$$R_X = \{x \in \mathbb{R} \mid \mathbb{P}(X = x) > 0\}.$$

Por ejemplo, al lanzar un dado podemos definir una variable aleatoria X que tome el valor de la cara obtenida, entonces:

$$\Omega = \{1, 2, 3, 4, 5, 6\}, \quad R_X = \{1, 2, 3, 4, 5, 6\}.$$

De forma similar, al lanzar una moneda podríamos definir una variable aleatoria Y con los valores:

$$R_Y = \{0, 1\},$$

donde $Y = 1$ representa "cara" y $Y = 0$ representa "sello".

5.3 Variable Aleatoria Discreta

Podemos diferenciar las variables aleatorias en dos categorías principales: aquellas cuyos valores pueden contarse y aquellas que no. Las primeras son conocidas como *variables aleatorias discretas*. Estas variables son ideales para modelar situaciones en las que los resultados son cuantificables, como el número de clientes que visitan una tienda, los defectos en una línea de producción, o el número de éxitos en una serie de intentos.

Definición 5.3 (Variable Aleatoria Discreta) Una *variable aleatoria discreta* X es aquella variable aleatoria cuyo rango R_X es un conjunto contable, es decir, un conjunto finito o infinitamente numerable.

Las variables aleatorias discretas se emplean para modelar situaciones donde los posibles resultados del experimento pueden listarse de manera explícita. Esto permite asignar probabilidades específicas a cada valor del rango, lo que facilita el análisis de eventos y el cálculo de probabilidades.

El hecho de que el rango sea contable permite describir completamente la distribución de X mediante una función de probabilidad, que asigna a cada valor posible x la probabilidad $\mathbb{P}(\{X = x\})$.

Ejemplo:

Supongamos que contamos el número de llamadas recibidas en una central telefónica en una hora. Definimos X como el número de llamadas. Aquí, el rango es:

$$R_X = \{0, 1, 2, 3, \dots\}.$$

Este conjunto es contable, por lo que X es una variable aleatoria discreta.

5.4 Función de Masa de Probabilidad (PMF)

Cuando trabajamos con variables aleatorias discretas, es esencial describir cómo se distribuyen sus valores posibles. Para ello, utilizamos la *función de masa de probabilidad (PMF)*, que asigna a cada valor $x \in R_X$ la probabilidad de que la variable aleatoria X tome dicho valor.

Definición 5.4 (Función de Masa de Probabilidad) Sea X una variable aleatoria discreta con rango $R_X = \{x_1, x_2, x_3, \dots\}$, donde R_X es un conjunto finito o contablemente infinito. La *función de masa de probabilidad (PMF*, por sus siglas en inglés) de X se define como:

$$P_X(x) = \mathbb{P}(X = x), \quad \forall x \in R_X.$$

La PMF asigna una probabilidad a cada valor posible que puede tomar la variable aleatoria X. Formalmente, es una *medida de probabilidad* definida en el conjunto de valores R_X, que satisface los axiomas y propiedades de un espacio de probabilidad.

Aunque la PMF se define habitualmente para valores en el rango, a veces es conveniente extender la PMF de X a todos los números reales. Si $x \notin R_X$, podemos escribir simplemente $P_X(x) = \mathbb{P}(X = x) = 0$. Por lo tanto, se tiene que:

$$P_X(x) = \begin{cases} \mathbb{P}(X = x) & \text{si } x \in R_X \\ 0 & \text{de lo contrario} \end{cases}$$

Rango

Si X es una variable aleatoria discreta, su rango R_X es un conjunto numerable de valores donde la función de masa de probabilidad es positiva:

$$R_X = \{x \mid P_X(x) > 0\}.$$

Propiedades

Como mencionamos anteriormente, la PMF es una medida de probabilidad, por lo que algunas de las propiedades que veremos a continuación pueden resultar evidentes. Sin embargo, es importante mencionarlas.

- **No negatividad:**

  $$0 \leq P_X(x) \leq 1, \quad \forall x \in R_X.$$

  Esto refleja que una probabilidad no puede ser negativa ni superar el valor de 1, ya que mide una proporción dentro del espacio muestral.

- **Suma total de probabilidades:**

  $$\sum_{x \in R_X} P_X(x) = 1.$$

  Así, asegura que la suma de todas las probabilidades asociadas a los valores posibles de X cubra todo el espacio muestral.

- **Probabilidad en subconjuntos:** Para cualquier conjunto $A \subseteq R_X$:

$$\mathbb{P}(X \in A) = \sum_{x \in A} P_X(x).$$

De esta forma, se permite calcular la probabilidad de que X tome valores en un subconjunto particular del rango, sumando las probabilidades individuales.

Ejemplo:

Sea X la variable aleatoria que representa el número obtenido al lanzar un dado justo. Su rango es $R_X = \{1, 2, 3, 4, 5, 6\}$, y la PMF está dada por:

$$P_X(x) = \frac{1}{6}, \quad \text{para cada } x \in R_X.$$

De modo que cada resultado tiene la misma probabilidad, dado que el dado es justo.

5.5 Función de Distribución Acumulada (CDF)

La función de masa de probabilidad (PMF) describe cómo se distribuyen los valores de una variable aleatoria discreta, asignando una probabilidad específica a cada valor posible. Sin embargo, en muchos casos necesitamos una representación que acumule estas probabilidades para describir cómo se distribuyen en conjunto los valores de la variable aleatoria. Aquí es donde entra en juego la *función de distribución acumulada (CDF)*.

Definición 5.5 (Función de Distribución Acumulada) La *función de distribución acumulada (CDF)* de una variable aleatoria X se define como:

$$F_X(x) = \mathbb{P}(X \leq x), \quad \forall x \in \mathbb{R}.$$

La intuición detrás de la CDF radica en que, en lugar de asignar probabilidades a valores puntuales, acumula las probabilidades de que la variable aleatoria tome un valor menor o igual a cierto umbral. En otras palabras, la CDF mide cuán probable es que la variable esté "hasta" o "antes" de un valor específico en su dominio. Esto proporciona una descripción global de la distribución de probabilidades, mostrando toda la información relevante sobre cómo se comporta la variable aleatoria en el intervalo considerado.

Propiedades

- **Monotonía:** $F_X(x)$ es una función no decreciente, lo que significa que si $x_1 \leq x_2$, entonces $F_X(x_1) \leq F_X(x_2)$.

  Este comportamiento se debe a que $F_X(x)$ acumula probabilidades. Al aumentar x, se incorporan más valores del rango de X, haciendo que la probabilidad acumulada nunca disminuya.

- **Límites en el infinito:**

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

El límite superior asegura que la probabilidad acumulada abarque todo el rango de X, alcanzando el valor total de 1.

Por otro lado, el límite inferior indica que, para valores muy pequeños de x, no se ha acumulado probabilidad significativa, ya que todavía no hemos alcanzado ningún valor probable del rango.

- **Relación con la PMF:** Para una variable aleatoria discreta, la probabilidad de que X tome un valor específico x puede determinarse a partir de la CDF mediante:
$$P_X(x) = F_X(x) - F_X(x^-),$$
donde $F_X(x^-)$ representa el límite de la CDF por la izquierda, definido como:
$$F_X(x^-) = \lim_{t \to x^-} F_X(t).$$
Para valores discontinuos de la CDF (característica de variables discretas), la diferencia entre $F_X(x)$ y $F_X(x^-)$ corresponde exactamente a $P_X(x)$.

Ejemplo:

Consideremos una variable aleatoria X que representa el número de caras obtenidas al lanzar dos monedas. El espacio muestral es:

$$\Omega = \{(C, C), (C, S), (S, C), (S, S)\}.$$

El rango de X es $R_X = \{0, 1, 2\}$, con la PMF:

$$P_X(x) = \begin{cases} \frac{1}{4} & \text{si } x = 0, \\ \frac{1}{2} & \text{si } x = 1, \\ \frac{1}{4} & \text{si } x = 2. \end{cases}$$

La CDF de X se calcula acumulando las probabilidades:

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0, \\ 0 + \frac{1}{4} = \frac{1}{4} & \text{si } 0 \leq x < 1, \\ \frac{1}{4} + \frac{1}{2} = \frac{3}{4} & \text{si } 1 \leq x < 2, \\ \frac{3}{4} + \frac{1}{4} = 1 & \text{si } x \geq 2. \end{cases}$$

5.6 Esperanza

El valor esperado de una variable aleatoria X es una medida de tendencia central que describe el promedio ponderado de los valores posibles de X, donde cada valor está ponderado por su probabilidad de ocurrencia. Intuitivamente, la esperanza representa el "promedio" de los resultados de un experimento aleatorio si se repite un gran número de veces.

Definición 5.6 (Esperanza) La *esperanza* (o valor esperado) de una variable aleatoria discreta X, denotada por $\mathbb{E}[X]$ o μ, se define como:

$$\mathbb{E}[X] = \sum_{x \in R_X} x \cdot P_X(x).$$

La expresión de la esperanza surge al combinar los valores que puede tomar X con sus probabilidades asociadas. Cada valor $x \in R_X$ contribuye al promedio en proporción a la probabilidad de que ocurra, $P_X(x)$. Este concepto es útil en una variedad de contextos, como evaluar el rendimiento esperado de una inversión, calcular el tiempo promedio de espera en una cola, o estimar el pago esperado en un juego de azar.

Ejemplo:

Si lanzamos un dado justo, los valores posibles de X son $R_X = \{1, 2, 3, 4, 5, 6\}$ con probabilidades iguales $P_X(x) = \frac{1}{6}$. El valor esperado sería:

$$\mathbb{E}[X] = \sum_{x=1}^{6} x \cdot \frac{1}{6} = \frac{1+2+3+4+5+6}{6} = 3{,}5.$$

Esto indica que, aunque ningún lanzamiento puede dar 3,5, este es el promedio que esperaríamos si lanzamos el dado muchas veces.

Propiedades

- **Linealidad:** La esperanza de una combinación lineal de variables es igual a la combinación lineal de sus esperanzas:

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y], \quad \text{para constantes } a, b \in \mathbb{R}.$$

- **Independencia:** Si X y Y son independientes, entonces la esperanza del producto es el producto de las esperanzas:

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

- **Transformaciones:** Para una función $g(X)$, la esperanza de $g(X)$ se calcula como:

$$\mathbb{E}[g(X)] = \sum_{x \in R_X} g(x) \cdot P_X(x).$$

  Esto permite calcular la esperanza de transformaciones de X, como X^2 o e^X.

Observación.

- La esperanza no siempre coincide con un valor que X pueda tomar.

  Esto es común, por ejemplo, en variables discretas como el lanzamiento de un dado, donde el promedio esperado 3,5 no es un resultado posible, pero describe el comportamiento agregado de múltiples repeticiones del experimento.

- Podría darse el caso que $\mathbb{E}[X] = +\infty$.

  Por ejemplo sea X una variable aleatoria discreta tal que, $X = 2^n$ con probabilidad $P_X(2^n) = 2^{-n}$, donde n es un entero positivo ($n \geq 1$). Notar que:

$$\sum_{n=1}^{\infty} P_X(2^n) = \sum_{n=1}^{\infty} 2^{-n} = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1.$$

Por lo tanto, $P_X(x)$ es una función de probabilidad, ahora, calculemos el valor esperado de X:

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} 2^n \cdot P_X(2^n) = \sum_{n=1}^{\infty} 2^n \cdot 2^{-n} = \sum_{n=1}^{\infty} 1.$$

Dado que esta serie diverge (suma infinitos unos), concluimos que:

$$\mathbb{E}[X] = \infty.$$

5.7 Varianza

La varianza de una variable aleatoria X mide cómo se dispersan los valores de X alrededor de su esperanza. Es decir, cuantifica qué tan alejados están, en promedio, los valores de X de su valor esperado $\mathbb{E}[X]$. Una varianza pequeña indica que los valores están concentrados cerca de la esperanza, mientras que una varianza grande refleja mayor dispersión.

Definición 5.7 (Varianza) La *varianza* de una variable aleatoria X, denotada por σ_X^2 o $\mathrm{Var}(X)$, está definida como:

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

La varianza calcula el promedio del cuadrado de las desviaciones de los valores de X respecto a su esperanza. De esta forma, se evita que las desviaciones positivas y negativas se cancelen, ya que eleva al cuadrado las diferencias antes de promediarlas. Así, $\mathrm{Var}(X)$ refleja la magnitud general de las desviaciones, independientemente de su signo.

Proposición 5.1

La varianza puede ser expresada más convenientemente como:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Demostración: La varianza de una variable aleatoria X se define como:

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Expandemos el cuadrado en el interior de la esperanza:

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

Ahora, aplicando la linealidad de la esperanza, descomponemos la esperanza término a término:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - 2\mathbb{E}[X \cdot \mathbb{E}[X]] + \mathbb{E}[\mathbb{E}[X]^2].$$

Observemos que $\mathbb{E}[X \cdot \mathbb{E}[X]] = \mathbb{E}[X] \cdot \mathbb{E}[X]$, ya que $\mathbb{E}[X]$ es una constante. Asimismo, $\mathbb{E}[(\mathbb{E}[X])^2] = (\mathbb{E}[X])^2$ por la misma razón. Por lo tanto, la expresión se simplifica a:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2.$$

Simplificando los términos, obtenemos la fórmula estándar de la varianza:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Propiedades

- **Escalamiento lineal:** Si X se escala linealmente como $aX + b$, entonces:

  $$\mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X).$$

  El término b no afecta la varianza, ya que desplazar X no altera la dispersión, mientras que a amplifica o reduce la varianza según su magnitud.

- **Independencia:** Si X y Y son independientes, la varianza de su suma es la suma de sus varianzas:

  $$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

  Esto muestra que, en variables independientes, las dispersiones se combinan de forma aditiva.

Ejemplo:

Sea X una variable aleatoria que representa el resultado del lanzamiento de un dado justo. Entonces:

$$R_X = \{1, 2, 3, 4, 5, 6\}, \quad P_X(x) = \frac{1}{6}.$$

La esperanza es:

$$\mathbb{E}[X] = \sum_{x=1}^{6} x \cdot \frac{1}{6} = 3{,}5.$$

Luego, se tiene que:

$$\mathbb{E}[X^2] = \sum_{x=1}^{6} x^2 \cdot \frac{1}{6} = \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2}{6} = \frac{91}{6}.$$

Finalmente, la varianza se calcula como:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{91}{6} - (3{,}5)^2 = \frac{35}{12}.$$

5.8 Desviación Estándar

La desviación estándar es una medida de dispersión que indica, en promedio, qué tan lejos están los valores de una variable aleatoria respecto a su esperanza. A diferencia de la varianza, la desviación estándar se encuentra en las mismas unidades que la variable aleatoria, lo que facilita su interpretación en un contexto práctico.

Definición 5.8 (Desviación Estándar) La *desviación estándar*, denotada por σ_X, se define como:

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}.$$

Como se mencionaba, mientras que la varianza mide la dispersión como un promedio de las desviaciones cuadráticas, la desviación estándar nos devuelve una medida en la escala original de los datos, siendo más directa de interpretar. Por ejemplo, si una variable X representa la altura en centímetros, su desviación estándar también estará en centímetros.

Observación. La desviación estándar no es aditiva como la varianza. Para variables independientes X y Y:

$$\mathrm{SD}(X+Y) \neq \mathrm{SD}(X) + \mathrm{SD}(Y).$$

Ejemplo:

Si $\mathrm{Var}(X) = 16$, entonces:

$$\mathrm{SD}(X) = \sqrt{16} = 4.$$

5.9 Otras Medidas de Tendencia Central

5.9.1 Coeficiente de Variación

El coeficiente de variación (CV) es una medida relativa de la dispersión, expresada como la proporción entre la desviación estándar y la esperanza de una variable aleatoria. Se utiliza para comparar la variabilidad de diferentes variables, incluso cuando tienen magnitudes o escalas diferentes.

Definición 5.9 (Coeficiente de Variación) El coeficiente de variación se define como:

$$\mathrm{CV}(X) = \frac{\mathrm{SD}(X)}{\mathbb{E}[X]}.$$

El CV mide la dispersión relativa de los valores en función de su media. Por ejemplo, Si $\mathrm{CV}(X) = 0{,}1$, la dispersión es del 10 % de la media. Mientras que si $\mathrm{CV}(X) = 1$, la dispersión es igual a la media.

El CV es particularmente útil cuando se comparan variables en diferentes escalas o unidades, ya que elimina el efecto de la magnitud absoluta.

Ejemplo:

Si $\mathrm{SD}(X) = 4$ y $\mathbb{E}[X] = 20$, entonces:

$$\mathrm{CV}(X) = \frac{4}{20} = 0{,}2 = 20\,\%.$$

5.10 Distribuciones Discretas Conocidas

Las distribuciones discretas conocidas son modelos probabilísticos que nos permiten describir y analizar fenómenos aleatorios en diferentes contextos. Cada una de estas distribuciones surge de situaciones específicas, como experimentos con resultados dicotómicos, conteos de eventos, o selección de elementos de una población.

5.10.1 Distribución Bernoulli

Definición 5.10 (Distribución Bernoulli) La *distribución Bernoulli* modela un experimento dicotómico, es decir, un experimento con dos posibles resultados numéricos: 1 para éxito (con probabilidad p) y 0 para fracaso (con probabilidad $1-p$).

- **Función de Masa de Probabilidad (PMF):**

$$P_X(x) = \begin{cases} p & \text{si } x = 1, \\ 1-p & \text{si } x = 0, \end{cases}$$

  donde $x \in \{0, 1\}$ y $0 \leq p \leq 1$.

- **Función de Distribución Acumulada (CDF):**

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

La distribución Bernoulli es útil para modelar situaciones donde solo hay dos posibles resultados. Algunos ejemplos incluyen:

- Lanzar una moneda justa ($p = 0{,}5$).
- Determinar si un cliente realiza una compra (p es la probabilidad de éxito, es decir, de que compre).
- Evaluar si una máquina en una línea de producción está operativa (p es la probabilidad de que funcione correctamente).

Proposición 5.2

Sea X una variable aleatoria con distribución Bernoulli de parámetro p. Entonces, la esperanza y la varianza de X están dadas por:

$$\mathbb{E}[X] = p, \quad \text{Var}(X) = p(1-p).$$

Demostración:

- **Esperanza:** Por definición, si $X \sim \text{Bernoulli}(p)$, entonces X toma el valor 1 con probabilidad p y el valor 0 con probabilidad $1-p$. Así tenemos que:

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

- **Varianza:** La varianza de X se define como:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Primero, calculamos $\mathbb{E}[X^2]$. Dado que $X^2 = X$ para una variable de tipo Bernoulli, se tiene que:

$$\mathbb{E}[X^2] = \mathbb{E}[X] = p.$$

Ahora, sustituimos en la fórmula de la varianza:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1-p).$$

Ejemplo:

Supongamos que X representa el resultado de un lanzamiento de moneda, con $p = 0{,}7$ para cara. Entonces:

$$P_X(1) = 0{,}7, \quad P_X(0) = 0{,}3.$$

La esperanza es:

$$\mathbb{E}[X] = 0{,}7,$$

y la varianza es:

$$\text{Var}(X) = 0{,}7 \cdot 0{,}3 = 0{,}21.$$

5.10.2 Distribución Geométrica

Definición 5.11 (Distribución Geométrica) La *distribución geométrica* modela el número de ensayos necesarios hasta el primer éxito en una secuencia de ensayos independientes, donde cada ensayo sigue una distribución Bernoulli con probabilidad de éxito p.

- **Función de Masa de Probabilidad (PMF):**

$$P_X(k) = (1-p)^{k-1}p, \quad k = 1, 2, 3, \ldots,$$

donde $0 \leq p \leq 1$.

La distribución geométrica es útil para modelar situaciones en las que se desea contar el número de intentos necesarios para alcanzar un éxito. Ejemplos comunes incluyen:

- Lanzar una moneda hasta que aparezca çara".
- Contar cuántos clientes visitan una tienda hasta que uno realice una compra.
- Determinar el número de intentos necesarios para que un proceso de producción genere un producto sin defectos.

Propiedades

- **Perdida de Memoria:** La distribución geométrica es una distribución con *perdida de memoria*, es decir:

$$\mathbb{P}(X > k + m \mid X > k) = \mathbb{P}(X > m).$$

Esto significa que, dado que aún no hemos tenido éxito después de k intentos, el proceso comienza de nuevo como si fuera el primer intento.

Proposición 5.3

Sea $X \sim \text{Geom}(p)$ una variable aleatoria con distribución geométrica. Entonces, la CDF, la esperanza y la varianza de X están dadas por:

$$F_X(k) = 1 - (1-p)^k, \quad \mathbb{E}[X] = \frac{1}{p}, \quad \text{Var}(X) = \frac{1-p}{p^2}.$$

donde $k = 1, 2, 3, \ldots$.

Demostración:

- **CDF:** Por definición, la CDF acumula las probabilidades desde el valor más bajo posible ($X = 1$) hasta $X = k$:

$$F_X(k) = \sum_{i=1}^{k} \mathbb{P}(X = i).$$

Sustituyendo la función de masa de probabilidad (PMF) para la distribución geométrica:

$$P_X(i) = (1-p)^{i-1} \cdot p \implies F_x(x) = \sum_{i=1}^{k} (1-p)^{i-1} \cdot p = p \cdot \sum_{i=1}^{k} (1-p)^{i-1}.$$

La suma $\sum_{i=1}^{k} (1-p)^{i-1}$ es una suma geométrica, y recordando que $\sum_{i=0}^{k-1} r^i = \frac{1-r^k}{1-r}, \quad$ donde $r \neq 1.$, así que:

$$\sum_{i=1}^{k} (1-p)^{i-1} = \frac{1-(1-p)^k}{1-(1-p)} = \frac{1-(1-p)^k}{p}.$$

Sustituyendo en $F_X(k)$, obtenemos:

$$F_X(k) = p \cdot \frac{1-(1-p)^k}{p}.$$

Cancelando p en el numerador y denominador, la CDF para la distribución geométrica es:

$$F_X(k) = 1 - (1-p)^k, \quad k = 1, 2, 3, \ldots$$

- **Esperanza:** La esperanza de X se define como:

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k \cdot P_X(k) = \sum_{k=1}^{\infty} k(1-p)^{k-1} p.$$

Sacando p como factor común, y utilizando la serie geométrica de primer orden $\sum_{k=1}^{\infty} k r^{k-1} = \frac{1}{(1-r)^2}, \quad$ para $|r| < 1$. La esperanza, entonces es tal que:

$$\mathbb{E}[X] = p \cdot \frac{1}{(1-(1-p))^2} = p \cdot \frac{1}{p^2} = \frac{1}{p}.$$

- **Varianza:** Ya sabemos que $\mathbb{E}[X] = \frac{1}{p}$. Ahora, debemos calcular $\mathbb{E}[X^2]$, que se expresa como:

$$\mathbb{E}[X^2] = \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} p.$$

Sacando p como factor común, nos queda:

$$\mathbb{E}[X^2] = p \sum_{k=1}^{\infty} k^2 (1-p)^{k-1}.$$

Sabiendo que la suma de una serie geométrica de segundo orden es tal que $\sum_{k=1}^{\infty} k^2 r^{k-1} = \frac{1+r}{(1-r)^3}, \quad \text{para } |r| < 1.$, tenemos que:

$$\mathbb{E}[X^2] = p \cdot \frac{1 + (1-p)}{p^3} = \frac{2-p}{p^2}.$$

Finalmente, calculamos la varianza:

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ &= \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 \\ &= \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}. \end{aligned}$$

Ejemplo:

Supongamos que estamos lanzando una moneda con probabilidad de cara $p = 0{,}25$. La probabilidad de que el primer éxito (cara) ocurra en el cuarto lanzamiento ($k = 4$) es:

$$P_X(4) = (1 - 0{,}25)^{4-1} \cdot 0{,}25 = (0{,}75)^3 \cdot 0{,}25 \approx 0{,}1055.$$

La probabilidad acumulada de que el primer éxito ocurra en los primeros 4 intentos es:

$$F_X(4) = 1 - (1 - 0{,}25)^4 = 1 - (0{,}75)^4 \approx 0{,}6836.$$

5.10.3 Distribución Binomial

Definición 5.12 (Distribución Binomial) La *distribución binomial* modela el número de éxitos en una secuencia de n ensayos independientes, donde cada ensayo sigue una distribución Bernoulli con probabilidad de éxito p.

- **Función de Masa de Probabilidad (PMF):**

$$P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n,$$

donde $\binom{n}{k}$ es el coeficiente binomial y $0 \leq p \leq 1$.

- **Función de Distribución Acumulada (CDF):**

$$F_X(k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}.$$

La distribución binomial es útil para modelar situaciones donde se realizan múltiples experimentos independientes bajo las mismas condiciones. Ejemplos incluyen:

- Contar cuántas veces aparece çara.ªl lanzar una moneda n veces.
- Determinar cuántos productos pasan una prueba de calidad en una muestra de n productos.
- Evaluar cuántos clientes realizan una compra en una muestra de n visitas a una tienda.

Proposición 5.4

Sea $X \sim \text{Binomial}(n, p)$ una variable aleatoria con distribución binomial. Entonces, la esperanza y la varianza de X están dadas por:

$$\mathbb{E}[X] = np, \quad \text{Var}(X) = np(1-p).$$

Demostración:

- **Esperanza:** Sea $X = \sum_{i=1}^{n} X_i$, donde cada $X_i \sim \text{Bernoulli}(p)$ y los X_i son independientes. Por linealidad de la esperanza, se tiene:

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} p = np.$$

- **Varianza:** Como los X_i son independientes, la varianza de la suma es la suma de las varianzas individuales:

$$\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(X_i).$$

  Dado que $\text{Var}(X_i) = p(1-p)$, se obtiene:

$$\text{Var}(X) = \sum_{i=1}^{n} p(1-p) = np(1-p).$$

Ejemplo:

Supongamos que lanzamos una moneda justa ($p = 0{,}5$) 5 veces y definimos X como el número de caras. La PMF es:

$$P_X(k) = \binom{5}{k} (0{,}5)^k (0{,}5)^{5-k}, \quad k = 0, 1, 2, 3, 4, 5.$$

Algunos valores específicos son:

$$P_X(2) = \binom{5}{2} (0{,}5)^2 (0{,}5)^3 = 10 \cdot 0{,}03125 = 0{,}3125.$$

La esperanza es:

$$\mathbb{E}[X] = 5 \cdot 0{,}5 = 2{,}5,$$

y la varianza es:

$$\text{Var}(X) = 5 \cdot 0{,}5 \cdot 0{,}5 = 1{,}25.$$

5.10.4 Distribución Binomial Negativa (o de Pascal)

Definición 5.13 (Distribución Binomial Negativa) La *distribución binomial negativa* (también conocida como distribución de Pascal) modela el número de ensayos necesarios para obtener m éxitos en un experimento donde cada ensayo es independiente y tiene una probabilidad de éxito p.

- **Función de Masa de Probabilidad (PMF):**

$$P_X(k) = \binom{k-1}{m-1} p^m (1-p)^{k-m}, \quad k = m, m+1, m+2, \ldots,$$

  donde $\binom{k-1}{m-1}$ es el coeficiente binomial, k es el número de ensayos realizados y $0 \leq p \leq 1$.

- **Función de Distribución Acumulada (CDF):**

$$F_X(k) = \sum_{i=m}^{k} \binom{i-1}{m-1} p^m (1-p)^{i-m}.$$

La distribución binomial negativa es útil para modelar situaciones donde se repiten experimentos independientes hasta que se alcanzan un número fijo de éxitos. Algunos ejemplos incluyen:

- Contar el número de lanzamientos de una moneda hasta obtener un número fijo de caras.
- Evaluar cuántos clientes deben visitar una tienda antes de que r de ellos realicen una compra.
- Determinar cuántos intentos de reparación se necesitan hasta lograr un número determinado de éxitos.

Proposición 5.5

Sea $X \sim$ Binomial Negativa(m, p) una variable aleatoria con distribución binomial negativa. Entonces, la esperanza y la varianza de X están dadas por:

$$\mathbb{E}[X] = \frac{m}{p}, \quad \mathrm{Var}(X) = \frac{m(1-p)}{p^2}.$$

Demostración:

- **Esperanza:** Sea X una variable aleatoria que modela el número de ensayos necesarios para obtener m éxitos en una secuencia de ensayos independientes, donde cada ensayo tiene una probabilidad de éxito p.

  Podemos escribir X como la suma de m variables aleatorias independientes $Y_1, Y_2, \ldots, Y_m$, donde cada Y_i representa el número de ensayos necesarios para

obtener un éxito, después de haber obtenido los éxitos anteriores. Cada Y_i sigue una distribución geométrica con parámetro p. Por lo tanto, se tiene:

$$X = Y_1 + Y_2 + \cdots + Y_m.$$

La esperanza de una variable aleatoria con distribución geométrica $Y_i \sim \text{Geom}(p)$ es $\mathbb{E}[Y_i] = \frac{1}{p}$. Utilizando la linealidad de la esperanza:

$$\mathbb{E}[X] = \mathbb{E}\left(\sum_{i=1}^{m} Y_i\right) = \sum_{i=1}^{m} \mathbb{E}[Y_i] = \sum_{i=1}^{m} \frac{1}{p} = m \cdot \frac{1}{p}.$$

Por lo tanto, la esperanza de X es:

$$\mathbb{E}[X] = \frac{m}{p}.$$

- **Varianza:** Dado que X es la suma de m variables aleatorias independientes $Y_1, Y_2, \ldots, Y_m$, su varianza es la suma de las varianzas individuales. Cada Y_i sigue una distribución geométrica con parámetro p, cuya varianza es:

$$\text{Var}(Y_i) = \frac{1-p}{p^2}.$$

Aplicando la propiedad de aditividad de la varianza para variables independientes:

$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^{m} Y_i\right) = \sum_{i=1}^{m} \text{Var}(Y_i) = \sum_{i=1}^{m} \frac{1-p}{p^2} = m \cdot \frac{1-p}{p^2}.$$

Por lo tanto, la varianza de X es:

$$\text{Var}(X) = \frac{m(1-p)}{p^2}.$$

Ejemplo:

Supongamos que lanzamos una moneda con probabilidad de cara $p = 0{,}5$. Queremos calcular la probabilidad de que el quinto éxito ocurra en el séptimo lanzamiento. Aquí $m = 5$ y $k = 7$. Usamos la PMF:

$$P_X(7) = \binom{6}{4}(0{,}5)^5(0{,}5)^2 = \binom{6}{4}(0{,}5)^7.$$

5.10.5 Distribución Hipergeométrica

Definición 5.14 (Distribución Hipergeométrica) La *distribución hipergeométrica* modela el número de elementos favorables en una muestra de tamaño n extraída sin reemplazo de una población finita de N elementos, donde K de ellos poseen una característica específica.

- **Función de Masa de Probabilidad (PMF):**

$$P_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \quad k = \max(0, n+K-N), \ldots, \min(K, n),$$

  donde $\binom{N}{n}$ es el coeficiente binomial.

- **Función de Distribución Acumulada (CDF):**

$$F_X(k) = \sum_{i=0}^{k} \frac{\binom{K}{i}\binom{N-K}{n-i}}{\binom{N}{n}}.$$

La distribución hipergeométrica es útil en situaciones donde se selecciona una muestra sin reemplazo de una población finita, lo que introduce una dependencia entre los resultados. Ejemplos comunes incluyen:

- Seleccionar una muestra de productos de un lote para determinar cuántos son defectuosos.
- Extraer cartas de una baraja sin reposición y contar cuántas son de un palo específico.
- Seleccionar estudiantes de una clase donde ciertos alumnos cumplen un criterio específico (por ejemplo, aprobar un examen).

Proposición 5.6

Sea $X \sim \text{Hipergeométrica}(N, K, n)$ una variable aleatoria con distribución hipergeométrica. Entonces, la esperanza y la varianza de X están dadas por:

$$\mathbb{E}[X] = n \cdot \frac{K}{N}, \quad \text{Var}(X) = n \cdot \frac{K}{N} \cdot \left(1 - \frac{K}{N}\right) \cdot \frac{N-n}{N-1}.$$

Demostración:

- **Esperanza:** Definamos variables indicadoras X_i para cada elemento en la muestra, donde $X_i = 1$ si el i-ésimo elemento cumple la condición favorable y 0 en caso contrario. La variable X se expresa como:

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

La esperanza de X es:

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

La probabilidad de que cualquier elemento de la población cumpla la condición es $\frac{K}{N}$. Por lo tanto, $\mathbb{E}[X_i] = \frac{K}{N}$, y se tiene:

$$\mathbb{E}[X] = \sum_{i=1}^{n} \frac{K}{N} = n \cdot \frac{K}{N}.$$

- **Varianza:** Dado que el concepto de *covarianza* aún no ha sido presentado, la demostración de esta expresión será realizada posteriormente.

Ejemplo:

Supongamos que en un aula hay 30 estudiantes ($N = 30$), de los cuales 12 son mujeres ($K = 12$). Si seleccionamos al azar a 5 estudiantes ($n = 5$), la distribución hipergeométrica describe la probabilidad de que k de ellos sean mujeres:

$$P_X(k) = \frac{\binom{12}{k}\binom{18}{5-k}}{\binom{30}{5}}, \quad k = 0, 1, \ldots, 5.$$

5.10.6 Distribución Poisson

Definición 5.15 (Distribución Poisson) La *distribución Poisson* modela la cantidad de eventos que ocurren en un intervalo fijo de tiempo o espacio, bajo las siguientes condiciones:

- Los eventos ocurren de forma independiente entre sí.
- La tasa media de ocurrencia es constante, igual a $\lambda > 0$.
- No ocurren dos eventos al mismo tiempo.
- **Función de Masa de Probabilidad (PMF):**

$$P_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

- **Función de Distribución Acumulada (CDF):**

$$F_X(k) = \sum_{i=0}^{k} \frac{e^{-\lambda}\lambda^i}{i!}.$$

La distribución Poisson es útil para modelar situaciones donde se cuentan eventos en un periodo o área, como:

- El número de llamadas recibidas por un centro de atención en una hora.
- La cantidad de clientes que llegan a una tienda en un día.

Proposición 5.7

Sea $X \sim \text{Poisson}(\lambda)$ una variable aleatoria con distribución Poisson. Entonces, la esperanza y la varianza de X están dadas por:

$$\mathbb{E}[X] = \lambda, \quad \text{Var}(X) = \lambda.$$

- El número de errores en una página de texto.

Demostración:

- **Esperanza:** Por definición, la esperanza de X es:

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k \cdot P_X(k) = \sum_{k=0}^{\infty} k \cdot \frac{e^{-\lambda}\lambda^k}{k!}.$$

  Realizando un cambio de índice en la serie, se obtiene:

$$\mathbb{E}[X] = e^{-\lambda}\lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}.$$

  Ahora por definición del número e como serie exponencial, se tiene que $\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{\lambda}$. Así finalmente:

$$\mathbb{E}[X] = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

- **Varianza:** La varianza de X se define como:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

  Sabemos que $\mathbb{E}[X] = \lambda$. Ahora, procedemos a calcular $\mathbb{E}[X^2]$. Por definición, tenemos:

$$\mathbb{E}[X^2] = \sum_{k=0}^{\infty} k^2 \cdot \frac{e^{-\lambda}\lambda^k}{k!}.$$

  Para simplificar este cálculo, utilizamos la descomposición $k^2 = k(k-1) + k$, lo que nos permite dividir la expresión en dos términos:

$$\mathbb{E}[X^2] = e^{-\lambda} \left(\sum_{k=2}^{\infty} k(k-1) \frac{\lambda^k}{k!} + \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} \right).$$

  Utilizando el hecho de que el denominador son factoriales, y además sacando convenientemente factores de λ para cada sumatoria, se tiene que:

$$\mathbb{E}[X^2] = e^{-\lambda} \left(\lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right).$$

  Notamos que al hacer un cambio de indices obtenemos la definición del número e como serie exponencial:

$$\mathbb{E}[X^2] = e^{-\lambda} \left(\lambda^2 e + \lambda e \right) = \lambda^2 + \lambda.$$

  Finalmente, calculamos la varianza:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda.$$

Ejemplo:

Supongamos que el número promedio de clientes que llegan a un restaurante es $\lambda = 5$ por hora. La probabilidad de que lleguen exactamente 3 clientes en una hora es:

$$P_X(3) = \frac{e^{-5} \cdot 5^3}{3!} = \frac{e^{-5} \cdot 125}{6}.$$

Proposición 5.8

La suma de dos variables Poisson $X_1 \sim \text{Poisson}(\lambda_1)$ y $X_2 \sim \text{Poisson}(\lambda_2)$ es también una variable Poisson:

$$X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2).$$

Demostración: Sean X_1 y X_2 dos variables aleatorias independientes con distribuciones Poisson de parámetros λ_1 y λ_2, respectivamente. Queremos demostrar que $X = X_1 + X_2$ sigue una distribución Poisson con parámetro $\lambda = \lambda_1 + \lambda_2$. Para ello, calcularemos la función de masa de probabilidad (PMF) de X.

La probabilidad de que $X = k$ se puede obtener sumando sobre todos los posibles pares (k_1, k_2) tales que $k_1 + k_2 = k$:

$$P(X = k) = \sum_{k_1=0}^{k} P(X_1 = k_1, X_2 = k - k_1).$$

Debido a la independencia de X_1 y X_2, esta probabilidad se puede descomponer en el producto de las probabilidades individuales:

$$P(X = k) = \sum_{k_1=0}^{k} P(X_1 = k_1) \cdot P(X_2 = k - k_1).$$

Sabemos que las PMFs de X_1 y X_2 son:

$$P(X_1 = k_1) = \frac{e^{-\lambda_1}\lambda_1^{k_1}}{k_1!}, \quad P(X_2 = k - k_1) = \frac{e^{-\lambda_2}\lambda_2^{k-k_1}}{(k-k_1)!}.$$

Sustituyendo estas expresiones en la suma, obtenemos:

$$P(X = k) = e^{-(\lambda_1+\lambda_2)} \sum_{k_1=0}^{k} \frac{\lambda_1^{k_1}}{k_1!} \cdot \frac{\lambda_2^{k-k_1}}{(k-k_1)!}.$$

El término dentro de la suma es equivalente a una expansión del binomio $(\lambda_1 + \lambda_2)^k$. Aplicando el teorema multinomial, obtenemos:

$$\sum_{k_1=0}^{k} \frac{\lambda_1^{k_1}\lambda_2^{k-k_1}}{k_1!(k-k_1)!} = \frac{(\lambda_1 + \lambda_2)^k}{k!}.$$

Sustituyendo este resultado, se tiene:

$$P(X = k) = \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1 + \lambda_2)^k}{k!}.$$

La PMF obtenida corresponde a la de una variable Poisson con parámetro $\lambda = \lambda_1 + \lambda_2$. Por lo tanto:

$$X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2).$$

Teorema Límite de Poisson

El *Teorema Límite de Poisson* establece que, bajo ciertas condiciones, la distribución binomial se aproxima a la distribución Poisson. Esto ocurre cuando el número de ensayos n es grande, la probabilidad de éxito p es pequeña, y el producto $np = \lambda$ se mantiene constante. Esta propiedad es valiosa en contextos donde modelamos eventos raros, como defectos en una línea de producción, llamadas a un centro de atención o fallas en un sistema técnico.

La aproximación es útil en la teoría de *decisiones bajo incertidumbre*, curso que se aborda en el siguiente semestre, y en modelos estocásticos, como el proceso de Poisson, donde se estudian la ocurrencia aleatoria de eventos en intervalos de tiempo o espacio.

Teorema 5.1 Teorema Límite de Poisson

Sea $X \sim \text{Bin}(n, p)$, una variable aleatoria binomial con n ensayos y probabilidad de éxito $p = \frac{\lambda}{n}$, donde $\lambda > 0$ es un valor fijo. Entonces, para cualquier $k \in \{0, 1, 2, \dots\}$, se cumple que:

$$\lim_{n\to\infty} P_X(k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \to \frac{e^{-\lambda}\lambda^k}{k!}.$$

Demostración: Sea la función de masa de probabilidad de la distribución binomial:

$$P_X(k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n!}{(n-k)!\,k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

Queremos calcular el límite de esta expresión cuando $n \to \infty$, manteniendo $\lambda = np$ constante. Comenzamos factorizando y reagrupando los terminos:

$$\begin{aligned}\lim_{n\to\infty} P_X(k) &= \lim_{n\to\infty} \frac{n!}{(n-k)!\,k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \lim_{n\to\infty} \frac{n!}{(n-k)!\,n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k}.\end{aligned}$$

Analizando cada uno de los factores se tiene que:

$$\frac{n!}{(n-k)!\,n^k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} = 1(1 - \frac{1}{n})(1 - \frac{2}{n}) \dots (1 - \frac{k-1}{n}).$$

Cuando $n \to \infty$ y k es fijo, este cociente tiende a 1. Finalmente, analizamos el factor $\left(1 - \frac{\lambda}{n}\right)^{n-k}$.

$$\left(1 - \frac{\lambda}{n}\right)^{n-k} = \left(1 - \frac{\lambda}{n}\right)^{n} \left(1 - \frac{\lambda}{n}\right)^{-k}.$$

Ahora por definición, sabemos que cuando $n \to \infty$ por definición $\left(1 - \frac{\lambda}{n}\right)^n = e^{\lambda}$, además, dado que k es fijo, el término adicional $\left(1 - \frac{\lambda}{n}\right)^{-k}$ tiende a 1. Sustituyendo todos estos resultados en la expresión original, obtenemos:

$$\lim_{n\to\infty} P_X(k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}.$$

Capítulo 6

Variables Aleatorias Continuas

En el capítulo anterior, estudiamos variables aleatorias discretas, que toman valores en un conjunto numerable, asignando a cada punto del conjunto una probabilidad positiva. Sin embargo, cuando se trabaja con variables aleatorias que pueden asumir valores en un conjunto no numerable, como los números reales en un intervalo, la situación cambia de manera fundamental. Este es el caso de las variables aleatorias continuas, las cuales requieren un enfoque distinto para definir y obtener sus probabilidades.

A lo largo de este capítulo, exploraremos las propiedades de estas variables, así como conceptos tales como la función de densidad de probabilidad (PDF), el cálculo de momentos y algunas de las distribuciones continuas más conocidas. Estos conceptos son esenciales en la modelización de fenómenos donde los resultados no se limitan a valores discretos, como en mediciones físicas continuas (peso, tiempo, temperatura) o en procesos estocásticos.

6.1 Variable Aleatoria Continua

Las variables aleatorias continuas se caracterizan por su capacidad de tomar cualquier valor dentro de un intervalo real. La probabilidad de que una variable aleatoria continua X tome un valor exacto es siempre cero:

$$\mathbb{P}(X = x) = 0, \quad \forall x \in \mathbb{R}.$$

Este hecho surge porque, en un conjunto no numerable, si cada punto tuviera una probabilidad positiva, la suma total de probabilidades sería infinita, lo que violaría el axioma de normalización. Por lo tanto, el cálculo de probabilidades se basa en intervalos y no en valores puntuales. En este contexto, el concepto de función de distribución acumulada (CDF) juega un papel central.

Definición 6.1 (Variable Aleatoria Continua) Una *variable aleatoria continua* X es aquella cuya función de distribución acumulada (CDF) $F_X(x)$ es continua para todo $x \in \mathbb{R}$.

Ejemplo:

Supongamos que medimos el tiempo que transcurre entre dos llamadas consecutivas recibidas en una central telefónica. Definimos X como el tiempo, en minutos, entre ambas llamadas. Aquí, el rango de X es el conjunto de todos los números reales no negativos:

$$R_X = [0, \infty).$$

Este conjunto es no numerable, ya que entre cualquier par de valores existe una cantidad infinita de decimales posibles. Por lo tanto, X es una variable aleatoria

continua. La probabilidad de que X tome un valor exacto, como $X = 3$ minutos, es igual a cero. En cambio, para calcular probabilidades, debemos considerar intervalos, por ejemplo, la probabilidad de que el tiempo entre llamadas sea entre 2 y 4 minutos:

$$\mathbb{P}(2 \leq X \leq 4).$$

6.2 Función de Densidad de Probabilidad (PDF)

Dado que no es posible asignar probabilidades a valores puntuales, ya que $\mathbb{P}(X = x) = 0$ para todo $x \in \mathbb{R}$, en su lugar, se utiliza la *función de densidad de probabilidad* (PDF, por sus siglas en inglés), la cual describe cómo se distribuye la probabilidad a lo largo del rango de la variable. La PDF no asigna probabilidades directas, sino que permite calcular la probabilidad de que la variable aleatoria tome un valor dentro de un intervalo.

Definición 6.2 (Función de Densidad de Probabilidad (PDF)) Sea X una variable aleatoria continua con función de distribución acumulada absolutamente continua $F_X(x)$. La *función de densidad de probabilidad (PDF)* de X, denotada por $f_X(x)$, se define como:

$$f_X(x) = \frac{dF_X(x)}{dx} = F'_X(x), \quad \text{si } F_X(x) \text{ es diferenciable en } x.$$

Aunque $f_X(x)$ no representa directamente una probabilidad, proporciona información sobre qué regiones del rango de X son más o menos probables. La PDF también satisface ciertas propiedades que garantizan que se comporte como una medida válida de probabilidad en los intervalos del rango.

Rango

Si X es una variable aleatoria continua, su rango se define como el conjunto de números reales donde la función de densidad de probabilidad es mayor que cero:

$$R_X = \{x \mid f_X(x) > 0\}.$$

Propiedades

La función de densidad de probabilidad cumple las siguientes propiedades fundamentales:

- **No negatividad:**

  $$f_X(x) \geq 0, \quad \forall x \in \mathbb{R}.$$

  Esto refleja que la densidad de probabilidad no puede ser negativa.

- **Área total bajo la curva:**

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

  El área total bajo la curva de la función de densidad debe ser igual a 1, garantizando que se cubra todo el espacio muestral.

- **Probabilidad en un intervalo:** La probabilidad de que la variable aleatoria X tome un valor en un intervalo $[a, b]$ se calcula como:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\,dx.$$

  Esta propiedad nos muestra como calcular probabilidades en variables continuas, dado que $\mathbb{P}(X = x) = 0$ para cualquier valor puntual x.

- **Probabilidades estrictas y no estrictas:** Para cualquier $a, b \in \mathbb{R}$, se tiene que:

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

  Esta pr

- **Probabilidad en subconjuntos:** Para cualquier conjunto $A \subseteq \mathbb{R}$:

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

  Esta propiedad generaliza el cálculo de probabilidades a cualquier conjunto medible A, no limitado a intervalos específicos.

Ejemplo:

Supongamos que X es una variable aleatoria que representa el tiempo (en minutos) que un cliente espera en una fila. Si la PDF de X está dada por:

$$f_X(x) = \begin{cases} \frac{1}{10}, & \text{si } 0 \leq x \leq 10, \\ 0, & \text{en otro caso,} \end{cases}$$

entonces el rango de X es $R_X = [0, 10]$. Para calcular la probabilidad de que el cliente espere entre 2 y 5 minutos, se realiza la siguiente integral:

$$\mathbb{P}(2 \leq X \leq 5) = \int_2^5 f_X(x)\,dx = \int_2^5 \frac{1}{10}\,dx = \frac{1}{10} \cdot (5 - 2) = 0{,}3.$$

Así, la probabilidad de que el cliente espere entre 2 y 5 minutos es 0,3 o un 30 %.

6.3 Esperanza

En el caso de variables aleatorias continuas, la esperanza se calcula mediante la función de densidad de probabilidad (PDF).

Definición 6.3 (Esperanza) La *esperanza* (o valor esperado) de una variable aleatoria continua X, denotada por $\mathbb{E}[X]$ o μ, se define como:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)\,dx,$$

donde $f_X(x)$ es la función de densidad de probabilidad de X.

Esta expresión combina los valores que puede tomar X con sus densidades de probabilidad asociadas. El valor x contribuye al promedio en proporción a $f_X(x)$, por lo que el cálculo de la esperanza refleja la distribución continua de X.

La esperanza de una variable aleatoria continua tiene propiedades análogas a las del caso discreto, por ejemplo, para una función $g(X)$, la esperanza de $g(X)$ en el caso continuo se define mediante la integral de la función de densidad de probabilidad de X:

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\, dx.$$

Ejemplo:

Supongamos que una variable aleatoria X representa el tiempo de espera de un cliente en una fila, y su función de densidad está dada por:

$$f_X(x) = \begin{cases} \frac{1}{10}, & \text{si } 0 \leq x \leq 10, \\ 0, & \text{en otro caso.} \end{cases}$$

Para calcular el valor esperado de X, realizamos la siguiente integral:

$$\mathbb{E}[X] = \int_0^{10} x \cdot \frac{1}{10}\, dx = \frac{1}{10} \int_0^{10} x\, dx.$$

Resolviendo la integral, obtenemos:

$$\mathbb{E}[X] = \frac{1}{10} \cdot \left[\frac{x^2}{2}\right]_0^{10} = \frac{1}{10} \cdot \frac{100}{2} = 5.$$

Esto indica que el tiempo de espera promedio es de 5 minutos.

6.4 Varianza

La varianza de una variable aleatoria continua X se define de la misma manera que en el caso discreto.

Definición 6.4 (Varianza) La *varianza* de una variable aleatoria continua X, denotada por σ_X^2 o $\mathrm{Var}(X)$, se define como:

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f_X(x)\, dx,$$

donde $f_X(x)$ es la función de densidad de probabilidad de X.

Esta presenta las mismas propiedades a las del caso discreto y también puede expresarse de manera equivalente como $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Ejemplo:

Sea X una variable aleatoria continua con función de densidad:

$$f_X(x) = \begin{cases} \frac{1}{10}, & \text{si } 0 \leq x \leq 10, \\ 0, & \text{en otro caso.} \end{cases}$$

La esperanza de X ya fue calculada en el ejemplo anterior y es $\mathbb{E}[X] = 5$. Ahora, calculamos $\mathbb{E}[X^2]$:

$$\mathbb{E}[X^2] = \int_0^{10} x^2 \cdot \frac{1}{10}\, dx = \frac{1}{10} \int_0^{10} x^2\, dx.$$

Resolviendo la integral:

$$\mathbb{E}[X^2] = \frac{1}{10} \cdot \left[\frac{x^3}{3}\right]_0^{10} = \frac{1}{10} \cdot \frac{1000}{3} = \frac{100}{3}.$$

Finalmente, la varianza es:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{100}{3} - 5^2 = \frac{100}{3} - 25 = \frac{25}{3}.$$

Por lo tanto, la varianza de X es $\frac{25}{3}$.

6.5 Otras Medidas de Tendencia Central

6.5.1 Asimetría (Skewness)

La asimetría de una variable aleatoria X es una medida que describe la simetría o falta de ella en la distribución de X alrededor de su valor esperado. Si la distribución es perfectamente simétrica, como ocurre en ciertas distribuciones conocidas, la asimetría es igual a cero. Por el contrario, una asimetría positiva indica que la cola derecha de la distribución es más larga o está más dispersa que la izquierda, mientras que una asimetría negativa sugiere lo opuesto.

Definición 6.5 (Asimetría) Sea X una variable aleatoria con esperanza $\mathbb{E}[X] = \mu$ y varianza $\mathrm{Var}(X) = \sigma^2$. La *asimetría* de X, denotada por $\mathrm{Skew}(X)$, se define como:

$$\mathrm{Skew}(X) = \mathbb{E}\left[\left(\frac{X - \mu}{\sigma}\right)^3\right].$$

La definición utiliza la desviación estándar σ para estandarizar la variable, y eleva al cubo las desviaciones para reflejar la dirección del sesgo en la distribución. Si la asimetría es positiva, los valores mayores que la media tienen mayor influencia; si es negativa, los valores menores tienen mayor influencia.

Ejemplo:

Sea X una variable aleatoria con esperanza $\mathbb{E}[X] = 10$, desviación estándar $\sigma = 2$ y función de densidad sesgada hacia valores mayores. Su asimetría se calcula como:

$$\mathrm{Skew}(X) = \mathbb{E}\left[\left(\frac{X - 10}{2}\right)^3\right].$$

Un valor de $\mathrm{Skew}(X) = 1{,}2$ sugiere una distribución con una cola derecha más extensa.

6.5.2 Curtosis (Kurtosis)

La curtosis de una variable aleatoria X es una medida que describe la concentración de los valores en torno a la media y la presencia de colas más o menos pronunciadas en la distribución. Una distribución con curtosis similar a la normal se denomina *mesocúrtica*. Una curtosis alta (*leptocúrtica*) indica una distribución con una mayor concentración de valores cerca de la media y colas más largas, mientras que una curtosis baja (*platicúrtica*) refleja colas más cortas y una menor concentración en la media.

Definición 6.6 (Curtosis) Sea X una variable aleatoria con esperanza $\mathbb{E}[X] = \mu$ y varianza $\text{Var}(X) = \sigma^2$. La *curtosis* de X, denotada por $\text{Kurt}(X)$, se define como:

$$\text{Kurt}(X) = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right].$$

La definición utiliza la desviación estándar σ para estandarizar la variable. El exponente 4 acentúa las desviaciones grandes respecto a la media, haciendo que las colas largas tengan una mayor influencia en el valor de la curtosis.

Ejemplo:

Supongamos que X es una variable aleatoria con $\mathbb{E}[X] = 0$ y $\sigma = 1$, cuya distribución tiene colas largas en comparación con la normal. La curtosis se calcula como:

$$\text{Kurt}(X) = \mathbb{E}\left[\left(\frac{X}{1}\right)^4\right] = \mathbb{E}[X^4].$$

Un valor de $\text{Kurt}(X) = 5$ indicaría una distribución leptocúrtica con colas más pesadas de lo habitual.

6.5.3 Mediana

La mediana es una medida de posición central en la distribución de una variable aleatoria, dividiendo el rango en dos mitades de igual probabilidad.

Definición 6.7 (Mediana) La *mediana* de una variable aleatoria X es el valor m tal que la función de distribución acumulada satisface:

$$F_X(m) = \frac{1}{2}.$$

La mediana indica el punto donde el 50 % de la probabilidad acumulada está por debajo y el otro 50 % está por encima.

6.5.4 Moda

La moda es el valor o los valores donde la función de densidad de probabilidad alcanza su máximo, indicando dónde se concentra la mayor densidad de probabilidad.

Definición 6.8 (Moda) La *moda* de una variable aleatoria X es el valor x_{moda} que maximiza la función de densidad de probabilidad $f_X(x)$:

$$x_{\text{moda}} = \arg\max_x f_X(x).$$

En distribuciones unimodales hay una única moda, mientras que en distribuciones multimodales pueden existir varios máximos locales.

6.5.5 Cuantiles

Los cuantiles son puntos que dividen la distribución en intervalos de igual probabilidad, permitiendo analizar la posición relativa de los valores.

Definición 6.9 (Cuantiles) Un *q-cuantil* de una variable aleatoria X es el valor x_k que satisface:

$$F_X(x_k) = \frac{k}{q}, \quad \text{para } k = 1, 2, \ldots, q-1.$$

Dependiendo del valor de q, los cuantiles reciben nombres específicos:

- **Mediana**: Es el 2-cuantil.
- **Cuartiles**: Son los 4-cuantiles, que dividen la distribución en cuatro intervalos de igual probabilidad.
- **Quintiles**: Son los 5-cuantiles.
- **Percentiles**: Son los 100-cuantiles.

Ejemplo:

Supongamos que X es una variable aleatoria continua con función de distribución acumulada $F_X(x)$. Si queremos calcular el primer cuartil (x_1), buscamos el valor que cumpla:

$$F_X(x_1) = \frac{1}{4}.$$

Este valor indica que el 25 % de la probabilidad acumulada se encuentra por debajo de x_1.

6.6 Función de una Variable Aleatoria Continua

En muchas situaciones, es de interés analizar el comportamiento de una función de una variable aleatoria continua. Si X es una variable aleatoria continua y aplicamos una transformación $Y = g(X)$, surge la pregunta: ¿cómo se comporta la probabilidad asociada a la variable Y? Aunque ya conocemos cómo calcular el valor esperado de $g(X)$ mediante:

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx,$$

es fundamental determinar la función de densidad de probabilidad de Y, para así describir completamente su distribución.

Este problema es relevante en diversos contextos, como cambios de escala, funciones no lineales y otras transformaciones de datos. Por ejemplo, al transformar una variable aleatoria mediante una función exponencial o logarítmica, es crucial conocer cómo se distribuye la probabilidad a lo largo de los nuevos valores que toma la variable transformada.

En las siguientes secciones, definiremos formalmente la función de densidad de Y y desarrollaremos métodos para encontrarla a partir de la densidad de X, con especial atención al caso en que g es una función estrictamente monótona. [1]

6.6.1 Cambio de Variable

Cuando deseamos encontrar la función de distribución acumulada (CDF) de una transformación de una variable aleatoria continua, el proceso es relativamente sencillo. Si $Y = g(X)$ y g es estrictamente monótona, podemos expresar la CDF de Y en términos de la CDF de X, tal que, $\mathbb{P}(g(X) \leq y)$. Sin embargo, si en lugar de la CDF buscamos la función de densidad de probabilidad (PDF) de Y, el problema se vuelve más complejo, ya que la probabilidad debe ajustarse para reflejar cómo la función g afecta la distribución de los valores de X.

Teorema 6.1 Cambio de Variable

Si X es una variable aleatoria continua con PDF $f_X(x)$, y $Y = g(X)$ es una transformación de X, donde g es estrictamente monótona y diferenciable, la PDF de Y es:

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right|.$$

Demostración: Consideremos una variable aleatoria continua X con función de distribución acumulada $F_X(x)$ y PDF $f_X(x)$. Sea $Y = g(X)$ una transformación de X, donde g es una función estrictamente monótona y diferenciable. Queremos encontrar la función de densidad de probabilidad de Y, denotada por $f_Y(y)$. Partimos de la CDF de Y, definida como:

$$F_Y(y) = \mathbb{P}(Y \leq y).$$

Dado que $Y = g(X)$, esta expresión se puede reescribir como:

$$F_Y(y) = \mathbb{P}(g(X) \leq y).$$

Ahora, asumamos que g es una función estrictamente creciente. En este caso, la condición $g(X) \leq y$ es equivalente a $X \leq g^{-1}(y)$. Esto nos lleva a:

$$F_Y(y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

Si, en cambio, g fuera estrictamente decreciente, la condición $g(X) \leq y$ se traduciría en $X \geq g^{-1}(y)$, y la CDF se expresaría como:

$$F_Y(y) = 1 - F_X(g^{-1}(y)).$$

[1] Una función es *monótona* si es creciente o decreciente en todo su dominio. Una función $g(x)$ es creciente si $x_1 \leq x_2$ implica $g(x_1) \leq g(x_2)$, y es decreciente si $x_1 \leq x_2$ implica $g(x_1) \geq g(x_2)$. Si las desigualdades son estrictas ($<$ y $>$), se dice que la función es *estrictamente monótona*. Las funciones estrictamente monótonas tienen una inversa bien definida.

Esto ocurre porque la función g afecta el orden según su monotonía. Si g es creciente, para $x_1 < x_2$, se cumple que $g(x_1) < g(x_2)$, lo que preserva el orden y hace que $Y \leq y$ sea equivalente a $X \leq g^{-1}(y)$. Por tanto, $F_Y(y) = F_X(g^{-1}(y))$. Si g es decreciente, el orden se invierte ($g(x_1) > g(x_2)$), y $Y \leq y$ implica $X \geq g^{-1}(y)$, resultando en $F_Y(y) = 1 - F_X(g^{-1}(y))$.

Dado que buscamos la PDF de Y, procedemos a derivar la CDF $F_Y(y)$ con respecto a y. En el caso de g estrictamente creciente, tenemos:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(g^{-1}(y)).$$

Aplicando la regla de la cadena, obtenemos:

$$f_Y(y) = F'_X(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y).$$

Sabemos que $F'_X(x) = f_X(x)$, por lo que:

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y).$$

Aquí es donde la monotonía de g juega un papel crucial. Si g es estrictamente creciente, $\frac{d}{dy} g^{-1}(y) > 0$, y la expresión anterior es válida tal cual. Sin embargo, si g es estrictamente decreciente, entonces $\frac{d}{dy} g^{-1}(y) < 0$. Para asegurar que la densidad de probabilidad sea no negativa, tomamos el valor absoluto de la derivada:

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right|.$$

Esta expresión generaliza ambos casos de monotonía. El valor absoluto se asegura de que la probabilidad no se vea afectada por la dirección de la transformación, manteniendo la interpretación correcta de la densidad.

Generalización Cambio de Variable

¿Qué sucede si la función sobre la que estamos trabajando no es estrictamente monótona en todo su dominio? En este caso, podemos dividir el dominio en intervalos donde la función sea estrictamente monótona y diferenciable. De esta forma, en cada uno de estos intervalos se puede aplicar el resultado obtenido previamente para el cambio de variable.

La demostración de este teorema sigue una lógica similar al caso de una función monótona. Dividimos el dominio de g en intervalos donde es estrictamente monótona. En cada uno de estos intervalos, se aplica el resultado anterior, calculando la contribución a la densidad $f_Y(y)$ de las preimágenes $g_i^{-1}(y)$ que satisfacen $g_i(x) = y$. Posteriormente, se suman estas contribuciones. El valor absoluto en la fórmula asegura que la densidad permanezca no negativa, independientemente de si g es creciente o decreciente en cada intervalo.

Teorema 6.2 Generalización Cambio de Variable

Sea X una variable aleatoria continua y $g : \mathbb{R} \to \mathbb{R}$ una función definida en una partición de intervalos $\{I_i\}_{i=1}^{N}$, donde g es estrictamente monótona y diferenciable en el interior de cada uno de dichos intervalos. Entonces, si $Y = g(X)$, la PDF de Y está dada por:

$$f_Y(y) = \sum_{i=1}^{N} f_X(g_i^{-1}(y)) \cdot \left| \frac{d}{dy} g_i^{-1}(y) \right|,$$

donde $g_i(x) = g(x)$ para todo $x \in I_i$.

6.7 Distribuciones Continuas Conocidas

6.7.1 Distribución Uniforme

Definición 6.10 (Distribución Uniforme) La *distribución uniforme* continua modela situaciones donde todos los valores posibles de una variable aleatoria tienen la misma densidad de probabilidad en un intervalo dado $[a, b]$.

- **Función de Densidad de Probabilidad (PDF):**

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{si } a \leq x \leq b, \\ 0, & \text{en otro caso.} \end{cases}$$

Aquí, a y b son los límites del intervalo, con $a < b$.

La distribución uniforme es útil para modelar situaciones donde no existe preferencia alguna entre los valores dentro de un intervalo. Algunos ejemplos incluyen:

- El tiempo de llegada aleatorio de un cliente a una tienda, cuando puede llegar en cualquier momento dentro de un horario dado.
- La posición de un punto seleccionado al azar en una línea de longitud $b - a$.
- El resultado de una variable aleatoria que representa la selección aleatoria de un número entre dos límites conocidos.

Proposición 6.1

Sea $X \sim \text{Uniforme}(a, b)$ una variable aleatoria con distribución uniforme en el intervalo $[a, b]$. Entonces, la CDF, la esperanza y la varianza de X están dadas por:

$$F_X(x) = \begin{cases} 0, & \text{si } x < a, \\ \frac{x-a}{b-a}, & \text{si } a \leq x \leq b, \\ 1, & \text{si } x > b, \end{cases} \quad \mathbb{E}[X] = \frac{a+b}{2}, \quad \text{Var}(X) = \frac{(b-a)^2}{12}.$$

Demostración:

- **CDF:** Por definición, la función de distribución acumulada es:

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt.$$

La PDF de la distribución uniforme es:

$$f_X(t) = \begin{cases} \frac{1}{b-a}, & \text{si } a \leq t \leq b, \\ 0, & \text{en otro caso.} \end{cases}$$

Ahora, evaluemos la integral en los distintos casos. Si $x < a$, $f_X(t) = 0$ para todo $t \leq x$, lo que implica $F_X(x) = 0$.

Por otro lado, si $a \leq x \leq b$, la integral se evalúa dentro del intervalo $[a, x]$:

$$F_X(x) = \int_a^x \frac{1}{b-a}\, dt = \frac{1}{b-a}(x-a).$$

Si $x > b$, toda el área bajo la PDF ha sido acumulada, por lo que $F_X(x) = 1$. Por lo tanto, la CDF es:

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

- **Esperanza:** La esperanza de X se define como:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_a^b x \cdot \frac{1}{b-a}\, dx.$$

Evaluamos en la integral:

$$\mathbb{E}[X] = \frac{1}{b-a}\int_a^b x\, dx = \frac{1}{b-a}\frac{x^2}{2}\Bigg|_a^b.$$

Sustituyendo los valores de los límites:

$$\mathbb{E}[X] = \frac{1}{b-a}\left(\frac{b^2}{2} - \frac{a^2}{2}\right) = \frac{b+a}{2}.$$

- **Varianza:** La varianza se define como:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Calculamos $\mathbb{E}[X^2]$:

$$\mathbb{E}[X^2] = \int_a^b x^2 \cdot \frac{1}{b-a}\, dx = \frac{1}{b-a}\frac{x^3}{3}\Bigg|_a^b.$$

Sustituyendo los valores de los límites:

$$\mathbb{E}[X^2] = \frac{1}{b-a} \cdot \frac{b^3 - a^3}{3}.$$

La varianza es entonces:

$$\begin{aligned} \mathrm{Var}(X) &= \frac{1}{3} \cdot \frac{b^3 - a^3}{b-a} - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{(b-a)^2}{12}. \end{aligned}$$

Ejemplo:

Supongamos que se selecciona al azar un número entre 2 y 10, siguiendo una distribución uniforme. La PDF es:

$$f_X(x) = \begin{cases} \frac{1}{8}, & \text{si } 2 \leq x \leq 10, \\ 0, & \text{en otro caso.} \end{cases}$$

La probabilidad de que el número seleccionado esté entre 4 y 7 es:

$$\mathbb{P}(4 \leq X \leq 7) = \int_4^7 \frac{1}{8}\, dx = \frac{1}{8}(7-4) = \frac{3}{8} = 0{,}375.$$

6.7.2 Distribución Exponencial

Definición 6.11 (Distribución Exponencial) La *distribución exponencial* modela el tiempo entre eventos, donde los eventos ocurren de forma aleatoria pero con una tasa constante.

- **Función de Densidad de Probabilidad (PDF):**

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{si } x \geq 0, \\ 0, & \text{si } x < 0, \end{cases}$$

donde $\lambda > 0$ es el parámetro de tasa, que indica la frecuencia promedio de ocurrencia de eventos.

La distribución exponencial es útil en situaciones donde se modela la espera o el tiempo entre eventos aleatorios. Algunos ejemplos incluyen:

- El tiempo entre llamadas consecutivas a un centro de atención al cliente.
- La duración de vida de un dispositivo electrónico antes de fallar.
- El tiempo entre llegadas de clientes a un sistema de servicio.

Propiedades

- **Perdida de Memoria:** La distribución exponencial es una distribución con *perdida de memoria*, es decir:

$$\mathbb{P}(X > t + s \mid X > s) = \mathbb{P}(X > t).$$

Esto implica que el tiempo previo transcurrido (s) no afecta la probabilidad futura. Por ejemplo, si un dispositivo electrónico ha funcionado 100 horas, su probabilidad de durar 50 horas más es igual a la de uno nuevo durar 50 horas.

Proposición 6.2

Sea $X \sim \text{Exp}(\lambda)$ una variable aleatoria con distribución exponencial de parámetro λ. Entonces, la CDF, la esperanza y la varianza de X están dadas por:

$$F_X(x) = \begin{cases} 0, & \text{si } x < 0, \\ 1 - e^{-\lambda x}, & \text{si } x \geq 0, \end{cases} \quad \mathbb{E}[X] = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}.$$

Demostración:

- **CDF:** La CDF se define como:

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt.$$

Evaluemos la integral en los diferentes casos, si $x < 0$, $f_X(t) = 0$ para todo $t \leq x$, lo que implica que $F_X(x) = 0$. Si $x \geq 0$, la integral se evalúa dentro del intervalo $[0, x]$:

$$F_X(x) = \int_0^x \lambda e^{-\lambda t}\, dt = -e^{-\lambda t} \Big|_0^x = 1 - e^{-\lambda x}.$$

Por lo tanto, la CDF es:

$$F_X(x) = \begin{cases} 0, & \text{si } x < 0, \\ 1 - e^{-\lambda x}, & \text{si } x \geq 0. \end{cases}$$

- **Esperanza:** La esperanza de X se define como:

$$\mathbb{E}[X] = \int_0^\infty x\lambda e^{-\lambda x}\, dx.$$

Utilizando integración por partes. Sean $u = x$ y $dv = \lambda e^{-\lambda x}\, dx$. Entonces:

$$du = dx, \quad v = -e^{-\lambda x}.$$

Aplicando la expresión de integración por partes:

$$\begin{aligned} \mathbb{E}[X] &= -xe^{-\lambda x} \Big|_0^\infty + \int_0^\infty e^{-\lambda x}\, dx \\ &= 0 - \frac{e^{-\lambda x}}{\lambda} \Big|_0^\infty = \frac{1}{\lambda}. \end{aligned}$$

- **Varianza:** La varianza se define como:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Calculamos $\mathbb{E}[X^2]$ para $X \sim \text{Exponencial}(\lambda)$:

$$\mathbb{E}[X^2] = \int_0^\infty x^2 \lambda e^{-\lambda x}\, dx.$$

Utilizamos integración por partes, definimos:

$$u = x^2, \quad dv = \lambda e^{-\lambda x}\, dx.$$

Calculamos las derivadas e integrales:

$$du = 2x\, dx, \quad v = -e^{-\lambda x}.$$

Sustituimos en la expresión de integración por partes:

$$\begin{aligned}\mathbb{E}[X^2] &= -x^2 e^{-\lambda x}\big|_0^\infty + \int_0^\infty 2xe^{-\lambda x}\, dx \\ &= 0 + \int_0^\infty 2xe^{-\lambda x}\, dx.\end{aligned}$$

El primer término se anula, ya que $x^2 e^{-\lambda x} \to 0$ cuando $x \to \infty$, y es 0 cuando $x = 0$. Ahora, calculamos $\int_0^\infty 2xe^{-\lambda x}\, dx$ mediante nuevamente integración por partes, definimos:

$$u = x, \quad dv = 2e^{-\lambda x}\, dx.$$

Calculamos las derivadas e integrales:

$$du = dx, \quad v = -\frac{2}{\lambda} e^{-\lambda x}.$$

Sustituyendo nuevamente:

$$\begin{aligned}\int_0^\infty 2xe^{-\lambda x}\, dx &= -\frac{2}{\lambda} xe^{-\lambda x}\Big|_0^\infty + \int_0^\infty \frac{2}{\lambda} e^{-\lambda x}\, dx \\ &= 0 + \frac{2}{\lambda^2}.\end{aligned}$$

Por lo tanto, se tiene que $\mathbb{E}[X^2]$:

$$\mathbb{E}[X^2] = \frac{2}{\lambda^2}.$$

La varianza es entonces:

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ &= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.\end{aligned}$$

Ejemplo:

Supongamos que el tiempo entre llegadas de clientes a un banco sigue una distribución exponencial con $\lambda = 0{,}5$ llegadas por minuto. La probabilidad de que el tiempo de espera sea menor o igual a 2 minutos es:

$$\mathbb{P}(X \leq 2) = F_X(2) = 1 - e^{-0{,}5 \cdot 2} = 1 - e^{-1} \approx 0{,}632.$$

6.7.3 Distribución Normal

Definición 6.12 (Distribución Normal) La *distribución normal* es una de las distribuciones más importantes en probabilidad y estadística, modelando fenómenos naturales que involucran la suma de múltiples factores independientes. Una variable aleatoria X sigue una distribución normal con media μ y varianza σ^2, denotada por $X \sim \mathcal{N}(\mu, \sigma^2)$, si su PDF es:

- **Función de Densidad de Probabilidad (PDF):**

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Aquí, μ es la media y σ^2 la varianza.

La distribución normal es útil en una variedad de situaciones donde las variables aleatorias representan fenómenos acumulativos. Algunos ejemplos incluyen:

- Los errores de medición en experimentos.
- Los puntajes en exámenes estandarizados.
- Las fluctuaciones en los precios de activos financieros.

Proposición 6.3

Sea $X \sim \mathcal{N}(\mu, \sigma^2)$ una variable aleatoria con distribución normal. Entonces, la CDF no tiene una forma cerrada, pero la esperanza y la varianza están dadas por:

$$\mathbb{E}[X] = \mu, \quad \text{Var}(X) = \sigma^2.$$

Además, si $X \sim \mathcal{N}(0, 1)$ (normal estándar), la CDF se denota $\Phi(x)$ y es:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt.$$

Los valores de $\Phi(x)$ no se expresan en forma cerrada, pero se encuentran tabulados en tablas de distribución normal estándar o calculados mediante herramientas estadísticas.

Demostración:

- **Esperanza:** La esperanza de X se define como:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx.$$

Hacemos un cambio de variable: $z = \frac{x-\mu}{\sigma}$, lo que implica $dx = \sigma\, dz$. La integral se convierte en:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} (\sigma z + \mu) \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz.$$

El primer término, $\int_{-\infty}^{\infty} ze^{-\frac{z^2}{2}}\,dz$, es cero debido a la simetría de la función exponencial. Por lo tanto, queda:

$$\mathbb{E}[X] = \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz = \mu.$$

- **Varianza:** La varianza se define como:

$$\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2].$$

Sabemos que:

$$\mathbb{E}[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx.$$

Haciendo el mismo cambio de variable $z = \frac{x-\mu}{\sigma}$, obtenemos:

$$\mathrm{Var}(X) = \sigma^2 \int_{-\infty}^{\infty} z^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz.$$

La integral es conocida y se verá en la función gamma más adelante, donde se obtiene un valor de 1. Por lo tanto:

$$\mathrm{Var}(X) = \sigma^2.$$

Proposición 6.4 Transformación a Normal Estándar

Sea $X \sim \mathcal{N}(\mu, \sigma^2)$ una variable aleatoria con media μ y varianza σ^2. Si definimos la variable estandarizada:

$$Z = \frac{X-\mu}{\sigma},$$

entonces Z sigue una distribución normal estándar $\mathcal{N}(0,1)$.

Demostración: Definimos la variable $Z = \frac{X-\mu}{\sigma}$. Queremos encontrar la PDF de Z. Utilizamos el teorema de cambio de variable. La transformación inversa es:

$$X = \sigma Z + \mu.$$

La derivada de X respecto a Z es:

$$\frac{dX}{dZ} = \sigma.$$

La PDF de $X \sim \mathcal{N}(\mu, \sigma^2)$ es:

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

Por el teorema de cambio de variable sabemos que:

$$f_Z(z) = f_X(\sigma z + \mu) \cdot \left|\frac{dX}{dZ}\right|^{-1}.$$

Sustituyendo $X = \sigma z + \mu$ y $\frac{dX}{dZ} = \sigma$:

$$\begin{aligned} f_Z(z) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\sigma z+\mu-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sigma} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}. \end{aligned}$$

Por lo tanto, $Z \sim \mathcal{N}(0, 1)$.

Propiedades

- **Simetría:** La distribución normal estándar es simétrica respecto a cero, lo que implica:
$$\Phi(-x) = 1 - \Phi(x).$$
- **Propiedad de escalamiento:** Si $X \sim \mathcal{N}(\mu, \sigma^2)$, la probabilidad acumulada puede expresarse en términos de Φ al estandarizar X:
$$\mathbb{P}(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$
- **Transformaciones lineales:** Si $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ y aplicamos una transformación lineal $Y = aX + b$, donde $a, b \in \mathbb{R}$, entonces Y sigue una distribución normal $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, donde:
$$\mu_Y = a\mu_X + b, \quad \sigma_Y^2 = a^2\sigma_X^2.$$

Ejemplo:

Supongamos que los tiempos de respuesta en un experimento siguen una distribución normal con media $\mu = 10$ segundos y desviación estándar $\sigma = 2$ segundos. La probabilidad de que un tiempo de respuesta sea menor o igual a 12 segundos es:

$$\mathbb{P}(X \leq 12) = \Phi\left(\frac{12-10}{2}\right) = \Phi(1).$$

Utilizando una tabla de la distribución normal estándar, obtenemos $\Phi(1) \approx 0{,}8413$. Por lo tanto, la probabilidad es aproximadamente 0,8413.

6.7.4 Distribución Gamma

La *distribución Gamma* está asociada a procesos que modelan el tiempo hasta la ocurrencia de k eventos independientes, cuando los tiempos entre eventos siguen una distribución exponencial. Para definir esta distribución, es fundamental entender la función Gamma, que generaliza el concepto de factorial para números reales y complejos positivos. La función Gamma se define como:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx, \quad \alpha > 0.$$

Cuando α es un entero positivo, se cumple que $\Gamma(\alpha) = (\alpha - 1)!$.

Definición 6.13 (Distribución Gamma) Una variable aleatoria X sigue una distribución Gamma con parámetros $\alpha > 0$ y $\lambda > 0$, denotada $X \sim \text{Gamma}(\alpha, \lambda)$, si su PDF es:

- **Función de Densidad de Probabilidad (PDF):**

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & \text{si } x \geq 0, \\ 0, & \text{si } x < 0, \end{cases}$$

donde α es el parámetro de forma y λ es el parámetro de escala.

La distribución Gamma es útil en situaciones donde se modela el tiempo de espera acumulado para múltiples eventos. Algunos ejemplos incluyen:

- El tiempo hasta que ocurren k fallos en un sistema.
- La duración total de servicio en un sistema donde los tiempos de atención siguen una distribución exponencial.
- El modelado de fenómenos en ingeniería de confiabilidad y procesos de riesgo.

Propiedades

- **Relación recursiva:** La función Gamma satisface la relación:

$$\Gamma(\alpha + 1) = \alpha \cdot \Gamma(\alpha),$$

la cual es análoga al comportamiento del factorial.

- **Valor especial:**

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Este resultado es fundamental en el cálculo de probabilidades asociadas a la distribución normal.

- **Propiedad integral:** Para cualquier $\alpha > 0$ y $\lambda > 0$, se tiene que:

$$\int_0^\infty x^{\alpha-1} e^{-\lambda x}\, dx = \frac{\Gamma(\alpha)}{\lambda^\alpha}.$$

Proposición 6.5

Sea $X \sim \text{Gamma}(\alpha, \lambda)$ una variable aleatoria con distribución Gamma. Entonces, la CDF no tiene forma cerrada, pero la esperanza y la varianza están dadas por:

$$\mathbb{E}[X] = \frac{\alpha}{\lambda}, \quad \text{Var}(X) = \frac{\alpha}{\lambda^2}.$$

Demostración:

- **Esperanza:** La esperanza de X se define como:

$$\mathbb{E}[X] = \int_0^\infty x f_X(x)\, dx = \int_0^\infty x \cdot \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\, dx.$$

Factorizamos los términos constantes fuera de la integral:

$$\mathbb{E}[X] = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-\lambda x}\, dx.$$

Realizando el cambio de variable $t = \lambda x$ ($x = \frac{t}{\lambda}$, $dx = \frac{dt}{\lambda}$), obtenemos:

$$\mathbb{E}[X] = \frac{\lambda^\alpha}{\Gamma(\alpha)\lambda^{\alpha+1}} \int_0^\infty t^\alpha e^{-t}\, dt = \frac{1}{\lambda} \cdot \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)}.$$

Dado que $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, se tiene que:

$$\mathbb{E}[X] = \frac{\alpha}{\lambda}.$$

- **Varianza:** La varianza se define como:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Calculamos $\mathbb{E}[X^2]$ de manera similar:

$$\mathbb{E}[X^2] = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha+1} e^{-\lambda x}\, dx.$$

Haciendo el cambio de variable $t = \lambda x$, obtenemos:

$$\mathbb{E}[X^2] = \frac{1}{\lambda^2} \cdot \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)}.$$

Sabemos que $\Gamma(\alpha+2) = (\alpha+1)\alpha\Gamma(\alpha)$, por lo que:

$$\mathbb{E}[X^2] = \frac{\alpha(\alpha+1)}{\lambda^2}.$$

La varianza es entonces:

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ &= \frac{\alpha(\alpha+1)}{\lambda^2} - \left(\frac{\alpha}{\lambda}\right)^2 \\ &= \frac{\alpha}{\lambda^2}.\end{aligned}$$

Ejemplo:

Supongamos que el tiempo total de espera para tres eventos consecutivos sigue una distribución Gamma con parámetros $\alpha = 3$ y $\lambda = 2$. La esperanza es:

$$\mathbb{E}[X] = \frac{3}{2} = 1{,}5.$$

La varianza es:

$$\mathrm{Var}(X) = \frac{3}{2^2} = \frac{3}{4}.$$

Capítulo 7

Distribución Conjunta

En este capítulo, exploraremos el concepto de distribución conjunta, que nos permite describir el comportamiento simultáneo de dos o más variables aleatorias. Este análisis nos permite responder preguntas como: ¿Qué sucede cuando las variables aleatorias están relacionadas? ¿Cómo se comporta un fenómeno cuya probabilidad aumenta o disminuye a medida que se relaciona con otro, como ocurre con la estatura y el peso?

La distribución conjunta nos permite modelar estas dependencias, analizar relaciones estadísticas y calcular medidas, como la covarianza y la correlación. Comenzaremos estudiando el caso de dos variables aleatorias, para luego generalizar los conceptos a un número arbitrario de variables.

7.1 Función de Distribución Acumulada de Dos Variables Aleatorias

Para introducir el estudio de variables aleatorias relacionadas, es natural comenzar con el caso más simple, dos variables aleatorias. La función de distribución acumulada conjunta (CDF) es el primer paso en esta dirección, ya que permite describir la relación entre ambas variables de manera general. Su definición se mantiene inalterada tanto en el caso discreto como en el continuo, proporcionando una base unificada para analizar la probabilidad conjunta de múltiples eventos. Comprender esta función es primordial, pues a partir de ella se derivan conceptos como independencia, correlación y distribuciones marginales.

Definición 7.1 (Función de Distribución Acumulada Conjunta) La *función de distribución acumulada conjunta* (CDF conjunta) de dos variables aleatorias X e Y se define como:

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y).$$

Propiedades

La función de distribución acumulada conjunta satisface las siguientes propiedades fundamentales:

- **Monotonía:**

  $$F_{X,Y}(x, y) \text{ es no decreciente en } x \text{ y en } y.$$

  A medida que aumentan x o y, la probabilidad acumulada no disminuye.

- **Límites:**

$$\lim_{x\to-\infty} F_{X,Y}(x,y) = 0, \qquad \lim_{y\to-\infty} F_{X,Y}(x,y) = 0,$$
$$\lim_{x\to\infty} F_{X,Y}(x,y) = F_Y(y), \qquad \lim_{y\to\infty} F_{X,Y}(x,y) = F_X(x),$$
$$\lim_{x\to\infty,y\to\infty} F_{X,Y}(x,y) = 1.$$

  Estos límites reflejan el comportamiento natural de la función de distribución acumulada conjunta. Cuando x o y tienden a $-\infty$, la probabilidad de que las variables tomen valores menores se aproxima a cero, ya que prácticamente ningún valor del espacio muestral cumple dicha condición. Por otro lado, si solo una de las variables crece hacia ∞, la función de distribución conjunta converge a la distribución marginal de la otra variable, pues en ese caso la probabilidad de la primera ya ha sido completamente acumulada. Finalmente, cuando ambas variables tienden a ∞, la función alcanza el valor 1, lo que indica que toda la probabilidad del espacio muestral ha sido considerada.

- **Probabilidad de un subconjunto:** La probabilidad de que (X, Y) tome valores en un rectángulo $[x_1, x_2] \times [y_1, y_2]$ es:

$$\mathbb{P}(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1).$$

7.2 Distribución Conjunta de Dos Variables Aleatorias Discretas

En el caso de variables aleatorias discretas, es importante no solo conocer las probabilidades individuales de cada variable, sino también la probabilidad de que ambas tomen ciertos valores específicos al mismo tiempo. Por ejemplo, si definimos X como el número de clientes que llegan a una tienda en la mañana y Y como los que llegan en la tarde, podríamos preguntarnos cuál es la probabilidad de que lleguen exactamente 3 clientes en la mañana y 2 en la tarde.

Además, estas variables no necesariamente son independientes. Es posible que el comportamiento de una influya en el de la otra. Por ejemplo, si hay una gran afluencia de clientes en la mañana, esto podría indicar, según ciertas condiciones, que en la tarde haya menos personas debido a la disponibilidad limitada de recursos o al comportamiento típico del flujo de clientes.

7.2.1 Función de Masa de Probabilidad Conjunta

Para modelar estos casos, utilizamos la *función de masa de probabilidad conjunta* (PMF conjunta), que asigna una probabilidad a cada par de valores posibles (x, y) de las variables aleatorias X e Y.

Definición 7.2 (Distribución Conjunta para Variables Discretas) La *función de masa de probabilidad conjunta* (*joint PMF*) $P_{X,Y}(x, y)$ describe la probabilidad de que dos variables aleatorias discretas X e Y tomen simultáneamente los valores x e y:

$$P_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y),$$

donde x y y recorren el rango conjunto de X e Y.

Rango

En el caso de dos variables aleatorias discretas X e Y, el rango conjunto, denotado por $R_{X,Y}$, es un subconjunto del producto cartesiano $R_X \times R_Y$, es decir, no todos los pares (x, y) de valores posibles necesariamente tienen una probabilidad positiva. Esto implica que, en general, $R_{X,Y} \neq R_X \times R_Y$, ya que algunos pares pueden tener probabilidad cero.

Formalmente, el rango conjunto se define como:

$$R_{X,Y} = \{(x, y) \in R_X \times R_Y \mid P_{X,Y}(x, y) > 0\}.$$

Por ejemplo, consideremos que X es el número de tareas completadas y Y es el número de errores detectados en un día. Supongamos que, en este contexto, es imposible que se detecten más errores que tareas completadas. Si los rangos individuales son $R_X = \{0, 1, 2, 3\}$ y $R_Y = \{0, 1, 2, 3\}$, el rango conjunto estará restringido a los pares (x, y) que cumplen $0 \leq y \leq x$. Así, aunque $(1, 3)$ pertenece al producto cartesiano $R_X \times R_Y$, no pertenece al rango conjunto $R_{X,Y}$, ya que sería imposible detectar 3 errores habiendo completado solo 1 tarea, o por ejemplo sería imposible detectar algún error si es que no se ha realizado ninguna tarea.

Propiedades

La función de masa de probabilidad conjunta (PMF) $P_{X,Y}(x, y)$ cumple las siguientes propiedades fundamentales:

- **No negatividad:**

  $$P_{X,Y}(x, y) \geq 0, \quad \forall (x, y) \in R_X \times R_Y.$$

  En todo el producto cartesiano de los rangos R_X y R_Y, las probabilidades conjuntas no pueden ser negativas.

- **Positividad en el rango conjunto:**

  $$P_{X,Y}(x, y) > 0, \quad \forall (x, y) \in R_{X,Y}.$$

  En el rango conjunto $R_{X,Y}$, todas las probabilidades conjuntas son estrictamente positivas.

- **Suma total de probabilidades:**

  $$\sum_{(x,y) \in R_{X,Y}} P_{X,Y}(x, y) = 1.$$

  La suma de todas las probabilidades en el rango conjunto es igual a 1, lo que garantiza que se cubre todo el espacio muestral.

- **Probabilidad en subconjuntos:** La probabilidad de que (X, Y) tome valores dentro de un subconjunto $A \subseteq R_X \times R_Y$ está dada por:

  $$\mathbb{P}((X, Y) \in A) = \sum_{(x,y) \in R_{X,Y} \cap A} P_{X,Y}(x, y).$$

Ejemplo:

Consideremos dos variables aleatorias discretas X e Y que modelan el siguiente experimento: X representa el número de llamadas recibidas por una central telefónica en la mañana, tal que $R_X = \{0, 1\}$, mientras que Y representa el número de llamadas recibidas en la tarde, tal que $R_Y = \{0, 1, 2\}$.

Supongamos que la función de masa de probabilidad conjunta $P_{X,Y}(x, y)$ está dada por la siguiente tabla:

$P_{X,Y}(x,y)$	$y=0$	$y=1$	$y=2$
$x=0$	0.10	0.15	0.05
$x=1$	0.20	0.30	0.20

La función $P_{X,Y}(x, y)$ describe la probabilidad de que la central telefónica reciba simultáneamente x llamadas en la mañana y y llamadas en la tarde. Por ejemplo:

- $P_{X,Y}(0, 1) = 0{,}15$ indica una probabilidad de 0,15 de recibir 0 llamadas en la mañana y 1 llamada en la tarde.
- $P_{X,Y}(1, 2) = 0{,}20$ significa que la probabilidad de recibir 1 llamada en la mañana y 2 en la tarde es 0,20.

Verificamos que la suma de todas las probabilidades sea igual a 1:

$$\begin{aligned}\sum_{x=0}^{1}\sum_{y=0}^{2} P_{X,Y}(x, y) &= 0{,}10 + 0{,}15 + 0{,}05 + 0{,}20 + 0{,}30 + 0{,}20 \\ &= 1.\end{aligned}$$

Esto confirma que $P_{X,Y}(x, y)$ es una función de masa de probabilidad conjunta válida.

7.2.2 Función de Masa de Probabilidad Marginal

Si queremos conocer la probabilidad de que una de las dos variables aleatorias tome un valor específico, sin importar qué valor asuma la otra variable, debemos sumar las probabilidades conjuntas sobre todos los valores posibles de esa segunda variable. Este procedimiento es análogo al concepto de probabilidades totales, donde se consideran todos los escenarios posibles para obtener la probabilidad de un evento en particular.

Definición 7.3 (PMF Marginal) La *función de masa de probabilidad marginal* de X, denotada $P_X(x)$, se obtiene sumando las probabilidades conjuntas sobre todos los valores posibles de Y:

$$P_X(x) = \sum_{y \in R_Y} P_{X,Y}(x, y).$$

De forma similar, la PMF marginal de Y es:

$$P_Y(y) = \sum_{x \in R_X} P_{X,Y}(x, y).$$

Ejemplo:

Consideremos el mismo ejemplo de la central telefónica. La función de masa de probabilidad conjunta $P_{X,Y}(x, y)$ está dada por la siguiente tabla:

$P_{X,Y}(x,y)$	$y=0$	$y=1$	$y=2$
$x=0$	0.10	0.15	0.05
$x=1$	0.20	0.30	0.20

Queremos calcular las funciones de masa de probabilidad marginales para X e Y.

Para calcular la PMF marginal de X, sumamos las probabilidades conjuntas para cada valor de x sobre todos los valores posibles de y:

$$\begin{aligned} P_X(0) &= P_{X,Y}(0,0) + P_{X,Y}(0,1) + P_{X,Y}(0,2) \\ &= 0{,}10 + 0{,}15 + 0{,}05 = 0{,}30, \\ P_X(1) &= P_{X,Y}(1,0) + P_{X,Y}(1,1) + P_{X,Y}(1,2) \\ &= 0{,}20 + 0{,}30 + 0{,}20 = 0{,}70. \end{aligned}$$

Mientras que, para calcular la PMF marginal de Y, sumamos las probabilidades conjuntas para cada valor de y sobre todos los valores posibles de x:

$$\begin{aligned} P_Y(0) &= P_{X,Y}(0,0) + P_{X,Y}(1,0) \\ &= 0{,}10 + 0{,}20 = 0{,}30, \\ P_Y(1) &= P_{X,Y}(0,1) + P_{X,Y}(1,1) \\ &= 0{,}15 + 0{,}30 = 0{,}45, \\ P_Y(2) &= P_{X,Y}(0,2) + P_{X,Y}(1,2) \\ &= 0{,}05 + 0{,}20 = 0{,}25. \end{aligned}$$

7.2.3 Esperanza

La *esperanza* en el contexto de dos variables aleatorias puede interpretarse de dos maneras, como el vector de esperanzas individuales o como la esperanza de una función definida en $\mathbb{R}^2 \to \mathbb{R}$.

Esperanza de un Vector

La esperanza de un vector aleatorio se define como el vector cuyas componentes son las esperanzas de las variables individuales.

Definición 7.4 (Esperanza de un Vector Aleatorio) Si (X, Y) es un vector aleatorio, su *esperanza* se define como:

$$\mathbb{E}[(X,Y)] = (\mathbb{E}[X], \mathbb{E}[Y]),$$

donde:

$$\mathbb{E}[X] = \sum_{x \in R_X} x\, P_X(x), \quad \mathbb{E}[Y] = \sum_{y \in R_Y} y\, P_Y(y).$$

Esta interpretación es útil cuando se analizan las propiedades combinadas de dos variables, como el promedio de sus comportamientos individuales.

Esperanza de una Función

La esperanza de una función de dos variables aleatorias describe el promedio ponderado de los valores que toma la función, basado en la probabilidad conjunta.

Definición 7.5 (Esperanza de una Función) Si $g(X, Y)$ es una función de dos variables aleatorias, donde $g : \mathbb{R}^2 \to \mathbb{R}$, su *esperanza* se define como:

$$\mathbb{E}[g(X, Y)] = \sum_{x \in R_X} \sum_{y \in R_Y} g(x, y)\, P_{X,Y}(x, y).$$

Esta definición es clave para calcular propiedades como la esperanza de X^2, XY o cualquier otra transformación que involucre ambas variables.

Ejemplo:

Supongamos que X e Y son dos variables aleatorias discretas con rango $R_X = \{0, 1\}$ y $R_Y = \{0, 1\}$, y la función de masa de probabilidad conjunta $P_{X,Y}(x, y)$ está dada por:

$P_{X,Y}(x, y)$	$y = 0$	$y = 1$
$x = 0$	0.10	0.20
$x = 1$	0.30	0.40

Queremos calcular $\mathbb{E}[g(X, Y)]$ para la función $g(X, Y) = X + Y$. Aplicamos la definición de esperanza de una función:

$$\begin{aligned}
\mathbb{E}[g(X, Y)] &= \sum_{x \in R_X} \sum_{y \in R_Y} (x + y)\, P_{X,Y}(x, y) \\
&= (0 + 0) \cdot 0{,}10 + (0 + 1) \cdot 0{,}20 + (1 + 0) \cdot 0{,}30 + (1 + 1) \cdot 0{,}40 \\
&= 0 + 0{,}20 + 0{,}30 + 0{,}80 \\
&= 1{,}30.
\end{aligned}$$

7.2.4 Varianza

La *varianza* en el contexto de dos variables aleatorias puede entenderse de dos maneras: como la varianza de cada componente de un vector aleatorio o como la varianza de una función definida en $\mathbb{R}^2 \to \mathbb{R}$. En ambos casos, la varianza mide la dispersión de los valores respecto a su esperanza.

Más adelante, introduciremos el concepto de *varianza de un vector aleatorio*, que involucra la relación entre las componentes del vector mediante la matriz de covarianza.

Varianza de una Función

La varianza de una función mide la dispersión de los valores de la función respecto a su esperanza, proporcionando información sobre cómo varían las transformaciones de las variables aleatorias.

Definición 7.6 (Varianza de una Función) Si $g(X, Y)$ es una función de dos variables aleatorias, donde $g : \mathbb{R}^2 \to \mathbb{R}$, su *varianza* se define como:

$$\mathrm{Var}(g(X, Y)) = \mathbb{E}[(g(X, Y) - \mathbb{E}[g(X, Y)])^2].$$

Ejemplo:

Usando el mismo ejemplo anterior, donde $g(X,Y) = X + Y$ y $P_{X,Y}(x,y)$ está dado por:

$P_{X,Y}(x,y)$	$y=0$	$y=1$
$x=0$	0.10	0.20
$x=1$	0.30	0.40

Calculamos la varianza de $g(X,Y)$. Primero, recordemos que en el ejemplo anterior obtuvimos $\mathbb{E}[g(X,Y)] = 1{,}30$. Ahora, necesitamos calcular $\mathbb{E}[(g(X,Y))^2]$:

$$\begin{aligned}\mathbb{E}[(g(X,Y))^2] &= \sum_{x\in R_X}\sum_{y\in R_Y}(x+y)^2\,P_{X,Y}(x,y)\\ &= (0+0)^2\cdot 0{,}10 + (0+1)^2\cdot 0{,}20 + (1+0)^2\cdot 0{,}30 + (1+1)^2\cdot 0{,}40\\ &= 0 + 0{,}20 + 0{,}30 + 1{,}60\\ &= 2{,}10.\end{aligned}$$

Finalmente, la varianza de $g(X,Y)$ es:

$$\begin{aligned}\mathrm{Var}(g(X,Y)) &= \mathbb{E}[(g(X,Y))^2] - (\mathbb{E}[g(X,Y)])^2\\ &= 2{,}10 - (1{,}30)^2\\ &= 2{,}10 - 1{,}69\\ &= 0{,}41.\end{aligned}$$

7.2.5 PMF Condicional

En ocasiones, nos interesa conocer la probabilidad de que una variable aleatoria discreta X tome un valor específico x, dado que ha ocurrido un evento A con $\mathbb{P}(A) > 0$. Este tipo de análisis es relevante cuando el evento A proporciona información adicional sobre el experimento aleatorio.

La función de masa de probabilidad condicional, denotada por $P_{X|A}(x)$, describe esta situación al ajustar las probabilidades teniendo en cuenta la ocurrencia del evento A. Matemáticamente, la PMF condicional se expresa como:

$$P_{X|A}(x) = \mathbb{P}(X = x \mid A) = \frac{\mathbb{P}(X = x, A)}{\mathbb{P}(A)}.$$

Esto se utiliza en situaciones donde el evento condicionante modifica las probabilidades originales. Por ejemplo, podríamos querer calcular la probabilidad de que un cliente realice una compra, dado que ha permanecido en la tienda por más de 10 minutos.

De manera similar, en ocasiones, nos interesa conocer la probabilidad conjunta de que dos variables aleatorias discretas X e Y tomen valores específicos x e y, dado que una de ellas ya ha asumido un valor conocido. Este enfoque nos permite modelar situaciones donde el conocimiento de una variable influye en la probabilidad de otra, capturando así la dependencia entre ambas.

Cuando condicionamos a un valor arbitrario y de Y, obtenemos la función de masa de probabilidad condicional de X, denotada por $P_{X|Y}(x \mid y)$. Esto representa la probabilidad de que X tome el valor x, dado que $Y = y$, y se define de la siguiente manera:

Definición 7.7 (PMF Condicional) Sean X e Y dos variables aleatorias discretas. La *función de masa de probabilidad condicional* de X dado que $Y = y$, denotada por $P_{X|Y}(x \mid y)$, se expresa como:

$$P_{X|Y}(x \mid y) = \frac{P_{X,Y}(x, y)}{P_Y(y)},$$

para todo $x \in R_X$ y $y \in R_Y$, siempre que $P_Y(y) > 0$.

Este concepto ajusta las probabilidades conjuntas para reflejar la información adicional proporcionada por el valor conocido de Y.

Ejemplo:

Supongamos que X es la variable aleatoria que representa el número de productos comprados por un cliente en una tienda, y Y es una variable que indica el tipo de cliente, donde $Y = 1$ representa un cliente habitual y $Y = 2$ un cliente nuevo. La función de masa de probabilidad conjunta $P_{X,Y}(x, y)$ está dada por la siguiente tabla:

$P_{X,Y}(x,y)$	$x=0$	$x=1$	$x=2$
$y=1$	0.1	0.3	0.1
$y=2$	0.2	0.2	0.1

Queremos calcular la función de masa de probabilidad condicional de X dado que $Y = 1$, es decir, $P_{X|Y}(x \mid 1)$.

Para esto, primero calculamos $P_Y(1)$ sumando las probabilidades conjuntas para $y = 1$:

$$P_Y(1) = \sum_{x \in R_X} P_{X,Y}(x, 1) = 0{,}1 + 0{,}3 + 0{,}1 = 0{,}5.$$

Ahora, aplicamos la definición de PMF condicional para cada valor de x:

$$P_{X|Y}(0 \mid 1) = \frac{P_{X,Y}(0,1)}{P_Y(1)} = \frac{0{,}1}{0{,}5} = 0{,}2,$$
$$P_{X|Y}(1 \mid 1) = \frac{P_{X,Y}(1,1)}{P_Y(1)} = \frac{0{,}3}{0{,}5} = 0{,}6,$$
$$P_{X|Y}(2 \mid 1) = \frac{P_{X,Y}(2,1)}{P_Y(1)} = \frac{0{,}1}{0{,}5} = 0{,}2.$$

Por lo tanto, la PMF condicional de X dado que $Y = 1$ es:

$$P_{X|Y}(x \mid 1) = \begin{cases} 0{,}2, & \text{si } x = 0, \\ 0{,}6, & \text{si } x = 1, \\ 0{,}2, & \text{si } x = 2. \end{cases}$$

7.2.6 Independencia

Hemos introducido las funciones de masa de probabilidad conjunta, marginal y condicional, las cuales describen el comportamiento de dos variables aleatorias que pueden estar relacionadas entre sí. Ahora retomaremos el concepto de independencia discutido en el capítulo 3, aplicado al caso de variables aleatorias.

Dos variables aleatorias X e Y son *independientes* si conocer el valor de una no proporciona información sobre la otra. En otras palabras, el conocimiento de Y no afecta las probabilidades de X. Como resultado, la función de masa de probabilidad conjunta se puede expresar como el producto de las funciones marginales:

Definición 7.8 (Independencia) Dos variables aleatorias X e Y son *independientes* si su función de masa de probabilidad conjunta es:

$$P_{X,Y}(x, y) = P_X(x) \cdot P_Y(y), \quad \forall x \in R_X, \forall y \in R_Y.$$

Esta definición formaliza el hecho de que, en el caso de independencia, las probabilidades conjuntas se descomponen en probabilidades individuales. Además, para variables independientes, la PMF condicional se simplifica:

$$P_{X|Y}(x \mid y) = \frac{P_{X,Y}(x, y)}{P_Y(y)} = \frac{P_X(x) \cdot P_Y(y)}{P_Y(y)} = P_X(x).$$

Ejemplo:

Supongamos que X representa el resultado del lanzamiento de un dado justo y Y el resultado de lanzar una moneda equilibrada. El rango de X es $R_X = \{1, 2, 3, 4, 5, 6\}$, mientras que el rango de Y es $R_Y = \{\text{cara}, \text{sello}\}$. Dado que ambos experimentos son independientes, la función de masa de probabilidad conjunta se calcula como:

$$P_{X,Y}(x, y) = P_X(x) \cdot P_Y(y),$$

donde:

$$P_X(x) = \frac{1}{6}, \quad P_Y(\text{cara}) = P_Y(\text{sello}) = \frac{1}{2}.$$

Por ejemplo, la probabilidad de obtener un 3 en el dado y "cara" en la moneda es:

$$P_{X,Y}(3, \text{cara}) = P_X(3) \cdot P_Y(\text{cara}) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}.$$

Del mismo modo, la probabilidad de obtener un 5 en el dado y "sello" en la moneda es:

$$P_{X,Y}(5, \text{sello}) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}.$$

En este caso, el conocimiento del resultado del dado no altera la probabilidad de obtener un resultado específico en la moneda, lo que confirma que las variables X e Y son independientes.

7.3 Distribución Conjunta de Dos Variables Aleatorias Continuas

En la sección anterior, abordamos la distribución conjunta de variables aleatorias discretas, donde es posible calcular la probabilidad de que las variables tomen valores específicos simultáneamente. En el caso de variables aleatorias continuas, el análisis cambia, ya que la probabilidad de que ambas variables tomen un valor exacto es siempre cero. En lugar de esto, utilizamos una función de densidad conjunta para describir cómo las probabilidades se distribuyen en el espacio continuo.

Para ilustrar esta situación, consideremos X como la temperatura máxima diaria en una ciudad y Y como la humedad relativa correspondiente. Aunque la probabilidad de que X sea exactamente $30°C$ y Y sea exactamente $60\,\%$ es cero, podemos estudiar la probabilidad de que ambas variables se encuentren en un rango determinado, como $29 \leq X \leq 31$ y $58 \leq Y \leq 62$. La función de densidad conjunta nos proporciona la información necesaria para realizar este cálculo.

Asimismo, como en las variables aleatorias discretas, es posible que estas variables estén relacionadas entre sí. Por ejemplo, cuando la temperatura es alta, es más probable que la humedad disminuya, lo que indica una dependencia entre X e Y.

7.3.1 Función de Densidad de Probabilidad Conjunta

Para variables aleatorias continuas, la *función de densidad de probabilidad conjunta* (PDF conjunta) describe cómo las probabilidades están distribuidas en un espacio continuo. A diferencia del caso discreto, la probabilidad de que las variables aleatorias X e Y tomen simultáneamente valores exactos es siempre cero. En cambio, calculamos la probabilidad de que X e Y se encuentren en un subconjunto del plano $\mathbb{R}^2$ integrando la PDF conjunta sobre dicho subconjunto.

Definición 7.9 (Distribución Conjunta para Variables Continuas) La *función de densidad de probabilidad conjunta* (*joint PDF*) $f_{X,Y}(x, y)$ se define como la derivada parcial de la función de distribución acumulada conjunta:

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x\, \partial y} F_{X,Y}(x, y),$$

donde $F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ es la CDF conjunta.

Rango

En el caso de dos variables aleatorias continuas X e Y, el rango conjunto, denotado por $R_{X,Y}$, es un subconjunto del producto cartesiano $R_X \times R_Y$. Esto implica que algunos pares (x, y) pueden tener densidad de probabilidad nula.

Formalmente, el rango conjunto se define como:

$$R_{X,Y} = \{(x, y) \in R_X \times R_Y \mid f_{X,Y}(x, y) > 0\}.$$

Por ejemplo, supongamos que X e Y representan las coordenadas de un punto aleatorio dentro de un círculo de radio 1. Aunque los rangos individuales pueden ser $R_X = [-1, 1]$ y $R_Y = [-1, 1]$, el rango conjunto $R_{X,Y}$ está restringido a los puntos que cumplen $x^2 + y^2 \leq 1$. Así, el punto $(1, 1)$, aunque pertenece al producto cartesiano, no está en el rango conjunto, ya que $1^2 + 1^2 = 2 > 1$.

Propiedades

La función de densidad de probabilidad conjunta $f_{X,Y}(x, y)$ satisface las siguientes propiedades fundamentales:

- **No negatividad:**

$$f_{X,Y}(x, y) \geq 0, \quad \forall (x, y) \in R_X \times R_Y.$$

La densidad conjunta no puede ser negativa en ningún punto del producto cartesiano.

- **Integrabilidad:** La integral de $f_{X,Y}(x,y)$ sobre todo el espacio es 1:
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1.$$
Esto garantiza que la densidad conjunta cubra todo el espacio muestral.
- **Probabilidad en subconjuntos:** La probabilidad de que (X,Y) tome valores dentro de un subconjunto $A \subseteq \mathbb{R}^2$ está dada por:
$$\mathbb{P}((X,Y)\in A) = \int\int_A f_{X,Y}(x,y)\,dx\,dy.$$

Ejemplo:

Supongamos que la posición de un punto aleatorio (X,Y) dentro de un cuadrado unitario sigue una distribución uniforme. La función de densidad conjunta está dada por:
$$f_{X,Y}(x,y) = \begin{cases} 1, & \text{si } 0 \le x \le 1 \text{ y } 0 \le y \le 1, \\ 0, & \text{en otro caso.} \end{cases}$$
Queremos calcular la probabilidad de que el punto caiga en el cuadrante definido por $0 \le X \le 0{,}5$ y $0 \le Y \le 0{,}5$. Esta probabilidad es:
$$\begin{aligned}\mathbb{P}((X,Y)\in[0,0{,}5]\times[0,0{,}5]) &= \int_0^{0{,}5}\int_0^{0{,}5} f_{X,Y}(x,y)\,dx\,dy \\ &= \int_0^{0{,}5}\int_0^{0{,}5} 1\,dx\,dy \\ &= (0{,}5-0)^2 = 0{,}25.\end{aligned}$$

7.3.2 Función de Densidad de Probabilidad Marginal

Si queremos conocer la densidad de probabilidad de una variable aleatoria continua, sin considerar el valor que asuma la otra variable, debemos integrar la función de densidad conjunta sobre todos los valores posibles de esa segunda variable. Este proceso es análogo al caso discreto, pero, considerando la naturaleza del caso continuo.

Definición 7.10 (PDF Marginal) La *función de densidad de probabilidad marginal* de X, denotada $f_X(x)$, se obtiene integrando la densidad conjunta sobre todos los valores posibles de Y:
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy.$$
De forma similar, la PDF marginal de Y es:
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx.$$

Ejemplo:

Supongamos que las variables aleatorias X e Y siguen una distribución uniforme en el cuadrado unitario, es decir:

$$f_{X,Y}(x,y) = \begin{cases} 1, & \text{si } 0 \leq x \leq 1 \text{ y } 0 \leq y \leq 1, \\ 0, & \text{en otro caso.} \end{cases}$$

Queremos calcular las funciones de densidad marginales de X e Y.

Para la PDF marginal de X, integramos la densidad conjunta sobre el rango de Y:

$$\begin{aligned} f_X(x) &= \int_0^1 f_{X,Y}(x,y)\,dy \\ &= \int_0^1 1\,dy \\ &= 1, \quad \text{para } 0 \leq x \leq 1. \end{aligned}$$

De manera similar, para la PDF marginal de Y, integramos la densidad conjunta sobre el rango de X:

$$\begin{aligned} f_Y(y) &= \int_0^1 f_{X,Y}(x,y)\,dx \\ &= \int_0^1 1\,dx \\ &= 1, \quad \text{para } 0 \leq y \leq 1. \end{aligned}$$

Ambas densidades marginales muestran que tanto X como Y siguen una distribución uniforme en el intervalo $[0,1]$. La probabilidad de que ambas variables estén en un intervalo específico se calcula a partir de estas PDFs.

7.3.3 Esperanza

En el caso de dos variables aleatorias continuas, el concepto de esperanza es análogo al presentado en el caso discreto, con la diferencia de que las sumas sobre los rangos se sustituyen por integrales sobre el rango continuo.

Esperanza de un Vector

La esperanza de un vector aleatorio continuo se define de manera similar a la del caso discreto, utilizando las funciones de densidad marginales.

Definición 7.11 (Esperanza de un Vector Aleatorio) Si (X,Y) es un vector aleatorio continuo, su *esperanza* es el vector:

$$\mathbb{E}[(X,Y)] = (\mathbb{E}[X], \mathbb{E}[Y]),$$

donde:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)\,dx, \quad \mathbb{E}[Y] = \int_{-\infty}^{\infty} y\, f_Y(y)\,dy.$$

Esperanza de una Función

Cuando se considera una función de dos variables aleatorias continuas, la esperanza se define integrando la función ponderada por la densidad conjunta de probabilidad.

Definición 7.12 (Esperanza de una Función) Si $g(X, Y)$ es una función de dos variables aleatorias continuas, donde $g : \mathbb{R}^2 \to \mathbb{R}$, su *esperanza* se define como:

$$\mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)\, f_{X,Y}(x, y)\, dx\, dy.$$

Esta definición nos permita calcular la esperanza de transformaciones no lineales de X e Y.

Ejemplo:

Supongamos que X e Y son dos variables aleatorias continuas definidas en el círculo unitario centrado en el origen, es decir, el conjunto $\{(x, y) \mid x^2 + y^2 \leq 1\}$. La función de densidad conjunta está dada por:

$$f_{X,Y}(x, y) = \frac{1}{\pi}, \quad \text{para } x^2 + y^2 \leq 1.$$

Queremos calcular $\mathbb{E}[g(X, Y)]$ para la función $g(X, Y) = X + Y$. Para resolver esta integral, aprovechamos la simetría del círculo, y podemos integrar respecto a una de las variables, expresando la otra en términos de los límites del círculo. Aplicamos la definición de esperanza:

$$\begin{aligned}\mathbb{E}[g(X, Y)] &= \int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} (x + y)\, f_{X,Y}(x, y)\, dy\, dx \\ &= \int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} (x + y) \cdot \frac{1}{\pi}\, dy\, dx.\end{aligned}$$

Dividimos la integral en dos términos:

$$\mathbb{E}[g(X, Y)] = \frac{1}{\pi} \left(\int_{-1}^{1} x \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy + \int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y\, dy\, dx \right).$$

Calculamos las integrales, para el primer término:

$$\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy = 2\sqrt{1 - x^2}, \quad \text{por lo que} \quad \int_{-1}^{1} x \cdot 2\sqrt{1 - x^2}\, dx = 0.$$

Mientras que para el segundo término:

$$\int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y\, dy\, dx = 0.$$

Ambos términos se anulan, debido a la propiedad de integración de funciones impares, [1] por lo que la esperanza es:

$$\mathbb{E}[g(X, Y)] = 0.$$

[1] Si una función $f(x)$ es *impar*, es decir, $f(-x) = -f(x)$, entonces su integral en un intervalo simétrico respecto al origen es cero: $\int_{-a}^{a} f(x)\, dx = 0$. Por otro lado, si una función $g(x)$ es *par*, es decir, $g(-x) = g(x)$, su integral en el mismo intervalo simétrico satisface $\int_{-a}^{a} g(x)\, dx = 2 \int_{0}^{a} g(x)\, dx$. Estas propiedades resultan útiles en cálculos que involucran simetrías en el dominio de integración.

Alternativamente, podríamos haber utilizado coordenadas polares para simplificar el dominio de integración. En ese caso, se obtendría el mismo resultado debido a la simetría del círculo.

7.3.4 Varianza

En el caso continuo, la *varianza* también puede interpretarse de dos maneras, como la varianza de cada componente de un vector aleatorio o como la varianza de una función definida en $\mathbb{R}^2 \to \mathbb{R}$. Todo lo mencionado sobre la interpretación y utilidad de la varianza en el caso discreto aplica de manera análoga aquí, con las diferencias propias de la naturaleza continua.

Varianza de una Función

La varianza de una función continua mide cómo se dispersan los valores de la función respecto a su esperanza en el dominio de integración.

Definición 7.13 (Varianza de una Función) Si $g(X, Y)$ es una función de dos variables aleatorias continuas, donde $g : \mathbb{R}^2 \to \mathbb{R}$, su *varianza* se define como:

$$\mathrm{Var}(g(X, Y)) = \mathbb{E}[(g(X, Y) - \mathbb{E}[g(X, Y)])^2].$$

Ejemplo:

Usando el mismo ejemplo anterior, donde X e Y son variables aleatorias definidas en el círculo unitario centrado en el origen, con la función de densidad conjunta:

$$f_{X,Y}(x, y) = \frac{1}{\pi}, \quad \text{para } x^2 + y^2 \leq 1,$$

y donde $g(X, Y) = X + Y$, calculamos la varianza de $g(X, Y)$.

Primero, recordemos que en el ejemplo de esperanza obtuvimos $\mathbb{E}[g(X, Y)] = 0$. Ahora, necesitamos calcular $\mathbb{E}[(g(X, Y))^2]$:

$$\begin{aligned}\mathbb{E}[(g(X, Y))^2] &= \int_{x^2+y^2\leq 1} (x + y)^2 \, f_{X,Y}(x, y) \, dx \, dy \\ &= \int_{x^2+y^2\leq 1} (x^2 + 2xy + y^2) \cdot \frac{1}{\pi} \, dx \, dy.\end{aligned}$$

Dado que el dominio es simétrico respecto al origen, el término $2xy$ se anula al integrar. Por lo tanto:

$$\mathbb{E}[(g(X, Y))^2] = \frac{1}{\pi} \int_{x^2+y^2\leq 1} (x^2 + y^2) \, dx \, dy.$$

Utilizando coordenadas polares, donde $x = r\cos\theta$, $y = r\sin\theta$, $r \in [0, 1]$, $\theta \in [0, 2\pi]$, y $dx\,dy = r\,dr\,d\theta$, la integral se convierte en:

$$\begin{aligned}\mathbb{E}[(g(X, Y))^2] &= \frac{1}{\pi} \int_0^{2\pi} \int_0^1 r^2 \, r \, dr \, d\theta \\ &= \frac{1}{\pi} \int_0^{2\pi} d\theta \int_0^1 r^3 \, dr \\ &= \frac{1}{\pi} \cdot 2\pi \cdot \frac{1}{4} = \frac{1}{2}.\end{aligned}$$

Finalmente, la varianza de $g(X,Y)$ es:

$$\begin{aligned}\mathrm{Var}(g(X,Y)) &= \mathbb{E}[(g(X,Y))^2] - (\mathbb{E}[g(X,Y)])^2 \\ &= \frac{1}{2} - 0^2 = \frac{1}{2}.\end{aligned}$$

7.3.5 Función de Densidad de Probabilidad Condicional

A menudo, al igual que en el caso discreto, es necesario determinar cómo se distribuye una variable aleatoria continua X dado que ha ocurrido un evento A con $\mathbb{P}(A) > 0$.

La *función de densidad de probabilidad condicional*, denotada $f_{X|A}(x)$, describe la densidad de probabilidad de X bajo la condición de que el evento A haya ocurrido. Se define como:

$$f_{X|A}(x) = \frac{\mathbb{P}(X \in dx, A)}{\mathbb{P}(A)}.$$

De manera equivalente, utilizando la función de densidad conjunta:

$$f_{X|A}(x) = \frac{\int_A f_{X,Y}(x,y)\,dy}{\mathbb{P}(A)}.$$

Sin embargo, nos interesa conocer también el cómo se comporta una variable aleatoria continua X cuando se tiene información adicional sobre otra variable aleatoria continua Y.

En el caso continuo, utilizamos la *función de densidad de probabilidad condicional* (PDF condicional), que describe la densidad de probabilidad de X dado que Y ha tomado un valor específico. Esta función se obtiene de la relación entre la densidad conjunta y la densidad marginal.

Definición 7.14 (PDF Condicional) Sean X e Y dos variables aleatorias continuas. La *función de densidad de probabilidad condicional* de X dado que $Y = y$, denotada por $f_{X|Y}(x \mid y)$, se define como:

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)},$$

para todo $x \in R_X$ y $y \in R_Y$, siempre que $f_Y(y) > 0$.

Este concepto refleja la idea de ajustar la densidad conjunta en función de la información adicional sobre Y. Como en el caso discreto, se busca capturar cómo cambia la probabilidad cuando conocemos un valor particular de la otra variable.

Ejemplo:

Supongamos que X e Y son dos variables aleatorias continuas definidas en el cuadrado $[0,1] \times [0,1]$, con una función de densidad conjunta:

$$f_{X,Y}(x,y) = 6xy, \quad \text{para } 0 \le x \le 1,\ 0 \le y \le 1.$$

Queremos calcular la función de densidad condicional de X dado que $Y = y$, es decir, $f_{X|Y}(x \mid y)$.

Primero, calculamos la densidad marginal de Y, $f_Y(y)$, integrando la densidad conjunta sobre todos los posibles valores de X:

$$\begin{aligned} f_Y(y) &= \int_0^1 f_{X,Y}(x,y)\,dx \\ &= \int_0^1 6xy\,dx \\ &= 6y\int_0^1 x\,dx \\ &= 6y\cdot\frac{1}{2} = 3y. \end{aligned}$$

Ahora, aplicamos la definición de PDF condicional:

$$\begin{aligned} f_{X|Y}(x \mid y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \frac{6xy}{3y} = 2x, \quad \text{para } 0 \le x \le 1 \text{ y } 0 < y \le 1. \end{aligned}$$

Por lo tanto, la función de densidad condicional es:

$$f_{X|Y}(x \mid y) = \begin{cases} 2x, & \text{si } 0 \le x \le 1 \text{ y } y > 0, \\ 0, & \text{en otro caso.} \end{cases}$$

7.3.6 Independencia

La independencia para variables aleatorias continuas sigue el mismo principio que en el caso discreto. Dos variables aleatorias X e Y son independientes si el conocimiento del valor de una no altera la distribución de la otra. Matemáticamente, esto se traduce en que la función de densidad conjunta puede descomponerse como el producto de las densidades marginales.

Definición 7.15 (Independencia) Dos variables aleatorias continuas X e Y son *independientes* si su función de densidad conjunta satisface:

$$f_{X,Y}(x,y) = f_X(x)\cdot f_Y(y), \quad \forall x, y \in \mathbb{R}.$$

Esta propiedad implica que la densidad condicional de X dado que $Y = y$ es igual a la densidad marginal de X:

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = f_X(x).$$

Ejemplo:

Consideremos que X representa la longitud (en metros) de un cable producido por una máquina, y Y el tiempo hasta una falla (en horas) de otra máquina. Supongamos que las funciones de densidad marginales están definidas en $[0,1]$ por:

$$f_X(x) = 2x, \quad f_Y(y) = 3y^2.$$

Si X e Y son independientes, la función de densidad conjunta es:

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) = 6xy^2.$$

Calculemos la probabilidad de que $X \leq 0{,}5$ y $Y \leq 0{,}5$:

$$\begin{aligned}
\mathbb{P}(X \leq 0{,}5, Y \leq 0{,}5) &= \int_0^{0{,}5} \int_0^{0{,}5} 6xy^2 \, dx \, dy \\
&= 6 \int_0^{0{,}5} y^2 \left(\int_0^{0{,}5} x \, dx \right) dy \\
&= 6 \int_0^{0{,}5} y^2 \cdot \left(\frac{x^2}{2} \Big|_0^{0{,}5} \right) dy \\
&= 6 \int_0^{0{,}5} y^2 \cdot \frac{0{,}25}{2} \, dy \\
&= 6 \cdot \frac{0{,}25}{2} \int_0^{0{,}5} y^2 \, dy \\
&= 0{,}75 \int_0^{0{,}5} y^2 \, dy \\
&= 0{,}75 \cdot \frac{y^3}{3} \Big|_0^{0{,}5} \\
&= 0{,}75 \cdot \frac{(0{,}5)^3}{3} \\
&= \frac{0{,}75}{24} = \frac{1}{32}.
\end{aligned}$$

Por lo tanto, la probabilidad buscada es $\frac{1}{32}$.

7.3.7 Función de Dos Variables Aleatorias Continuas

Como se vio en el caso de una variable aleatoria continua, es común que una transformación de una variable aleatoria produzca otra variable cuyo comportamiento probabilístico nos interesa caracterizar. De manera análoga, cuando trabajamos con dos variables aleatorias continuas (X, Y) y aplicamos una transformación $(U, V) = g(X, Y)$, donde $g : \mathbb{R}^2 \to \mathbb{R}^2$, surge la necesidad de determinar la distribución conjunta de las nuevas variables.

En esta sección, exploraremos cómo obtener la función de densidad conjunta de (U, V) a partir de la densidad conjunta de (X, Y).

Sea $h : \mathbb{R}^2 \to \mathbb{R}$ una función de dos variables aleatorias continuas X e Y. Definimos una nueva variable aleatoria Z como:

$$Z = h(X, Y).$$

Nuestro objetivo es determinar la distribución de Z. Para ello, comenzamos escribiendo su función de distribución acumulada (CDF):

$$\begin{aligned}
F_Z(z) &= \mathbb{P}(Z \leq z) \\
&= \mathbb{P}(h(X, Y) \leq z) \\
&= \iint_D f_{X,Y}(x, y) \, dx \, dy,
\end{aligned}$$

donde $D = \{(x, y) \mid h(x, y) \leq z\}$ es la región del plano donde la función $h(X, Y)$ toma valores menores o iguales a z.

Para obtener la función de densidad de probabilidad (PDF) de Z, simplemente derivamos $F_Z(z)$ con respecto a z:

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} \iint_D f_{X,Y}(x, y)\, dx\, dy.$$

Cambio de Variable

Si realizamos un cambio de variables a partir de (X, Y) mediante una transformación diferenciable $(U, V) = g(X, Y)$, donde $g : \mathbb{R}^2 \to \mathbb{R}^2$, la función de densidad de (U, V) se obtiene ajustando la densidad de (X, Y) según cómo la transformación redistribuye la probabilidad.

Esto nos permite obtener la densidad conjunta de las nuevas variables a partir de la densidad de las variables originales. Como en el caso unidimensional, necesitamos corregir el efecto de la transformación en la probabilidad, para ello utilizaremos el determinante del Jacobiano, asegurando que la densidad resultante conserve las propiedades de una función de densidad.

Teorema 7.1 Cambio de Variable en Dos Variables Aleatorias

Sean X, Y variables aleatorias continuas y sea $g : \mathbb{R}^2 \to \mathbb{R}^2$ una función diferenciable e invertible, definida como:

$$g(x, y) = \begin{bmatrix} g_1(x, y) \\ g_2(x, y) \end{bmatrix},$$

donde cada función componente $g_i : \mathbb{R}^2 \to \mathbb{R}$ es diferenciable. Entonces, si realizamos la transformación

$$(U, V) = g(X, Y), \quad \text{o equivalentemente} \quad (X, Y) = g^{-1}(U, V),$$

las variables aleatorias U, V son continuas y su densidad conjunta se obtiene mediante:

$$f_{U,V}(u, v) = f_{X,Y}(g^{-1}(u, v)) \cdot |\det J_g|^{-1},$$

donde $J_g(x, y)$ es el Jacobiano de la transformación g, dado por:

$$J_g(x, y) = \begin{bmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{bmatrix}.$$

Demostración: Consideremos dos variables aleatorias continuas X e Y con función de densidad conjunta $f_{X,Y}(x, y)$. Sea $g : \mathbb{R}^2 \to \mathbb{R}^2$ una transformación diferenciable e invertible, definida como:

$$g(x, y) = \begin{bmatrix} g_1(x, y) \\ g_2(x, y) \end{bmatrix}.$$

Definimos las nuevas variables aleatorias (U, V) mediante la transformación:

$$(U, V) = g(X, Y),$$

con inversa $(X, Y) = g^{-1}(U, V)$. Nuestro objetivo es encontrar la densidad conjunta de (U, V), es decir, $f_{U,V}(u, v)$. La probabilidad de que (U, V) tome valores en una región $S \subseteq \mathbb{R}^2$ está dada por:

$$\mathbb{P}((U, V) \in S) = \int\int_S f_{U,V}(u, v)\, du\, dv.$$

Dado que la transformación es invertible, esta probabilidad también se puede expresar en términos de (X, Y) como:

$$\mathbb{P}((X, Y) \in g^{-1}(S)) = \int\int_{g^{-1}(S)} f_{X,Y}(x, y)\, dx\, dy.$$

El teorema del cambio de variables en integrales nos dice que si realizamos la transformación $(x, y) = g^{-1}(u, v)$, el elemento diferencial cambia de acuerdo con el determinante del Jacobiano de la transformación inversa:

$$dx\, dy = |\det J_g(x, y)|^{-1}\, du\, dv,$$

donde $J_g(x, y)$ es la matriz Jacobiana de g:

$$J_g(x, y) = \begin{bmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{bmatrix}.$$

Sustituyendo en la integral de (X, Y), obtenemos:

$$\int\int_S f_{U,V}(u, v)\, du\, dv = \int\int_{g^{-1}(S)} f_{X,Y}(x, y)\, |\det J_g|^{-1}\, du\, dv.$$

Por el Teorema Fundamental del Cálculo aplicado a integrales dobles, al tomar derivadas parciales respecto a u y v, obtenemos la densidad conjunta de (U, V):

$$f_{U,V}(u, v) = f_{X,Y}(g^{-1}(u, v)) \cdot |\det J_g|^{-1}.$$

Ejemplo:

Supongamos que X e Y son variables aleatorias continuas con densidad conjunta:

$$f_{X,Y}(x, y) = \begin{cases} 1, & 0 \le x \le 1,\ 0 \le y \le 1, \\ 0, & \text{en otro caso.} \end{cases}$$

Definimos la transformación:

$$U = X + Y = g_1(X, Y), \quad V = X - Y = g_2(X, Y).$$

Nuestro objetivo es encontrar la densidad conjunta $f_{U,V}(u, v)$. Para expresar (X, Y) en términos de (U, V), resolvemos el sistema de ecuaciones:

$$X = \frac{U + V}{2}, \quad Y = \frac{U - V}{2}.$$

El Jacobiano de la transformación es:

$$J_g(x, y) = \begin{bmatrix} \frac{\partial g_1}{\partial X} & \frac{\partial g_1}{\partial Y} \\ \frac{\partial g_2}{\partial X} & \frac{\partial g_2}{\partial Y} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

El determinante de esta matriz es:

$$\det J_g = (-1) - (1) = -2.$$

Por lo tanto, tomamos su valor absoluto para la transformación:

$$|\det J_g|^{-1} = \frac{1}{2}.$$

Utilizando el teorema del cambio de variable:

$$f_{U,V}(u,v) = f_{X,Y}(g^{-1}(u,v)) \cdot |\det J_g|^{-1}.$$

Dado que la densidad original es $f_{X,Y}(x,y) = 1$ en $0 \leq x \leq 1$, $0 \leq y \leq 1$, verificamos cómo cambia el rango bajo la nueva transformación:

- $X \in [0,1]$ y $Y \in [0,1]$ implican $U = X + Y \in [0,2]$.
- Simultáneamente, $V = X - Y$ varía en el rango $[-1,1]$.

Entonces, la densidad conjunta en las nuevas variables es:

$$f_{U,V}(u,v) = \begin{cases} \frac{1}{2}, & 0 \leq u \leq 2,\ -1 \leq v \leq 1, \\ 0, & \text{en otro caso.} \end{cases}$$

Generalización Cambio de Variable

Si la transformación $(U,V) = g(X,Y)$ no es inyectiva en todo su dominio, podemos dividir el espacio en regiones donde sí lo sea y aplicar el resultado en cada una de ellas de manera separada.

Teorema 7.2 Generalización del Cambio de Variable

Sean X, Y variables aleatorias continuas con densidad conjunta $f_{X,Y}(x,y)$ y sea $g : \mathbb{R}^2 \to \mathbb{R}^2$ una función diferenciable que no es globalmente inyectiva. Supongamos que el dominio de g se puede particionar en regiones disjuntas R_i, donde g es inyectiva en cada una de ellas. En estas regiones, existen funciones inversas locales $\varrho_i : \mathbb{R}^2 \to \mathbb{R}^2$, bien definidas en sus respectivos dominios.
Bajo estas condiciones, la densidad conjunta de $(U,V) = g(X,Y)$ se obtiene sumando las contribuciones de cada región:

$$f_{U,V}(u,v) = \sum_i f_{X,Y}(\varrho_i(u,v)) \cdot |\det J_{\varrho_i}|^{-1},$$

donde $J_{\varrho_i}(x,y)$ es el Jacobiano de la transformación inversa ϱ_i, evaluado en la preimagen correspondiente:

$$J_{\varrho_i}(x,y) = \begin{bmatrix} \frac{\partial \varrho_{i,1}}{\partial x} & \frac{\partial \varrho_{i,1}}{\partial y} \\ \frac{\partial \varrho_{i,2}}{\partial x} & \frac{\partial \varrho_{i,2}}{\partial y} \end{bmatrix}.$$

Esta generalización es útil cuando la transformación tiene múltiples preimágenes en el espacio original. Casos comunes incluyen cambios de coordenadas en sistemas polares o cilíndricos, donde un mismo punto en el espacio transformado puede provenir de distintos puntos en el dominio original debido a la periodicidad o simetría de la transformación.

7.4 Distribución Conjunta de Múltiples Variables Aleatorias

Todo lo desarrollado en las secciones anteriores para una o dos variables aleatorias puede extenderse al caso de un número arbitrario de variables aleatorias. Para ello, introducimos el concepto de *vector aleatorio*, que nos permite describir simultáneamente un conjunto de variables aleatorias y analizar su distribución conjunta.

7.4.1 Vector Aleatorio

¿Por qué trabajar con vectores? El vector es una representación matemática que permite organizar múltiples variables de manera estructurada, facilitando su manipulación y análisis. En el contexto de la probabilidad, los vectores nos permiten modelar conjuntos de variables aleatorias relacionadas, proporcionando una notación compacta y herramientas analíticas para describir su comportamiento conjunto.

Utilizar vectores en lugar de tratar cada variable individualmente nos permite aprovechar propiedades algebraicas y conceptos que se verán posteriormente como la covarianza, la independencia y la transformación de distribuciones.

Definición 7.16 (Vector Aleatorio) Un *vector aleatorio* de dimensiones n es un conjunto de n variables aleatorias $(X_1, X_2, \ldots, X_n)$ definidas en un mismo espacio de probabilidad. Se denota usualmente como:

$$\mathbf{X} = (X_1, X_2, \ldots, X_n).$$

Al igual que en los casos de una o dos variables, podemos describir la distribución conjunta de n variables aleatorias de forma discreta o continua.

7.4.2 Función de Distribución Acumulada

El concepto de función de distribución acumulada conjunta extendido a n variables aleatorias, mantiene la misma interpretación probabilística que en el caso bidimensional. En este contexto, la función de distribución acumulada conjunta describe la probabilidad de que todas las variables tomen valores menores o iguales a ciertos umbrales simultáneamente. Es decir, permite cuantificar la probabilidad acumulada dentro de una región multidimensional definida por los valores de las variables.

Definición 7.17 (Función de Distribución Acumulada Conjunta) Sea $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ un vector de variables aleatorias. La *función de distribución acumulada conjunta* (CDF conjunta) se define como:

$$F_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n).$$

Propiedades

La CDF conjunta de n variables aleatorias cumple las mismas propiedades que se mencionaron anteriormente, adaptada al caso de las n variables aleatorias, estas son, las siguientes propiedades generales:

- **Monotonía:**

$$F_{\mathbf{X}}(x_1, x_2, \ldots, x_n) \text{ es no decreciente en cada } x_i.$$

A medida que aumenta cualquiera de las coordenadas x_i, la probabilidad acumulada no disminuye.

- **Límites:**

$$\lim_{x_i \to -\infty} F_{\mathbf{X}}(x_1, \ldots, x_n) = 0, \quad \forall i,$$
$$\lim_{x_i \to \infty} F_{\mathbf{X}}(x_1, \ldots, x_n) = F_{X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n),$$
$$\lim_{x_1,\ldots,x_n \to \infty} F_{\mathbf{X}}(x_1, \ldots, x_n) = 1.$$

La función crece desde 0 y converge a 1 a medida que todas las variables tienden a infinito.

- **Probabilidad en un subconjunto:** La probabilidad de que $\mathbf{X}$ tome valores dentro de un hiperrectángulo $[x_1^{(1)}, x_1^{(2)}] \times \cdots \times [x_n^{(1)}, x_n^{(2)}]$ se obtiene mediante:

$$\mathbb{P}(x_1^{(1)} < X_1 \leq x_1^{(2)}, \ldots, x_n^{(1)} < X_n \leq x_n^{(2)}) = \sum_{\boldsymbol{\epsilon} \in \{0,1\}^n} (-1)^{|\boldsymbol{\epsilon}|} F_{\mathbf{X}}(x_1^{(\epsilon_1)}, \ldots, x_n^{(\epsilon_n)}),$$

donde $|\boldsymbol{\epsilon}|$ es la suma de los elementos de $\boldsymbol{\epsilon}$, y los índices $x_i^{(\epsilon_i)}$ alternan entre los extremos del intervalo.

7.4.3 Función de Masa de Probabilidad o Densidad Conjunta

A continuación se presenta el concepto de distribución conjunta para múltiples variables aleatorias, para esto, necesitamos una función que describa la probabilidad de que todas ellas tomen valores específicos simultáneamente. Dependiendo de si las variables son discretas o continuas, esta función puede ser una función de masa de probabilidad conjunta (PMF conjunta) o una función de densidad de probabilidad conjunta (PDF conjunta).

Definición 7.18 (Distribución Conjunta para Múltiples Variables Aleatorias) Sea $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ un vector aleatorio. *Su distribución conjunta* se caracteriza de la siguiente manera:

- **Caso Discreto:** Si $X_1, X_2, \ldots, X_n$ son variables aleatorias discretas, la *función de masa de probabilidad conjunta* (PMF conjunta) se define como:

$$P_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n).$$

- **Caso Continuo:** Si $X_1, X_2, \ldots, X_n$ son variables aleatorias continuas, la *función de densidad de probabilidad conjunta* (PDF conjunta) se define como:

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \ldots \partial x_n} F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n),$$

donde $F_{X_1,X_2,\ldots,X_n}$ es la función de distribución acumulada conjunta.

Rango

El rango conjunto de un vector aleatorio $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ es el subconjunto del espacio $R_{X_1} \times R_{X_2} \times \cdots \times R_{X_n}$ donde la PMF o la PDF es positiva. En lugar de considerar pares (x, y) dentro de un subconjunto del plano, ahora trabajamos con tuplas $(x_1, x_2, \ldots, x_n)$ dentro de un subconjunto del espacio n-dimensional.

Definición 7.19 (Rango) El *rango conjunto* de $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ depende de si las variables aleatorias son discretas o continuas:

- **Caso Discreto:** Si $X_1, X_2, \ldots, X_n$ son variables aleatorias discretas, el rango conjunto es el conjunto de valores donde la función de masa de probabilidad conjunta es estrictamente positiva:

$$R_{X_1,\ldots,X_n} = \{(x_1, \ldots, x_n) \in R_{X_1} \times \cdots \times R_{X_n} \mid P_{X_1,\ldots,X_n}(x_1, \ldots, x_n) > 0\}.$$

- **Caso Continuo:** Si $X_1, X_2, \ldots, X_n$ son variables aleatorias continuas, el rango conjunto se define en términos de la función de densidad de probabilidad conjunta:

$$R_{X_1,\ldots,X_n} = \{(x_1, \ldots, x_n) \in R_{X_1} \times \cdots \times R_{X_n} \mid f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) > 0\}.$$

Como ya se ha mencionado, el rango conjunto no necesariamente coincide con el producto cartesiano de los rangos individuales, ya que ciertos valores pueden ser imposibles o tener densidad cero.

7.4.4 Funciones de Masa de Probabilidad o Densidad Marginal

Como se mencionó, si nos interesa analizar una sola variable aleatoria dentro de un conjunto de múltiples variables, es necesario obtener su distribución marginal, la cual describe su comportamiento sin considerar explícitamente a las demás. Esto se logra eliminando las demás variables del análisis mediante sumas (en el caso discreto) o integrales (en el caso continuo), de manera análoga a lo visto en el caso de dos variables.

Definición 7.20 (PMF o PDF Marginal) La *función de probabilidad* o *densidad marginal* de una variable X_i se obtiene sumando (en el caso discreto) o integrando (en el caso continuo) la función de probabilidad o densidad conjunta sobre las restantes variables:

$$P_{X_i}(x_i) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} P_{X_1,X_2,\ldots,X_n}(x_1, \ldots, x_n),$$

para el caso discreto, y

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,X_2,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n,$$

para el caso continuo.

7.4.5 Esperanza

El concepto de esperanza para un vector aleatorio $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, permite calcular la esperanza como un vector cuyas componentes son las esperanzas individuales de cada variable. Asimismo, cuando trabajamos con una función de múltiples variables aleatorias, su esperanza se define mediante sumas o integrales, dependiendo de si las variables son discretas o continuas.

Esperanza de un Vector

La esperanza de un vector aleatorio se define considerando la esperanza de cada una de sus componentes de manera individual.

Definición 7.21 (Esperanza de un Vector Aleatorio) Si $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ es un vector aleatorio, su *esperanza* se define como:

$$\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}] = (\mathbb{E}[X_1], \mathbb{E}[X_2], \ldots, \mathbb{E}[X_n]),$$

donde cada esperanza individual está dada por:

- **Caso Discreto:**
$$\mathbb{E}[X_i] = \sum_{x_i \in R_{X_i}} x_i \, P_{X_i}(x_i).$$
- **Caso Continuo:**
$$\mathbb{E}[X_i] = \int_{-\infty}^{\infty} x_i \, f_{X_i}(x_i) \, dx_i.$$

Esperanza de una Función

Cuando se trabaja con una función de múltiples variables aleatorias, su esperanza se obtiene mediante sumas o integrales, dependiendo del tipo de variables.

Definición 7.22 (Esperanza de una Función) Si $g(X_1, X_2, \ldots, X_n)$ es una función de n variables aleatorias, tal que $g : \mathbb{R}^n \to \mathbb{R}$, su *esperanza* se define como:

- **Caso Discreto:**
$$\mathbb{E}[g(X_1, \ldots, X_n)] = \sum_{x_1} \cdots \sum_{x_n} g(x_1, \ldots, x_n) P_{X_1,\ldots,X_n}(x_1, \ldots, x_n).$$
- **Caso Continuo:**
$$\mathbb{E}[g(X_1, \ldots, X_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \, dx_1 \ldots dx_n.$$

Esta expresión permite calcular el valor esperado de cualquier función que dependa de múltiples variables aleatorias.

7.4.6 Varianza

Para un conjunto de n variables aleatorias, la *varianza* puede analizarse en los contextos ya mencionados, en donde, seguiremos analizando el caso de una función.

Varianza de una Función

La varianza de una función de múltiples variables aleatorias, que mapea estas en el espacio de los reales ($\mathbb{R}$), esta dada por:

Definición 7.23 (Varianza de una Función) Si $g(X_1, X_2, \ldots, X_n)$ es una función de n variables aleatorias, tal que $g : \mathbb{R}^n \to \mathbb{R}$, su *varianza* se define de la siguiente manera:

- **Caso Discreto:**

$$\mathrm{Var}(g(X_1, \ldots, X_n)) = \mathbb{E}\left[(g(X_1, \ldots, X_n) - \mathbb{E}[g(X_1, \ldots, X_n)])^2\right],$$

donde la esperanza se calcula como

$$\mathbb{E}[g(X_1, \ldots, X_n)] = \sum_{x_1} \cdots \sum_{x_n} g(x_1, \ldots, x_n) P_{X_1,\ldots,X_n}(x_1, \ldots, x_n).$$

- **Caso Continuo:**

$$\mathrm{Var}(g(X_1, \ldots, X_n)) = \mathbb{E}\left[(g(X_1, \ldots, X_n) - \mathbb{E}[g(X_1, \ldots, X_n)])^2\right],$$

donde la esperanza se calcula mediante

$$\mathbb{E}[g(X_1, \ldots, X_n)] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n.$$

7.4.7 Función de Probabilidad o Densidad Condicional

La ocurrencia de un evento A con $\mathbb{P}(A) > 0$, como se ha visto, modifica la distribución de las variables aleatorias bajo estudio. Si $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, la función de probabilidad o densidad condicional describe su distribución dentro de esta nueva condición, como se define a continuación:

- **Caso Discreto:**

$$P_{\mathbf{X}|A}(x_1, x_2, \ldots, x_n \mid A) = \frac{\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n, A)}{\mathbb{P}(A)}.$$

- **Caso Continuo:**

$$f_{\mathbf{X}|A}(x_1, x_2, \ldots, x_n \mid A) = \frac{\int_A f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)\, d\mathbf{x}}{\mathbb{P}(A)},$$

donde la integral se toma sobre la región del espacio de $\mathbf{X}$ que corresponde al evento A.

Conocer el valor de algunas variables aleatorias también nos proporciona información sobre la distribución de otras. Cuando se tiene un vector aleatorio $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ y se conoce que algunas de sus componentes toman valores específicos, la función de probabilidad o densidad condicional nos permite redefinir la distribución de las restantes variables.

Definición 7.24 (PMF o PDF Condicional) Sea $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ un vector aleatorio y sea $\mathbf{Y} = (X_{k+1}, \ldots, X_n)$ un subconjunto de sus componentes. La *función de probabilidad* o *densidad condicional* de $\mathbf{X}_1 = (X_1, \ldots, X_k)$ dado que $\mathbf{Y} = \mathbf{y}$ se define como:

- **Caso Discreto:**

$$P_{X_1,\ldots,X_k|X_{k+1},\ldots,X_n}(x_1, \ldots, x_k \mid y_{k+1}, \ldots, y_n) = \frac{P_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)}{P_{X_{k+1},\ldots,X_n}(y_{k+1}, \ldots, y_n)},$$

  siempre que $P_{X_{k+1},\ldots,X_n}(y_{k+1}, \ldots, y_n) > 0$.

- **Caso Continuo:**

$$f_{X_1,\ldots,X_k|X_{k+1},\ldots,X_n}(x_1, \ldots, x_k \mid y_{k+1}, \ldots, y_n) = \frac{f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)}{f_{X_{k+1},\ldots,X_n}(y_{k+1}, \ldots, y_n)},$$

  siempre que $f_{X_{k+1},\ldots,X_n}(y_{k+1}, \ldots, y_n) > 0$.

7.4.8 Independencia

El concepto de independencia se mantiene independiente del número de variables, un conjunto de variables aleatorias es independiente si su distribución conjunta es el producto de sus distribuciones individuales, es decir, si conocer algunas de ellas no afecta la distribución de las demás. Para n variables, esto se define como sigue.

Definición 7.25 (Independencia de Múltiples Variables Aleatorias) Las variables aleatorias $X_1, X_2, \ldots, X_n$ son *independientes* si y solo si para cualquier subconjunto $\{i_1, i_2, \ldots, i_k\} \subseteq \{1, 2, \ldots, n\}$ con $1 \leq k \leq n$, se cumple que:

- **Caso Discreto:**

$$P_{X_{i_1},X_{i_2},\ldots,X_{i_k}}(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) = P_{X_{i_1}}(x_{i_1}) P_{X_{i_2}}(x_{i_2}) \ldots P_{X_{i_k}}(x_{i_k}),$$

  para todo $(x_{i_1}, x_{i_2}, \ldots, x_{i_k})$ en sus respectivos rangos.

- **Caso Continuo:**

$$f_{X_{i_1},X_{i_2},\ldots,X_{i_k}}(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) = f_{X_{i_1}}(x_{i_1}) f_{X_{i_2}}(x_{i_2}) \ldots f_{X_{i_k}}(x_{i_k}),$$

  para todo $(x_{i_1}, x_{i_2}, \ldots, x_{i_k})$ en el soporte de la distribución conjunta.

Es importante notar que esta condición no solo debe cumplirse para el conjunto completo de n variables, sino también para cualquier subconjunto de ellas. Si existe al menos un subconjunto de variables que no satisface esta propiedad, entonces las variables no son independientes. No obstante, es posible que solo algunos grupos de variables sean independientes entre sí, sin que todo el conjunto lo sea.

En términos de probabilidad condicional, las variables aleatorias $X_1, X_2, \ldots, X_n$ son independientes si y solo si, para cualquier partición del conjunto de índices en dos subconjuntos disjuntos $\{i_1, \ldots, i_k\}$ y $\{j_1, \ldots, j_m\}$, se cumple que:

- **Caso Discreto:**

$$P_{X_{i_1},\ldots,X_{i_k}|X_{j_1},\ldots,X_{j_m}}(x_{i_1},\ldots,x_{i_k} \mid x_{j_1},\ldots,x_{j_m}) = P_{X_{i_1},\ldots,X_{i_k}}(x_{i_1},\ldots,x_{i_k}),$$

siempre que $P_{X_{j_1},\ldots,X_{j_m}}(x_{j_1},\ldots,x_{j_m}) > 0$.

- **Caso Continuo:**

$$f_{X_{i_1},\ldots,X_{i_k}|X_{j_1},\ldots,X_{j_m}}(x_{i_1},\ldots,x_{i_k} \mid x_{j_1},\ldots,x_{j_m}) = f_{X_{i_1},\ldots,X_{i_k}}(x_{i_1},\ldots,x_{i_k}),$$

siempre que $f_{X_{j_1},\ldots,X_{j_m}}(x_{j_1},\ldots,x_{j_m}) > 0$.

Sin embargo, en la mayoría de los fenómenos naturales y sociales, las variables están interrelacionadas de alguna manera, por lo que la independencia es una condición fuerte que debe verificarse cuidadosamente antes de asumirla.

7.4.9 Función de Múltiples Variables Aleatorias Continuas

Si tenemos un vector aleatorio continuo $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ y aplicamos una transformación diferenciable $\mathbf{U} = g(\mathbf{X})$, donde $g : \mathbb{R}^n \to \mathbb{R}^n$, nos interesa determinar la densidad conjunta de $\mathbf{U}$.

Cambio de Variable

Si realizamos una transformación diferenciable e invertible $g : \mathbb{R}^n \to \mathbb{R}^n$, tal que $\mathbf{U} = g(\mathbf{X})$, entonces podemos expresar la densidad conjunta de $\mathbf{U}$ en términos de la densidad conjunta de $\mathbf{X}$.

Teorema 7.3 Cambio de Variable en n Variables

Sea $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ un vector de variables aleatorias continuas con densidad conjunta $f_{\mathbf{X}}(\mathbf{x})$, y sea $g : \mathbb{R}^n \to \mathbb{R}^n$ una transformación diferenciable e invertible. Definimos $\mathbf{U} = (U_1, U_2, \ldots, U_n)$ como la transformación:

$$\mathbf{U} = g(\mathbf{X}).$$

Entonces, la densidad conjunta de $\mathbf{U}$ se obtiene como:

$$f_{\mathbf{U}}(\mathbf{u}) = f_{\mathbf{X}}(g^{-1}(\mathbf{u})) \cdot \left|\det J_g\right|^{-1},$$

donde J_g es la matriz Jacobiana de la transformación g:

$$J_g(\mathbf{x}) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_n} \end{bmatrix}.$$

Generalización Cambio de Variable

Es importante recordar que algunas transformaciones $(U, V) = g(X, Y)$ no son inyectivas en todo su dominio. En estos casos, es posible dividir el espacio en regiones donde la transformación sea inyectiva y aplicar el resultado en cada una de ellas por separado.

Teorema 7.4 Generalización del Cambio de Variable

Sea $\mathbf{X}$ un vector aleatorio continuo con densidad conjunta $f_{\mathbf{X}}(\mathbf{x})$, y sea $g : \mathbb{R}^n \to \mathbb{R}^n$ una función diferenciable que no es globalmente inyectiva. Supongamos que el dominio de g se puede particionar en regiones disjuntas R_i, donde g es inyectiva en cada una de ellas. En estas regiones, existen funciones inversas locales $\varrho_i : \mathbb{R}^n \to \mathbb{R}^n$, bien definidas en sus respectivos dominios.
Bajo estas condiciones, la densidad conjunta de $\mathbf{U} = g(\mathbf{X})$ se obtiene sumando las contribuciones de cada región:

$$f_{\mathbf{U}}(\mathbf{u}) = \sum_i f_{\mathbf{X}}(\varrho_i(\mathbf{u})) \cdot |\det J_{\varrho_i}|^{-1},$$

donde J_{ϱ_i} es el Jacobiano de la transformación inversa ϱ_i, evaluado en la preimagen correspondiente:

$$J_{\varrho_i}(\mathbf{x}) = \begin{bmatrix} \frac{\partial \varrho_{i,1}}{\partial x_1} & \frac{\partial \varrho_{i,1}}{\partial x_2} & \cdots & \frac{\partial \varrho_{i,1}}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \varrho_{i,n}}{\partial x_1} & \frac{\partial \varrho_{i,n}}{\partial x_2} & \cdots & \frac{\partial \varrho_{i,n}}{\partial x_n} \end{bmatrix}.$$

7.5 Covarianza

En la mayoría de las veces en la vida cotidiana, las variables aleatorias no son independientes, sino que están relacionadas de alguna manera. En estos casos, el conocimiento del valor de una variable puede proporcionar información sobre la otra. Un caso común de relación entre variables es la dependencia lineal, donde un aumento o disminución en una variable puede estar asociado con un cambio proporcional en otra. Para cuantificar esta relación, introducimos la extitcovarianza, que mide la dependencia lineal entre dos variables aleatorias.

Definición 7.26 (Covarianza) Dadas dos variables aleatorias X e Y, la *covarianza* entre ellas se define como:

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

La covarianza es útil en estadística y probabilidad, ya que permite medir la dirección de la relación entre variables aleatorias. Sin embargo, su valor numérico depende de las unidades de las variables, por lo que resulta difícil interpretarla de manera absoluta sin normalización.

Proposición 7.1

Sea X e Y dos variables aleatorias. La covarianza se puede expresar como:

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Demostración: Por definición, la covarianza entre dos variables aleatorias X e Y se define como:

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Desarrollamos el producto en el interior de la esperanza:

$$\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]].$$

Aplicamos la linealidad de la esperanza:

$$\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y].$$

Observamos que los términos $\mathbb{E}[X]\mathbb{E}[Y]$ se cancelan:

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Propiedades

La covarianza posee diversas propiedades que facilitan su cálculo e interpretación:

- **Relación con la Varianza:** La covarianza de una variable consigo misma es su varianza:
$$\mathrm{Cov}(X,X) = \mathrm{Var}(X).$$
- **Independencia:** Si X e Y son independientes, entonces:
$$\mathrm{Cov}(X,Y) = 0.$$

Observación. Si bien la independencia entre dos variables aleatorias X e Y implica que su covarianza es cero, la inversa no necesariamente es cierta. Es decir, si $\mathrm{Cov}(X,Y) = 0$, esto no garantiza que X e Y sean independientes.

La covarianza mide únicamente la relación lineal entre las variables, por lo que su anulación solo indica que no existe una dependencia lineal entre ellas. Sin embargo, podrían estar relacionadas de manera no lineal y, por lo tanto, no ser independientes.

Un ejemplo de esto ocurre si X es una variable aleatoria uniformemente distribuida en $[-1,1]$ y definimos $Y = X^2$. En este caso, aunque $\mathrm{Cov}(X,Y) = 0$, X e Y no son independientes, ya que el valor de Y depende completamente de X.

- **Simetría:** La covarianza es simétrica, es decir:
$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X).$$
- **Homogeneidad:** Para cualquier constante a:
$$\mathrm{Cov}(aX,Y) = a\,\mathrm{Cov}(X,Y).$$
- **Invariancia ante traslaciones constantes:** Para cualquier constante c:
$$\mathrm{Cov}(X+c,Y) = \mathrm{Cov}(X,Y).$$

- **Aditividad:** Para tres variables aleatorias X, Y, Z:
$$\mathrm{Cov}(X+Y, Z) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z).$$
- **Expresión general para combinaciones lineales:** Sea $\{X_i\}_{i=1}^m$ y $\{Y_j\}_{j=1}^n$ dos conjuntos de variables aleatorias, y sean $\{a_i\}_{i=1}^m$ y $\{b_j\}_{j=1}^n$ constantes, entonces:
$$\mathrm{Cov}\left(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j\right) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j \,\mathrm{Cov}(X_i, Y_j).$$

Ejemplo:

Supongamos que X e Y son dos variables aleatorias discretas con la siguiente función de masa de probabilidad conjunta:

$P_{X,Y}(x,y)$	$y=0$	$y=1$	$y=2$
$x=0$	0.1	0.2	0.1
$x=1$	0.2	0.3	0.1

Primero, calculamos las funciones de masa de probabilidad marginales de X e Y sumando sobre la otra variable, para la variable aleatoria X se tiene que:

$$P_X(0) = P_{X,Y}(0,0) + P_{X,Y}(0,1) + P_{X,Y}(0,2) = 0{,}1 + 0{,}2 + 0{,}1 = 0{,}4,$$
$$P_X(1) = P_{X,Y}(1,0) + P_{X,Y}(1,1) + P_{X,Y}(1,2) = 0{,}2 + 0{,}3 + 0{,}1 = 0{,}6.$$

Mientras que para la variable aleatoria Y:

$$P_Y(0) = P_{X,Y}(0,0) + P_{X,Y}(1,0) = 0{,}1 + 0{,}2 = 0{,}3,$$
$$P_Y(1) = P_{X,Y}(0,1) + P_{X,Y}(1,1) = 0{,}2 + 0{,}3 = 0{,}5,$$
$$P_Y(2) = P_{X,Y}(0,2) + P_{X,Y}(1,2) = 0{,}1 + 0{,}1 = 0{,}2.$$

Ahora, calculamos las esperanzas marginales:

$$\mathbb{E}[X] = \sum_x x P_X(x) = (0 \cdot 0{,}4) + (1 \cdot 0{,}6) = 0{,}6,$$
$$\mathbb{E}[Y] = \sum_y y P_Y(y) = (0 \cdot 0{,}3) + (1 \cdot 0{,}5) + (2 \cdot 0{,}2) = 0{,}9.$$

A continuación, calculamos $\mathbb{E}[XY]$:

$$\begin{aligned}\mathbb{E}[XY] &= \sum_x \sum_y xy P_{X,Y}(x,y) \\ &= (0 \cdot 0 \cdot 0{,}1) + (0 \cdot 1 \cdot 0{,}2) + (0 \cdot 2 \cdot 0{,}1) \\ &\quad + (1 \cdot 0 \cdot 0{,}2) + (1 \cdot 1 \cdot 0{,}3) + (1 \cdot 2 \cdot 0{,}1) \\ &= 0 + 0 + 0 + 0 + 0{,}3 + 0{,}2 = 0{,}5.\end{aligned}$$

Finalmente, aplicamos la definición de covarianza:

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= 0{,}5 - (0{,}6 \times 0{,}9) \\ &= 0{,}5 - 0{,}54 \\ &= -0{,}04.\end{aligned}$$

Dado que la covarianza es negativa, esto indica que existe una relación inversa entre X e Y, es decir, cuando una de las variables aumenta, la otra tiende a disminuir.

7.5.1 Matriz de Covarianza

Hasta ahora, hemos trabajado con la varianza de variables aleatorias individuales, pero no hemos definido aún la varianza de un vector aleatorio. La razón es que, en este caso, no basta con medir la dispersión de cada componente de forma aislada, sino que también debemos capturar la relación entre ellas. La varianza de un vector aleatorio se define de la siguiente manera:

$$\mathrm{Var}(\mathbf{X}) = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^\top].$$

Al desarrollar esta expresión, se obtiene una relación que no solo mide la dispersión de cada variable individualmente, sino que también describe cómo las diferentes componentes de $\mathbf{X}$ interactúan entre sí. Es precisamente esta estructura la que da origen a la *matriz de covarianza*, la cual captura la relación lineal entre las variables en el sistema.

Definición 7.27 (Matriz de Covarianza) Sea $\mathbf{X} = (X_1, X_2, \ldots, X_n)^\top$ un vector de variables aleatorias. La *matriz de covarianza* de $\mathbf{X}$ es la matriz $n \times n$ definida por:

$$\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{X}) = \begin{bmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Cov}(X_2, X_2) & \cdots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \cdots & \mathrm{Cov}(X_n, X_n) \end{bmatrix}.$$

Cada entrada Σ_{ij} representa la covarianza entre X_i y X_j, es decir:

$$\Sigma_{ij} = \mathrm{Cov}(X_i, X_j).$$

Esta matriz proporciona una forma sistemática de analizar la relación de dependencia lineal entre las variables del vector aleatorio $\mathbf{X}$, extendiendo la noción de varianza unidimensional al caso multivariado.

Propiedades

- **Simetría:** La matriz de covarianza es simétrica, es decir, $\Sigma_{ij} = \Sigma_{ji}$.
- **Varianzas en la diagonal:** Los elementos diagonales de Σ corresponden a las varianzas de las variables aleatorias:

$$\Sigma_{ii} = \mathrm{Var}(X_i).$$

- **Semidefinida positiva:** Para cualquier vector $a \in \mathbb{R}^n$, se cumple que:

$$a^\top \Sigma a \geq 0.$$

- **Independencia e independencia lineal:** Si X_i y X_j son independientes, entonces $\mathrm{Cov}(X_i, X_j) = 0$.

7.6 Correlación

Uno de los inconvenientes de la covarianza es que su valor depende de las unidades de las variables, lo que dificulta la comparación entre diferentes pares de variables. Para solucionar esto, se introduce la correlación, una medida estandarizada que cuantifica la relación lineal entre dos variables aleatorias. A diferencia de la covarianza, la correlación es adimensional y está acotada en el intervalo $[-1, 1]$, lo que permite interpretar su magnitud de manera uniforme y compararla directamente en distintos contextos.

Definición 7.28 (Coeficiente de Correlación) Dadas dos variables aleatorias X e Y, el *coeficiente de correlación de Pearson* se define como:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y},$$

donde $\text{Cov}(X, Y)$ es la covarianza entre X e Y, y $\sigma_X = \sqrt{\text{Var}(X)}$ y $\sigma_Y = \sqrt{\text{Var}(Y)}$ son las desviaciones estándar de X e Y, respectivamente.

El coeficiente de correlación es útil porque estandariza la relación entre dos variables, facilitando su interpretación sin importar sus escalas. La correlación mide únicamente la relación lineal entre las variables, por lo que valores cercanos a cero no implican ausencia de relación, sino simplemente la falta de una dependencia lineal. En la práctica, se utiliza el coeficiente de correlación para evaluar la intensidad de la relación entre dos variables y su posible aplicación en modelos predictivos de Machine Learning.

Propiedades

El coeficiente de correlación posee varias propiedades fundamentales que permiten interpretar su comportamiento y utilidad en la relación entre variables aleatorias:

- **Acotación:**
$$-1 \leq \rho(X, Y) \leq 1.$$
Este resultado garantiza que la correlación siempre se mantenga dentro de este intervalo, actuando como una medida de relación normalizada.

- **Simetría:**
$$\rho(X, Y) = \rho(Y, X).$$
La correlación no depende del orden en el que se consideren las variables.

- **Relación con la independencia:** Si X e Y son independientes, entonces:
$$\rho(X, Y) = 0.$$
Sin embargo, la inversa no necesariamente es cierta: una correlación nula solo implica ausencia de dependencia lineal, pero las variables podrían estar relacionadas de forma no lineal.

- **Casos extremos:**
$$\rho(X, Y) = 1 \quad \text{si } Y = aX + b, \text{ con } a > 0.$$

$$\rho(X,Y) = -1 \quad \text{si } Y = aX + b, \text{ con } a < 0.$$

Un coeficiente de correlación de ± 1 indica una relación lineal perfecta entre las variables, es decir, todos los puntos de (X, Y) se encuentran sobre una línea recta.

- **Invarianza ante transformaciones afines:** Si $X^* = aX + b$ e $Y^* = cY + d$, con $a, c > 0$, entonces:
$$\rho(X^*, Y^*) = \rho(X, Y).$$
Esto significa que la correlación es invariante ante cambios de escala y traslaciones en las variables.

- **Dependencia de la varianza:** Si $\text{Var}(X) = 0$ o $\text{Var}(Y) = 0$, entonces:
$$\rho(X,Y) \text{ no está definida.}$$
Esto ocurre porque la correlación se normaliza por las desviaciones estándar, y si una variable tiene varianza cero, su desviación estándar es cero, lo que genera una indeterminación en la fórmula.

Ejemplo:

Utilizando el ejemplo anterior donde calculamos la covarianza entre X e Y, ahora determinaremos el coeficiente de correlación $\rho(X, Y)$. Supongamos que X e Y son dos variables aleatorias discretas con la siguiente función de masa de probabilidad conjunta:

$P_{X,Y}(x,y)$	$y=0$	$y=1$	$y=2$
$x=0$	0.1	0.2	0.1
$x=1$	0.2	0.3	0.1

Recordemos que la correlación se define como:

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}.$$

A partir del cálculo previo, ya conocemos:

$$\text{Cov}(X,Y) = -0{,}04, \quad \mathbb{E}[X] = 0{,}6, \quad \mathbb{E}[Y] = 0{,}9.$$

Ahora, calculamos la varianza de X e Y. Para calcular la varianza, determinamos $\mathbb{E}[X^2]$ y $\mathbb{E}[Y^2]$:

$$\mathbb{E}[X^2] = \sum_x x^2 P_X(x) = 0^2(0{,}4) + 1^2(0{,}6) = 0{,}6,$$
$$\mathbb{E}[Y^2] = \sum_y y^2 P_Y(y) = 0^2(0{,}3) + 1^2(0{,}5) + 2^2(0{,}2) = 1{,}3.$$

Finalmente, las varianzas son:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 0{,}6 - (0{,}6)^2 = 0{,}6 - 0{,}36 = 0{,}24,$$
$$\text{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = 1{,}3 - (0{,}9)^2 = 1{,}3 - 0{,}81 = 0{,}49.$$

Sustituyendo estos valores en la fórmula de la correlación:

$$\begin{aligned}\rho(X,Y) &= \frac{-0{,}04}{\sqrt{0{,}24}\cdot\sqrt{0{,}49}} \\ &= \frac{-0{,}04}{\sqrt{0{,}24}\times 0{,}7} \\ &= \frac{-0{,}04}{0{,}343} \approx -0{,}117.\end{aligned}$$

El coeficiente de correlación $\rho(X,Y) \approx -0{,}117$ indica que X e Y tienen una débil relación lineal negativa. Esto significa que, en promedio, cuando X aumenta, Y tiende a disminuir ligeramente.

7.7 Distribuciones Conjuntas Conocidas

7.7.1 Distribución Multinomial

La *distribución multinomial* es una generalización de la distribución binomial para múltiples categorías. Modela el número de veces que ocurren distintos resultados en un experimento con n ensayos independientes, donde cada ensayo puede tener uno de k posibles resultados con probabilidades fijas.

Definición 7.29 (Distribución Multinomial) Un vector aleatorio $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ sigue una *distribución multinomial* $\mathbf{X} \sim \text{Multinomial}(n, \boldsymbol{p})$ con $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)$ si cuenta la cantidad de veces que ocurren k posibles resultados en n ensayos independientes, donde cada resultado ocurre con probabilidad fija p_i y $\sum_{i=1}^{k} p_i = 1$.

- **Función de Masa de Probabilidad (PMF):**

$$P(X_1 = x_1, \ldots, X_k = x_k) = \frac{n!}{x_1! x_2! \ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k},$$

donde $x_1, x_2, \ldots, x_k$ son valores enteros no negativos tales que $\sum_{i=1}^{k} x_i = n$.

La distribución multinomial se utiliza en problemas donde se cuentan eventos que pueden caer en diferentes categorías. Algunos ejemplos incluyen:

- Clasificación de respuestas en una encuesta con múltiples opciones.
- Resultados de lanzar un dado n veces y contar cuántas veces cae cada número.
- Modelado de la frecuencia de palabras en un texto en función de categorías predefinidas.

Propiedades

- **Esperanza y Varianza:** Para cada categoría i, la esperanza y varianza de X_i están dadas por:

$$\begin{aligned}\mathbb{E}[X_i] &= np_i, \\ \text{Var}(X_i) &= np_i(1-p_i).\end{aligned}$$

- **Covarianza entre categorías:** Para cualquier par de categorías distintas i y j:
$$\text{Cov}(X_i, X_j) = -np_ip_j.$$
La covarianza es negativa porque un aumento en una categoría implica una reducción en otra, dado que $\sum X_i = n$.

- **Distribuciones marginales:** Si $\mathbf{X} \sim \text{Multinomial}(n, \boldsymbol{p})$, entonces cada variable individual X_i sigue una distribución binomial:
$$X_i \sim \text{Binomial}(n, p_i).$$

- **Suma de vectores multinomiales:** Si $\mathbf{X}_1 \sim \text{Multinomial}(n_1, \boldsymbol{p})$ y $\mathbf{X}_2 \sim \text{Multinomial}(n_2, \boldsymbol{p})$, y son independientes, entonces su suma también sigue una distribución multinomial:
$$\mathbf{X}_1 + \mathbf{X}_2 \sim \text{Multinomial}(n_1 + n_2, \boldsymbol{p}).$$

Ejemplo:

Supongamos que lanzamos un dado equilibrado de seis caras $n = 10$ veces y queremos conocer la probabilidad de obtener exactamente $x_1 = 2$ unos, $x_2 = 3$ doses, $x_3 = 1$ tres, $x_4 = 1$ cuatro, $x_5 = 2$ cincos y $x_6 = 1$ seis. Dado que cada cara tiene una probabilidad de $p_i = \frac{1}{6}$, la probabilidad se calcula con la función de masa de probabilidad:

$$\begin{aligned} P(X_1 = 2, X_2 = 3, X_3 = 1, X_4 = 1, X_5 = 2, X_6 = 1) &= \frac{10!}{2!3!1!1!2!1!} \left(\frac{1}{6}\right)^{2+3+1+1+2+1} \\ &= \frac{10!}{2!3!1!1!2!1!} \left(\frac{1}{6}\right)^{10}. \end{aligned}$$

7.7.2 Distribución Multinormal

La *distribución multinormal* es una generalización de la distribución normal a múltiples variables aleatorias.

Definición 7.30 (Distribución Multinormal) Un vector aleatorio $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ sigue una *distribución multinormal* $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, si cualquier combinación lineal de sus componentes tiene una distribución normal. Donde:

- $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$ es el vector de medias.
- $\boldsymbol{\Sigma}$ es la matriz de covarianza de tamaño $n \times n$, donde $\Sigma_{ij} = \text{Cov}(X_i, X_j)$.
- **Función de Densidad de Probabilidad (PDF):**
$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right),$$
donde $|\boldsymbol{\Sigma}|$ denota el determinante de la matriz de covarianza.

La distribución multinormal aparece en numerosas aplicaciones estadísticas, especialmente en problemas de estimación y clasificación. Se utiliza cuando las variables tienen una relación conjunta con estructura normal. Algunos ejemplos incluyen:

- Modelado de errores en métodos de regresión multivariada.
- Representación de la distribución de características en modelos de aprendizaje automático.
- Análisis de series de tiempo multivariadas con correlación entre variables.

Propiedades

- **Marginales normales:** Si $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces cada componente X_i sigue una distribución normal univariada:
$$X_i \sim \mathcal{N}(\mu_i, \Sigma_{ii}).$$
- **Distribución de subconjuntos:** Si $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, y tomamos un subconjunto de $m < n$ variables, este subconjunto también sigue una distribución normal:
$$\mathbf{X}_m \sim \mathcal{N}_m(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m),$$
donde $\boldsymbol{\mu}_m$ y $\boldsymbol{\Sigma}_m$ son los parámetros correspondientes en $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$.
- **Independencia y Covarianza:** Si $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces X_i y X_j son independientes si y solo si $\text{Cov}(X_i, X_j) = 0$, es decir, si $\Sigma_{ij} = 0$.
- **Transformaciones lineales:** Si $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y $\mathbf{A}$ es una matriz de dimensiones $m \times n$, entonces el vector transformado
$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$$
sigue una distribución normal de dimensión m:
$$\mathbf{Y} \sim \mathcal{N}_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top).$$

Ejemplo:

Supongamos que $\mathbf{X} = (X_1, X_2)$ sigue una distribución multinormal bivariada con:
$$\boldsymbol{\mu} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0{,}5 \\ 0{,}5 & 2 \end{bmatrix}.$$
Queremos calcular la probabilidad de que $\mathbf{X}$ tome valores en un subconjunto dado.

Para encontrar la densidad en un punto particular, evaluamos la función de densidad conjunta:
$$f_{\mathbf{X}}(x_1, x_2) = \frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right).$$
Calculando el determinante y la inversa de $\boldsymbol{\Sigma}$, podemos encontrar la densidad en un punto específico. Esto nos permite modelar fenómenos donde dos variables tienen una relación lineal con dependencia normal.

7.8 Estadísticos de Orden

En muchos problemas, no solo nos interesa conocer el comportamiento global de una muestra, sino también identificar sus valores más bajos, más altos o ciertos cuantiles intermedios. En problemas de confiabilidad, por ejemplo, es crucial determinar el primer tiempo de falla de un sistema para prever su mantenimiento. En economía, el análisis de ingresos no se enfoca únicamente en la media, sino en cómo se distribuyen los valores, lo que se refleja en cuantiles como la mediana o el percentil 90. En meteorología, comprender los valores extremos es esencial para modelar eventos climáticos adversos.

Dado un conjunto de n variables aleatorias independientes e idénticamente distribuidas *(i.i.d.)*, ordenar sus valores nos permite obtener una estructura que facilita el estudio de estos aspectos. Esta nueva secuencia de variables, conocidas como *estadísticos de orden*, no solo describe los valores más relevantes dentro de la muestra, sino que también juega un papel clave en inferencia estadística y teoría de valores extremos.

Definición 7.31 (Estadísticos de Orden) Sean $X_1, X_2, \ldots, X_n$ variables aleatorias continuas i.i.d. con función de densidad $f(x)$ y función de distribución acumulada $F(x)$. Definimos los *estadísticos de orden* como las variables aleatorias obtenidas al ordenar la muestra en orden creciente:

$$\begin{aligned} X_{(1)} &= \text{mín}(X_1, X_2, \ldots, X_n), && \text{(valor mínimo)} \\ X_{(2)} &= \text{segundo menor de } X_1, X_2, \ldots, X_n, && \text{(segundo menor)} \\ &\vdots \\ X_{(n)} &= \text{máx}(X_1, X_2, \ldots, X_n). && \text{(valor máximo)} \end{aligned}$$

Cada $X_{(i)}$ representa el i-ésimo valor más pequeño de la muestra, es decir, $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ forma la versión ordenada de la muestra original.

7.8.1 Función de Distribución Acumulada de los Estadísticos de Orden

Al ser variables aleatorias independientes, podría pensarse que la probabilidad de que el k-ésimo estadístico de orden sea menor que un valor dado simplemente sigue la función de distribución acumulada (CDF) de la distribución original. Sin embargo, el problema es más sutil, cuando analizamos la distribución de un estadístico de orden, estamos considerando múltiples configuraciones posibles de la muestra.

Para que $X_{(k)} \leq x$, necesitamos que al menos k valores de la muestra sean menores o iguales a x. En otras palabras, no basta con que un solo valor sea menor que x, sino que debe cumplirse esta condición para al menos k de ellos. A su vez, los valores restantes pueden ser menores, iguales o mayores a x.

Notamos que podemos modelar cada valor X_i de la muestra como una variable Bernoulli que indica si $X_i \leq x$. En este caso, cada X_i contribuye con un valor de 1 con probabilidad $F(x)$ y con un valor de 0 con probabilidad $1 - F(x)$. Dado que no nos interesa el orden específico en el que ocurren estos eventos, y aprovechando la propiedad de independencia, el número total de valores en la muestra que cumplen la

condición sigue una distribución binomial $\text{Bin}(n, F(x))$, lo que nos lleva a la siguiente expresión.

Definición 7.32 (Función de Distribución Acumulada de un Estadístico de Orden) Sea $X_{(k)}$ el k-ésimo estadístico de orden en una muestra de n variables aleatorias i.i.d. con función de distribución acumulada $F(x)$. Entonces, la *función de distribución acumulada* de $X_{(k)}$ está dada por:

$$F_{X_{(k)}}(x) = \mathbb{P}(X_{(k)} \leq x) = \sum_{j=k}^{n} \binom{n}{j} F(x)^j (1 - F(x))^{n-j}.$$

Cada término en la sumatoria representa la probabilidad de que exactamente j de los n valores sean menores o iguales a x, mientras que los restantes $n - j$ sean mayores. Como nos interesa la probabilidad de que *al menos* k valores cumplan esta condición, debemos considerar todos los casos en los que $j \geq k$.

La sumatoria acumula estas probabilidades, asegurando que contabilizamos todas las formas en que el k-ésimo menor valor pueda ser menor o igual a x. Si solo consideráramos el caso $j = k$, subestimaríamos la probabilidad total. Así, la CDF del estadístico de orden $X_{(k)}$ se obtiene sumando desde $j = k$ hasta $j = n$, cubriendo todas las configuraciones posibles.

7.8.2 Función de Densidad de los Estadísticos de Orden

Dado lo que conocemos sobre las variables aleatorias y su distribución, es posible determinar la función de densidad de probabilidad (PDF) de un estadístico de orden a partir de su función de distribución acumulada. En particular, al derivar la expresión obtenida para $F_{X_{(k)}}(x)$, obtenemos su densidad de probabilidad.

Proposición 7.2 Función de Densidad de un Estadístico de Orden

Sea $X_{(k)}$ el k-ésimo estadístico de orden en una muestra de n variables aleatorias i.i.d. con función de densidad $f(x)$ y función de distribución acumulada $F(x)$. Su *función de densidad de probabilidad* está dada por:

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} F(x)^{k-1} (1 - F(x))^{n-k} f(x).$$

Esta expresión describe la distribución del k-ésimo estadístico de orden en una muestra de tamaño n. El coeficiente combinatorio $\frac{n!}{(k-1)!(n-k)!}$ cuenta las formas de elegir los valores menores y mayores a x. La probabilidad de que $k-1$ valores sean menores es $F(x)^{k-1}$, mientras que la de que $n - k$ valores sean mayores es $(1 - F(x))^{n-k}$. Finalmente, el término $f(x)$ ajusta la densidad al intervalo infinitesimal, asegurando una distribución válida.

Demostración: Partimos de la función de distribución acumulada del estadístico de orden $X_{(k)}$:

$$F_{X_{(k)}}(x) = \sum_{j=k}^{n} \binom{n}{j} F(x)^j (1 - F(x))^{n-j}.$$

Para obtener la densidad de $X_{(k)}$, derivamos ambos lados con respecto a x:

$$f_{X_{(k)}}(x) = \frac{d}{dx} \sum_{j=k}^{n} \binom{n}{j} F(x)^j (1 - F(x))^{n-j}.$$

Aplicando la linealidad de la derivada:

$$f_{X_{(k)}}(x) = \sum_{j=k}^{n} \binom{n}{j} \frac{d}{dx} \left[F(x)^j (1 - F(x))^{n-j} \right].$$

Utilizamos la regla del producto para derivar el término dentro de la sumatoria:

$$\begin{aligned} &\frac{d}{dx} \left[F(x)^j (1 - F(x))^{n-j} \right] \\ &= jF(x)^{j-1}(1 - F(x))^{n-j} \frac{d}{dx} F(x) \ + (n-j)F(x)^j (1 - F(x))^{n-j-1} \frac{d}{dx}(1 - F(x)). \end{aligned}$$

Como $\frac{d}{dx}F(x) = f(x)$ y $\frac{d}{dx}(1 - F(x)) = -f(x)$, sustituimos:

$$= jF(x)^{j-1}(1 - F(x))^{n-j} f(x) - (n-j)F(x)^j (1 - F(x))^{n-j-1} f(x).$$

Factorizamos por $f(x)$:

$$\frac{d}{dx} \left[F(x)^j (1 - F(x))^{n-j} \right] = f(x) \left[jF(x)^{j-1}(1 - F(x))^{n-j} - (n-j)F(x)^j (1 - F(x))^{n-j-1} \right].$$

Sustituyendo en la sumatoria:

$$f_{X_{(k)}}(x) = f(x) \sum_{j=k}^{n} \binom{n}{j} \left[jF(x)^{j-1}(1 - F(x))^{n-j} - (n-j)F(x)^j (1 - F(x))^{n-j-1} \right].$$

Separando ahora la sumatoria en dos, tenemos:

$$f_{X_{(k)}}(x) = f(x) \left[\sum_{j=k}^{n} j \binom{n}{j} F(x)^{j-1}(1 - F(x))^{n-j} - \sum_{j=k}^{n-1} (n-j) \binom{n}{j} F(x)^j (1 - F(x))^{n-j-1} \right].$$

Notemos que, en los coeficiones binomiales se tiene lo siguiente:

$$j\binom{n}{j} = n\binom{n-1}{j-1}, \quad \text{y} \quad (n-j)\binom{n}{j} = n\binom{n-1}{j},$$

Así, podemos reescribir la expresión como:

$$= nf(x) \left[\sum_{j=k}^{n} \binom{n-1}{j-1} F(x)^{j-1}(1 - F(x))^{(n-1)-(j-1)} - \sum_{j=k}^{n-1} \binom{n-1}{j} F(x)^j (1 - F(x))^{(n-1)-j} \right].$$

Observamos que la primera sumatoria se puede descomponer como:

$$\begin{aligned} &\sum_{j=k}^{n} \binom{n-1}{j-1} F(x)^{j-1}(1 - F(x))^{(n-1)-(j-1)} \\ &= \binom{n-1}{k-1} F(x)^{k-1}(1 - F(x))^{n-k} + \sum_{j=k+1}^{n} \binom{n-1}{j-1} F(x)^{j-1}(1 - F(x))^{(n-1)-(j-1)}. \end{aligned}$$

Mientras que la segunda sumatoria se reescribe como:

$$\sum_{j=k}^{n-1} \binom{n-1}{j} F(x)^j (1-F(x))^{(n-1)-j} = \sum_{j=k+1}^{n} \binom{n-1}{j-1} F(x)^{j-1} (1-F(x))^{(n-1)-(j-1)}.$$

Vemos que las sumatorias restantes en ambas expresiones son idénticas y se cancelan entre sí, dejando únicamente:

$$f_{X_{(k)}}(x) = n\binom{n-1}{k-1} F(x)^{k-1} (1-F(x))^{n-k} f(x).$$

Finalmente, recordando que:

$$n\binom{n-1}{k-1} = \frac{n!}{(k-1)!(n-k)!},$$

Obtenemos la expresión final:

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} F(x)^{k-1} (1-F(x))^{n-k} f(x).$$

7.8.3 Función de Densidad Conjunta de los Estadísticos de Orden

Dado que el eje central de este capítulo es la distribución conjunta, resulta pertinente definir la distribución conjunta de los estadísticos de orden. Para ello, recordemos que si las variables aleatorias son independientes, su densidad conjunta se obtiene como el producto de sus densidades marginales.

En este caso, las variables originales $X_1, X_2, \ldots, X_n$ son independientes e idénticamente distribuidas (*i.i.d.*), por lo que su densidad conjunta es:

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = f(x_1) f(x_2) \cdots f(x_n).$$

Sin embargo, al considerar los estadísticos de orden $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$, la distinción entre las variables originales desaparece. Es decir, cualquier permutación de los valores iniciales genera la misma secuencia ordenada.

Dado que hay $n!$ formas de ordenar una muestra de n elementos, cada una de estas contribuye de manera equivalente a la probabilidad total de que los valores ordenados sean exactamente $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$.

Definición 7.33 (Densidad Conjunta de los Estadísticos de Orden)
La *densidad conjunta de los estadísticos de orden* se define como:

$$f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = n! f(x_1) f(x_2) \cdots f(x_n), \quad x_1 \leq x_2 \leq \cdots \leq x_n.$$

Esta expresión muestra que la densidad conjunta de los estadísticos de orden no solo depende de la densidad original $f(x)$, sino también del número de formas en que los valores de la muestra pueden organizarse. La restricción $x_1 \leq x_2 \leq \cdots \leq x_n$ es crucial, pues asegura que trabajamos con la versión ordenada de la muestra y no con cualquier permutación arbitraria.

7.8.4 Distribuciones del Mínimo y Máximo

Dado un conjunto de n variables aleatorias independientes e idénticamente distribuidas (i.i.d.) con función de distribución acumulada (CDF) $F(x)$ y función de densidad (PDF) $f(x)$, podemos determinar la distribución de los extremos de la muestra.

Distribución del Mínimo

El valor mínimo de la muestra, denotado como $X_{(1)}$, es el menor de los valores observados. Su función de distribución acumulada se obtiene considerando que $X_{(1)} \leq x$ si al menos un valor en la muestra es menor o igual que x, lo cual equivale al complemento del evento donde todos los valores son mayores que x:

$$F_{X_{(1)}}(x) = \mathbb{P}(X_{(1)} \leq x) = 1 - \mathbb{P}(X_1 > x, X_2 > x, \ldots, X_n > x).$$

Dado que las variables son independientes, la probabilidad conjunta es el producto de las probabilidades individuales:

$$F_{X_{(1)}}(x) = 1 - (1 - F(x))^n.$$

La función de densidad de $X_{(1)}$ se obtiene derivando su CDF:

$$f_{X_{(1)}}(x) = n(1 - F(x))^{n-1} f(x).$$

Distribución del Máximo

El valor máximo de la muestra, denotado como $X_{(n)}$, es el mayor de los valores observados. Su función de distribución acumulada se obtiene considerando que $X_{(n)} \leq x$ si y solo si todos los valores de la muestra son menores o iguales a x:

$$F_{X_{(n)}}(x) = \mathbb{P}(X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x).$$

Por independencia, esto es:

$$F_{X_{(n)}}(x) = F(x)^n.$$

La función de densidad del máximo se obtiene derivando la CDF:

$$f_{X_{(n)}}(x) = nF(x)^{n-1} f(x).$$

Capítulo 8

Propiedades de la Esperanza

El concepto de esperanza juega un papel esencial en la teoría de probabilidad, ya que nos permite describir el comportamiento promedio de una variable aleatoria. Sin embargo, su utilidad va mucho más allá de un simple promedio, en muchas aplicaciones, necesitamos entender cómo se comporta la esperanza cuando incorporamos nueva información, cuando combinamos variables aleatorias o cuando buscamos caracterizar una distribución de manera más profunda.

Por ejemplo, en pronósticos financieros, la esperanza condicional permite actualizar nuestras estimaciones con base en nueva información. En modelos de riesgo, la varianza total nos ayuda a distinguir entre incertidumbre inherente y variabilidad explicada por factores observables.

Además, descubriremos que la esperanza es una concepto que nos permite descifrar las distribuciones y sus relaciones. En particular, la función generadora de momentos (FGM) nos permitirá capturar información completa sobre una distribución a partir de sus momentos, lo que facilita el estudio de la suma de variables aleatorias y la identificación de familias de distribuciones con propiedades específicas.

8.1 Esperanza Condicional

En el capítulo anterior estudiamos la distribución conjunta de dos variables aleatorias y cómo la información sobre una de ellas afecta la otra a través de la probabilidad condicional. Ahora, nos interesa analizar cómo esta información influye en su valor esperado, lo que nos lleva al concepto de *esperanza condicional*.

Definición 8.1 (Esperanza Condicional) Dada una variable aleatoria X y otra variable Y, la esperanza condicional de X dado $Y = y$ representa el valor esperado de X bajo la condición de que Y ha tomado un valor específico. Su expresión depende del tipo de variables:

- Si X e Y son discretas:

$$\mathbb{E}[X \mid Y = y] = \sum_{x \in R_X} x \cdot P_{X|Y}(x \mid y).$$

- Si X e Y son continuas con densidad condicional $f_{X|Y}(x \mid y)$:

$$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x \mid y)\, dx.$$

La esperanza condicional extiende el concepto clásico de esperanza, ajustándolo en

función de la información adicional proporcionada por Y.

Propiedades

La esperanza condicional cumple las mismas propiedades de la esperanza tradicional, entre ellas:

- **Independencia:** Si X e Y son independientes, entonces:

$$\mathbb{E}[g(X) \mid Y] = \mathbb{E}[g(X)].$$

  Es decir, conocer Y no proporciona información adicional sobre X.

- **Linealidad:** Si $g(X)$ y $h(Y)$ son funciones, se cumple:

$$\mathbb{E}[g(X) \cdot h(Y) \mid Y] = h(Y) \cdot \mathbb{E}[g(X) \mid Y].$$

  Esto permite separar términos que dependen exclusivamente de Y, pues se consideran como constantes.

Ejemplo:

- **Caso Discreto:** Supongamos que el número de productos X comprados por un cliente depende del tipo de cliente Y, donde:
  - $Y = 0$ representa compradores ocasionales (60 % de los clientes).
  - $Y = 1$ representa compradores frecuentes (40 % de los clientes).

  Las distribuciones condicionales de X son:

X	0	1	2
$P(X \mid Y = 0)$	0.5	0.4	0.1
$P(X \mid Y = 1)$	0.2	0.3	0.5

  Calculamos la esperanza condicional de X dado que el cliente es ocasional ($Y = 0$):

$$\begin{aligned}\mathbb{E}[X \mid Y = 0] &= (0 \cdot 0{,}5) + (1 \cdot 0{,}4) + (2 \cdot 0{,}1) \\ &= 0 + 0{,}4 + 0{,}2 = 0{,}6.\end{aligned}$$

  Para los compradores frecuentes ($Y = 1$):

$$\begin{aligned}\mathbb{E}[X \mid Y = 1] &= (0 \cdot 0{,}2) + (1 \cdot 0{,}3) + (2 \cdot 0{,}5) \\ &= 0 + 0{,}3 + 1{,}0 = 1{,}3.\end{aligned}$$

  Estos valores representan el número esperado de productos comprados por cada tipo de cliente.

- **Caso Continuo:** Sea $X \mid Y = y$ una variable aleatoria con distribución normal:

$$X \mid Y = y \sim \mathcal{N}(y + 1,\, 4).$$

Es decir, dado $Y = y$, la densidad condicional de X es:

$$f_{X|Y}(x \mid y) = \frac{1}{\sqrt{8\pi}} \exp\left(-\frac{(x-(y+1))^2}{8}\right).$$

La esperanza condicional se obtiene como:

$$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x \mid y)\, dx.$$

Como $X \mid Y = y$ sigue una distribución normal $\mathcal{N}(y+1, 4)$, se conoce que su esperanza es el valor esperado de la normal:

$$\mathbb{E}[X \mid Y = y] = y + 1.$$

8.2 Varianza Condicional

En la sección anterior estudiamos la esperanza condicional como una forma de redefinir el valor esperado de una variable aleatoria cuando se conoce información sobre otra. Sin embargo, la esperanza no describe completamente la dispersión de una variable. Para comprender cómo varía X dado que Y ha tomado un valor específico, introducimos la *varianza condicional*.

Definición 8.2 (Varianza Condicional) Dada una variable aleatoria X y otra variable Y, la varianza condicional de X dado $Y = y$ mide la dispersión de X alrededor de su esperanza condicional. Se define como:

$$\mathrm{Var}(X \mid Y = y) = \mathbb{E}[(X - \mathbb{E}[X \mid Y = y])^2 \mid Y = y].$$

Al igual que la varianza clásica, esta cantidad nos permite cuantificar la incertidumbre de X, pero ahora considerando únicamente los valores de X dentro de cada subconjunto definido por Y.

Propiedades

La varianza condicional cumple con varias propiedades análogas a las ya vistas:

- **No negatividad:**

  $$\mathrm{Var}(X \mid Y) \geq 0.$$

  La varianza condicional siempre es no negativa, pues es una esperanza de un cuadrado.

- **Linealidad con escala:** Si a es una constante, entonces:

  $$\mathrm{Var}(aX \mid Y) = a^2 \,\mathrm{Var}(X \mid Y).$$

- **Independencia:** Si X e Y son independientes, entonces:

  $$\mathrm{Var}(X \mid Y) = \mathrm{Var}(X).$$

- **Relación con la Esperanza Condicional:** A partir de la definición, podemos expresar la varianza condicional como:

$$\mathrm{Var}(X \mid Y) = \mathbb{E}[X^2 \mid Y] - \mathbb{E}[X \mid Y]^2.$$

Ejemplo:

Usando los ejemplos anteriores:

- **Caso Discreto:** Recordemos que las esperanzas condicionales eran:

$$\mathbb{E}[X \mid Y = 0] = 0{,}6, \quad \mathbb{E}[X \mid Y = 1] = 1{,}3.$$

Ahora, calculamos la varianza condicional para $Y = 0$:

$$\begin{aligned}\mathrm{Var}(X \mid Y = 0) &= \sum_{x \in R_X} (x - 0{,}6)^2 P_{X|Y}(x \mid 0) \\ &= (0 - 0{,}6)^2(0{,}5) + (1 - 0{,}6)^2(0{,}4) + (2 - 0{,}6)^2(0{,}1) \\ &= (0{,}36)(0{,}5) + (0{,}16)(0{,}4) + (1{,}96)(0{,}1) \\ &= 0{,}18 + 0{,}064 + 0{,}196 = 0{,}44.\end{aligned}$$

Para $Y = 1$:

$$\begin{aligned}\mathrm{Var}(X \mid Y = 1) &= \sum_{x \in R_X} (x - 1{,}3)^2 P_{X|Y}(x \mid 1) \\ &= (0 - 1{,}3)^2(0{,}2) + (1 - 1{,}3)^2(0{,}3) + (2 - 1{,}3)^2(0{,}5) \\ &= (1{,}69)(0{,}2) + (0{,}09)(0{,}3) + (0{,}49)(0{,}5) \\ &= 0{,}338 + 0{,}027 + 0{,}245 = 0{,}61.\end{aligned}$$

- **Caso Continuo:** Para el caso en que $X \mid Y = y$ sigue una distribución normal:

$$X \mid Y = y \sim \mathcal{N}(y + 1, 4).$$

Como la varianza de una normal $\mathcal{N}(\mu, \sigma^2)$ es simplemente σ^2, tenemos:

$$\mathrm{Var}(X \mid Y = y) = 4.$$

8.3 Ley de la Esperanza Total

Al calcular la esperanza condicional de X dado $Y = y$, estamos fijando un valor específico de Y y evaluando el comportamiento de X bajo esa condición. Sin embargo, en la mayoría de los problemas, Y es una variable aleatoria cuyo valor aún no conocemos. En este sentido, la esperanza condicional $\mathbb{E}[X \mid Y]$ no es solo un número, sino una nueva variable aleatoria que depende de Y, y que toma dicho valor con probabilidad $\mathbb{P}(Y = y)$.

Por esto, para obtener la esperanza de X, debemos promediar la esperanza condicional $\mathbb{E}[X \mid Y]$ sobre todos los posibles valores de Y, por la probabilidad de cada uno de estos valores. Esto nos lleva a la *ley de la esperanza total*, la cual establece que la esperanza de X puede descomponerse en función de Y.

Teorema 8.1 Ley de la Esperanza Total

Sea X una variable aleatoria y Y otra variable aleatoria con función de probabilidad $P_Y(y)$ o densidad $f_Y(y)$. La esperanza total de X se expresa como:

$$\mathbb{E}[X] = \mathbb{E}_Y[\mathbb{E}_X[X|Y]]$$

Es decir:

- Si Y es una variable discreta con valores en R_Y:

$$\mathbb{E}[X] = \sum_{y \in R_Y} \mathbb{E}[X \mid Y = y] \cdot P_Y(y).$$

- Si Y es una variable continua con función de densidad $f_Y(y)$:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X \mid Y = y] \cdot f_Y(y)\, dy.$$

Demostración:

Demostraremos el resultado para los casos discreto y continuo por separado.

- **Caso Discreto:** Si Y es una variable discreta con valores en R_Y, la esperanza de X se define como:

$$\mathbb{E}[X] = \sum_{x \in R_X} x \cdot P_X(x).$$

Utilizando la probabilidad condicional $P_{X|Y}(x \mid y) = \frac{P_{X,Y}(x,y)}{P_Y(y)}$, podemos reescribir $P_X(x)$ en términos de Y:

$$P_X(x) = \sum_{y \in R_Y} P_{X|Y}(x \mid y) P_Y(y).$$

Sustituyendo esto en la expresión de la esperanza:

$$\mathbb{E}[X] = \sum_{x \in R_X} x \sum_{y \in R_Y} P_{X|Y}(x \mid y) P_Y(y).$$

Cambiando el orden de la sumatoria:

$$\mathbb{E}[X] = \sum_{y \in R_Y} P_Y(y) \sum_{x \in R_X} x P_{X|Y}(x \mid y).$$

Reconociendo la esperanza condicional:

$$\mathbb{E}[X] = \sum_{y \in R_Y} \mathbb{E}[X \mid Y = y] P_Y(y) = \mathbb{E}_Y[\mathbb{E}_X[X|Y]].$$

- **Caso Continuo:** Si X y Y son variables continuas, partimos de la definición de esperanza:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx.$$

Utilizando la densidad condicional $f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$, podemos escribir $f_X(x)$ en términos de Y:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x \mid y) f_Y(y)\, dy.$$

Sustituyendo en la expresión de la esperanza:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{X|Y}(x \mid y) f_Y(y)\, dy \right) dx.$$

Intercambiando el orden de integración:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} f_Y(y) \left(\int_{-\infty}^{\infty} x f_{X|Y}(x \mid y)\, dx \right) dy.$$

Reconociendo la esperanza condicional:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X \mid Y = y] f_Y(y)\, dy = \mathbb{E}_Y[\mathbb{E}_X[X|Y]].$$

Ejemplo:

- **Caso Discreto:** Supongamos que en un almacén, un cliente puede realizar una compra con una probabilidad de 0,7 o salir sin comprar con probabilidad 0,3. Definimos la variable aleatoria X como el monto gastado por el cliente, y Y como un indicador de si realizó una compra:

$$Y = \begin{cases} 1, & \text{si compra (con probabilidad 0,7)} \\ 0, & \text{si no compra (con probabilidad 0,3)} \end{cases}$$

Las distribuciones condicionales de X son:

$$\mathbb{E}[X \mid Y = 1] = 50, \quad \text{(si compra, gasta en promedio 50)}$$
$$\mathbb{E}[X \mid Y = 0] = 0, \quad \text{(si no compra, gasta 0).}$$

Aplicando la ley de la esperanza total:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{y \in R_Y} \mathbb{E}[X \mid Y = y] \cdot P_Y(y) \\ &= (50 \cdot 0{,}7) + (0 \cdot 0{,}3) = 35. \end{aligned}$$

Esto indica que el gasto promedio por cliente, considerando que algunos no compran, es de 35.

- **Caso Continuo:** Supongamos que el tiempo de vida (en años) de una batería sigue una distribución condicional normal:

$$X \mid Y = y \sim \mathcal{N}(y, 1).$$

donde Y representa una característica del material de la batería y sigue una distribución uniforme $Y \sim U(4, 6)$.

La esperanza condicional es:

$$\mathbb{E}[X \mid Y = y] = y.$$

Aplicando la ley de la esperanza total:

$$\begin{aligned}\mathbb{E}[X] &= \int_4^6 \mathbb{E}[X \mid Y = y] f_Y(y)\, dy \\ &= \int_4^6 y \cdot \frac{1}{2}\, dy \\ &= \frac{1}{2}\int_4^6 y\, dy.\end{aligned}$$

Resolviendo la integral:

$$\begin{aligned}\frac{1}{2}\left[\frac{y^2}{2}\right]_4^6 &= \frac{1}{2}\left(\frac{6^2}{2} - \frac{4^2}{2}\right) \\ &= \frac{1}{2}\left(\frac{36}{2} - \frac{16}{2}\right) = \frac{1}{2}\,(18 - 8) = \frac{1}{2}(10) = 5.\end{aligned}$$

Por lo tanto, la vida media de la batería, considerando la variabilidad en el material, es de 5 años.

8.4 Ley de la Varianza Total

Al igual que la esperanza, al condicionar X sobre una variable aleatoria Y, la varianza de X sigue siendo una variable aleatoria. Sin embargo, debido a la naturaleza cuadrática de la varianza, su descomposición no se obtiene de la misma manera que en la ley de la esperanza total. En lugar de una simple combinación lineal, la variabilidad total de X se divide en dos componentes, una que mide la dispersión de X dentro de los subconjuntos definidos por Y y otra que refleja la variabilidad entre estos subconjuntos. Esta descomposición es lo que conocemos como la *ley de la varianza total*.

Teorema 8.2 Ley de la Varianza Total

Sea X una variable aleatoria y Y otra variable aleatoria con la que puede estar relacionada. Entonces, la varianza de X satisface:

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathbb{E}[X \mid Y]).$$

El termino $\mathbb{E}[\mathrm{Var}(X \mid Y)]$ mide la variabilidad de X dentro de los subconjuntos definidos por Y, es decir, cuánta dispersión existe en X cuando Y es conocido. Mientras que,$\mathrm{Var}(\mathbb{E}[X \mid Y])$ captura la variabilidad de los valores esperados condicionales de X, es decir, cuán dispersa está la media condicional $\mathbb{E}[X \mid Y]$ entre distintos valores de Y.

Demostración: Sea la definición de varianza en términos de esperanza:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Utilizando la ley de la esperanza total:

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]], \quad \mathbb{E}[X^2] = \mathbb{E}[\mathbb{E}[X^2 \mid Y]].$$

Expresamos $\mathbb{E}[X^2 \mid Y]$ en términos de varianza y esperanza condicional:

$$\mathbb{E}[X^2 \mid Y] = \mathrm{Var}(X \mid Y) + \mathbb{E}[X \mid Y]^2.$$

Tomando esperanza en ambos lados:

$$\mathbb{E}[X^2] = \mathbb{E}[\mathrm{Var}(X \mid Y)] + \mathbb{E}[(\mathbb{E}[X \mid Y])^2].$$

Sustituyéndolo en la ecuación original de la varianza:

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[\mathrm{Var}(X \mid Y)] + \mathbb{E}[(\mathbb{E}[X \mid Y])^2] - (\mathbb{E}[X])^2 \\ &= \mathbb{E}[\mathrm{Var}(X \mid Y)] + \mathbb{E}[(\mathbb{E}[X \mid Y])^2] - \mathbb{E}[\mathbb{E}[X \mid Y]]^2 \\ &= \mathbb{E}[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathbb{E}[X \mid Y]).\end{aligned}$$

Ejemplo:

Utilizando los mismos ejemplos de la Ley de la Esperanza Total, calcularemos la varianza total de X aplicando la descomposición en términos de la varianza condicional y la varianza de la esperanza condicional.

- **Caso Discreto:** Recordemos que en el problema del gasto en un almacén, tenemos:

$$\begin{aligned}\mathbb{E}[X \mid Y = 1] &= 50, & \mathbb{E}[X \mid Y = 0] &= 0, \\ \mathrm{Var}(X \mid Y = 1) &= 100, & \mathrm{Var}(X \mid Y = 0) &= 0.\end{aligned}$$

  Calculamos la esperanza de X, que ya obtuvimos como:

$$\mathbb{E}[X] = 35.$$

  Ahora, calculamos la esperanza de la varianza condicional:

$$\begin{aligned}\mathbb{E}[\mathrm{Var}(X \mid Y)] &= \sum_{y \in R_Y} \mathrm{Var}(X \mid Y = y) P_Y(y) \\ &= (100 \cdot 0{,}7) + (0 \cdot 0{,}3) = 70.\end{aligned}$$

  Luego, calculamos la varianza de la esperanza condicional:

$$\begin{aligned}\mathrm{Var}(\mathbb{E}[X \mid Y]) &= \sum_{y \in R_Y} (\mathbb{E}[X \mid Y = y] - \mathbb{E}[X])^2 P_Y(y) \\ &= (50 - 35)^2 \cdot 0{,}7 + (0 - 35)^2 \cdot 0{,}3 \\ &= (15)^2 \cdot 0{,}7 + (-35)^2 \cdot 0{,}3 \\ &= 225 \cdot 0{,}7 + 1225 \cdot 0{,}3 = 157{,}5 + 367{,}5 = 525.\end{aligned}$$

  Aplicando la Ley de la Varianza Total:

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathbb{E}[X \mid Y]) = 70 + 525 = 595.$$

- **Caso Continuo:** En el ejemplo de la duración de las baterías, sabemos que $X \mid Y = y \sim \mathcal{N}(y, 1)$, por lo que su varianza condicional es:

$$\text{Var}(X \mid Y = y) = 1.$$

Como $Y \sim U(4, 6)$, la esperanza de la varianza condicional es:

$$\mathbb{E}[\text{Var}(X \mid Y)] = \int_4^6 1 \cdot \frac{1}{2}\, dy = 1.$$

Ahora, calculamos la varianza de la esperanza condicional. Sabemos que:

$$\mathbb{E}[X \mid Y = y] = y.$$

Y ya calculamos que:

$$\mathbb{E}[X] = 5.$$

La varianza de Y en un intervalo uniforme es:

$$\text{Var}(Y) = \frac{(b-a)^2}{12} = \frac{(6-4)^2}{12} = \frac{4}{12} = \frac{1}{3}.$$

Como $\mathbb{E}[X \mid Y] = Y$, se cumple que:

$$\text{Var}(\mathbb{E}[X \mid Y]) = \text{Var}(Y) = \frac{1}{3}.$$

Aplicando la Ley de la Varianza Total:

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X \mid Y)] + \text{Var}(\mathbb{E}[X \mid Y]) = 1 + \frac{1}{3} = \frac{4}{3}.$$

Por lo tanto, la varianza total de la vida útil de las baterías es $\frac{4}{3}$.

8.5 Suma de Variables Aleatorias

La suma de variables aleatorias es un concepto utilizado para modelar la combinación de múltiples influencias en un fenómeno. En muchas aplicaciones, es común que una cantidad de interés resulte de la acumulación de efectos individuales, como la suma de tiempos de espera en un sistema, la agregación de ingresos en un periodo de tiempo o la combinación de errores en mediciones sucesivas. Formalmente:

Definición 8.3 (Suma de Variables Aleatorias) Si Y representa la suma de n variables aleatorias $X_1, X_2, \ldots, X_n$, se define como:

$$Y = X_1 + X_2 + \cdots + X_n,$$

en esta sección analizaremos sus propiedades, incluyendo su esperanza, varianza y distribución.

8.5.1 Esperanza de la Suma

Anteriormente, definimos la esperanza de una función $g : \mathbb{R}^n \to \mathbb{R}$ aplicada a un conjunto de variables aleatorias. Un caso particular de interés ocurre cuando la función es simplemente la suma de estas variables.

Esto puede ser visto directamente desde la aplicación de la definición general de esperanza para funciones de variables aleatorias conjuntas. Sin embargo, si conocemos la distribución marginal de cada una de estas, para efectos prácticos, es más sencillo aplicar la linealidad de la esperanza, ya que el cálculo de la esperanza de una suma de variables aleatorias se comporta de manera aditiva independientemente de la relación entre las variables.

Teorema 8.3 Esperanza de la Suma

Para cualquier conjunto de n variables aleatorias $X_1, X_2, \ldots, X_n$, se cumple:

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

Demostración: Esta propiedad se sigue directamente de la linealidad de la esperanza, que establece que la esperanza de una combinación lineal de variables aleatorias es igual a la combinación lineal de sus esperanzas:

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

Esta propiedad se debe a la linealidad de la esperanza, lo que implica que la fórmula es válida independientemente de si las variables están correlacionadas o no.

8.5.2 Varianza de la Suma

La varianza en el caso de la suma de variables aleatorias, no solo depende de la varianza individual de cada término, sino también de cómo estas variables están relacionadas entre sí a través de su covarianza.

Teorema 8.4 Varianza de la Suma

Sea $Y = \sum_{i=1}^{n} X_i$ la suma de n variables aleatorias, entonces su varianza está dada por:

$$\mathrm{Var}(Y) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j).$$

Si las variables X_i son independientes, entonces $\mathrm{Cov}(X_i, X_j) = 0$ para $i \neq j$, lo que simplifica la expresión a:

$$\mathrm{Var}(Y) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Demostración: Por la definición de varianza, tenemos:

$$\text{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2.$$

Sustituyendo $Y = \sum_{i=1}^{n} X_i$, se obtiene:

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] - \left(\mathbb{E}\left[\sum_{i=1}^{n} X_i\right]\right)^2.$$

Expandiendo el cuadrado:

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \mathbb{E}\left[\sum_{i=1}^{n} X_i^2 + 2\sum_{i<j} X_i X_j\right] - \sum_{i=1}^{n} \mathbb{E}[X_i]^2.$$

Usando la linealidad de la esperanza:

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}[X_i^2] + 2\sum_{i<j} \mathbb{E}[X_i X_j] - \sum_{i=1}^{n} (\mathbb{E}[X_i])^2 - 2\sum_{i<j} \mathbb{E}[X_i]\mathbb{E}[X_j].$$

Observamos que los términos de la forma $\mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2$ corresponden a $\text{Var}(X_i)$, mientras que la diferencia en los términos cruzados da lugar a la covarianza:

$$\text{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j].$$

Sustituyendo en la ecuación obtenemos:

$$\text{Var}(Y) = \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum_{i<j} \text{Cov}(X_i, X_j).$$

Finalmente, si las variables X_i son independientes, entonces $\text{Cov}(X_i, X_j) = 0$ para $i \neq j$, lo que reduce la expresión a:

$$\text{Var}(Y) = \sum_{i=1}^{n} \text{Var}(X_i).$$

8.5.3 Distribución de la Suma de Variables Aleatorias Independientes

Al estudiar la suma de variables aleatorias, un caso de interés particular es cuando estas son independientes. La independencia permite que la distribución conjunta se exprese como el producto de sus distribuciones marginales, lo que facilita el análisis matemático y la obtención de resultados explícitos. Además, muchas propiedades fundamentales de las distribuciones de probabilidad, como la estabilidad bajo suma en ciertas familias de distribuciones, dependen de esta condición.

Definición 8.4 (Distribución de la Suma para Variables Independientes) Sea $Y = \sum_{i=1}^{n} X_i$ la suma de n variables aleatorias independientes. Su distribución se determina según el tipo de variables:

- **Caso Discreto:** Si X_1 y X_2 son variables aleatorias discretas independientes, la función de masa de probabilidad (PMF) de Y se obtiene mediante:

$$P_Y(y) = \sum_{x_2 \in R_{X_2}} P_{X_1}(y - x_2) \cdot P_{X_2}(x_2).$$

- **Caso Continuo:** Si X_1 y X_2 son variables aleatorias continuas independientes, la función de densidad de probabilidad (PDF) de Y se obtiene mediante la convolución:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(y - x_2) f_{X_2}(x_2)\, dx_2.$$

  Esta operación, es conocida como *convolución de densidades*, la cúal permite calcular la distribución de la suma a partir de las distribuciones individuales, es decir:

$$f_Y(y) = (f_{X_1} * f_{X_2})(y).$$

8.5.4 Distribuciones Resultantes de Sumas

Para algunas familias de distribuciones, la suma de variables aleatorias independientes de esa familia produce una nueva variable de la misma familia con distintos parámetros en función de los ya conocidos. Algunos casos comunes son:

- $\text{Poisson}(\lambda_1) + \text{Poisson}(\lambda_2) = \text{Poisson}(\lambda_1 + \lambda_2)$.
- $\text{Binomial}(n, p) + \text{Binomial}(m, p) = \text{Binomial}(n + m, p)$.
- $\text{Gamma}(s, \lambda) + \text{Gamma}(t, \lambda) = \text{Gamma}(s + t, \lambda)$.
- $\text{Exponencial}(\lambda) + \text{Exponencial}(\lambda) = \text{Gamma}(2, \lambda)$.
- $\text{Normal}(\mu_1, \sigma_1^2) + \text{Normal}(\mu_2, \sigma_2^2) = \text{Normal}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

El estudio de la suma de variables aleatorias tiene diversas aplicaciones. Por ejemplo, en teoría de colas, que se verá en cursos como *Decisiones Bajo Incertidumbre* o *Gestión de Operaciones*, los tiempos de espera acumulados en sistemas de servicio pueden modelarse mediante la suma de variables aleatorias independientes. Un aspecto interesante de estas sumas es que, en muchos casos, su distribución sigue perteneciendo a la misma familia que las variables originales, como ocurre con las distribuciones ejemplificadas anteriormente. Aunque podríamos demostrar esto directamente a partir de las definiciones, resulta más conveniente introducir un nuevo concepto, la *función generadora de momentos* (FGM).

8.6 Función Generadora de Momentos

El estudio de distribuciones de variables aleatorias requiere herramientas que permitan caracterizarlas completamente. Una de estas herramientas es la *función generadora de momentos* (MGF, por sus siglas en inglés), que encapsula en una única expresión toda la información sobre los momentos de una distribución. Su utilidad se extiende al cálculo de momentos, la identificación de distribuciones y el análisis de combinaciones de variables aleatorias.

8.6.1 Momentos

Los *momentos* de una variable aleatoria describen características de su distribución, como su tendencia central, dispersión y forma. Matemáticamente, el momento de orden k de una variable aleatoria X respecto al origen se define como:

$$\mathbb{E}[X^k]$$

Los momentos describen distintas propiedades de la distribución:

- El *momento de primer orden*, $\mathbb{E}[X]$, es la *media*.
- El *momento de segundo orden*, $\mathbb{E}[X^2]$, ayuda a cuantificar la dispersión de los valores alrededor del origen.

Además de los momentos respecto al origen, existen los *momentos centrales*, definidos en torno a la media de la variable aleatoria:

$$\mathbb{E}[(X - \mathbb{E}[X])^k].$$

Estos momentos capturan la variabilidad relativa a la media en lugar del origen:

- El *momento central de orden 2* es la *varianza*, $\mathbb{E}[(X - \mathbb{E}[X])^2]$, que mide la dispersión respecto a la media.
- El *momento central de orden 3* mide la **asimetría**, indicando si la distribución tiene una cola más pesada en una dirección.
- El *momento central de orden 4* mide la *curtosis*.

Así como hemos definido los momentos y momentos centrales, necesitamos una herramienta que no solo permita calcular los momentos de una variable aleatoria, sino que también proporcione características de su distribución para facilitar su identificación. Una función especialmente útil para este propósito es la exponencial, ya que su expansión en serie de Taylor permite expresar los momentos de manera natural.

Definición 8.5 (Función Generadora de Momentos) Sea X una variable aleatoria. La *función generadora de momentos (FGM)* de X se define como:

$$M_X(s) = \mathbb{E}[e^{sX}],$$

Decimos que la FGM de X existe si existe una constante positiva $a > 0$ tal que $M_X(s)$ es finita para todo $s \in [-a, a]$.

Así mismo, para un vector aleatorio $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, la función generadora de momentos (FGM) se define como:

$$M_{\mathbf{X}}(\mathbf{s}) = \mathbb{E}[e^{\mathbf{s}^\top \mathbf{X}}],$$

donde $\mathbf{s} = (s_1, s_2, \ldots, s_n)^\top$ es un vector de parámetros reales, por ende, $\mathbf{s}^\top \mathbf{X} = s_1X_1 + s_2X_2 + \cdots + s_nX_n$ es el producto interno, y la esperanza se toma con respecto a la distribución conjunta de $\mathbf{X}$.

Para un vector aleatorio, la FGM existe si $M_{\mathbf{X}}(\mathbf{s}) < \infty$ en alguna vecindad del origen $\mathbf{s} = \mathbf{0}$, es decir, si existe una constante $a > 0$ tal que $M_{\mathbf{X}}(\mathbf{s})$ es finita para todo $\mathbf{s}$ con $\|\mathbf{s}\| < a$. Esta función al igual que en su forma de una variable aleatoria, encapsula los momentos conjuntos de las componentes de $\mathbf{X}$, y si existe, caracteriza completamente la distribución conjunta del vector aleatorio.

Ejemplo:

Consideremos una variable aleatoria $X \sim \text{Exp}(\lambda)$ con parámetro $\lambda > 0$. Su función de densidad de probabilidad (PDF) es:

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

La función generadora de momentos $M_X(s)$ se define como:

$$M_X(s) = \mathbb{E}[e^{sX}] = \int_0^\infty e^{sx} \lambda e^{-\lambda x}\, dx.$$

Resolviendo la integral:

$$M_X(s) = \lambda \int_0^\infty e^{(s-\lambda)x} dx = \frac{\lambda}{s-\lambda} \cdot (e^{(s-\lambda)x}\Big|_0^\infty).$$

Para que la integral converja, necesitamos que $s - \lambda < 0$, es decir, $s < \lambda$. Evaluando la integral en este dominio:

$$M_X(s) = \frac{\lambda}{\lambda - s}, \quad s < \lambda.$$

Es decir, la FGM de una variable exponencial existe para $s < \lambda$.

8.6.2 Cálculo de Momentos a partir de la FGM

Como se mencionó anteriormente, la utilidad de utilizar la función exponencial en la definición de la función generadora de momentos radica en su expansión en serie de Taylor, lo que permite expresar los momentos de una distribución de manera natural a través de sus derivadas.

Proposición 8.1

La función generadora de momentos $M_X(s)$ de una variable aleatoria X puede expresarse como:

$$M_X(s) = \sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k]s^k}{k!}.$$

A partir de esta serie, los momentos de X se obtienen derivando $M_X(s)$ k-veces y evaluando en $s = 0$:

$$\mathbb{E}[X^k] = \frac{d^k}{ds^k} M_X(s)\Big|_{s=0}.$$

Demostración:

- **FGM en Función de Sumatoria:** Expandiendo la función exponencial en serie de Taylor:

  $$e^{sX} = \sum_{k=0}^{\infty} \frac{(sX)^k}{k!},$$

  y tomando la esperanza en ambos lados, obtenemos:

  $$M_X(s) = \mathbb{E}[e^{sX}] = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{(sX)^k}{k!}\right].$$

  Intercambiando la esperanza con la suma infinita, siempre que la serie converja absolutamente:

  $$M_X(s) = \sum_{k=0}^{\infty} \frac{s^k}{k!}\mathbb{E}[X^k].$$

- **Momentos a través de la FGM:** Dado que hemos obtenido la expansión en serie de Taylor de la función generadora de momentos:

  $$M_X(s) = \sum_{m=0}^{\infty} \frac{\mathbb{E}[X^m]}{m!}s^m,$$

  notamos que cada coeficiente de s^m en la serie corresponde a $\mathbb{E}[X^m]/m!$. Para extraer el k-ésimo momento, derivamos $M_X(s)$ k-veces con respecto a s:

  $$\begin{aligned}\frac{d^k}{ds^k}M_X(s) &= \sum_{m=k}^{\infty} \frac{\mathbb{E}[X^m]}{m!} \cdot m(m-1)\dots(m-k+1)s^{m-k} \\ &= \sum_{m=k}^{\infty} \frac{\mathbb{E}[X^m]}{(m-k)!}s^{m-k} \\ &= \mathbb{E}[X^k] + \sum_{m=k+1}^{\infty} \frac{\mathbb{E}[X^m]}{(m-k)!}s^{m-k}.\end{aligned}$$

  Al derivar k-veces, cada término de la forma s^m en la serie original genera un factor $m(m-1)\dots(m-k+1)$, que es precisamente la notación factorial reducida de los primeros k términos de $m!$, es decir, $m!/(m-k)!$. Los términos con

$m < k$ desaparecen porque, al derivar más veces de las que tienen exponentes de s, se transforman en constantes que se anulan. Evaluando en $s = 0$, todos los términos con s^{m-k} se anulan para $m > k$, dejando únicamente el término con $m = k$, el cúal fue extraído de la sumatoria, obteniendo así:

$$\mathbb{E}[X^k] = \frac{d^k}{ds^k} M_X(s)\Big|_{s=0}.$$

8.6.3 Unicidad de la FGM

Como hemos discutido previamente, una de las propiedades de la función generadora de momentos es su capacidad para caracterizar completamente una distribución. Intuitivamente, la FGM condensa toda la información sobre sus momentos en una única expresión.

El resultado se basa en la unicidad de la transformada de Laplace y en la expansión en serie de Taylor de la función generadora de momentos. A nivel de pregrado, suele aceptarse sin prueba formal debido a que una demostración rigurosa requiere conocimientos de análisis complejo, como el teorema de identidad para funciones holomorfas. En su lugar, se recomienda trabajar con la *función característica* la cúal será introducida en la siguiente sección, cuya unicidad se puede establecer mediante la fórmula de inversión y el teorema de Weierstrass. Sin embargo, la demostración se puede encontrar en *Feller (An Introduction to Probability Theory and Its Applications, Vol. 2)*

Teorema 8.5 Unicidad de la Función Generadora de Momentos

Sean X y Y dos variables aleatorias. Supongamos que existe una constante positiva c tal que las funciones generadoras de momentos de X y Y son finitas e idénticas para todo $s \in [-c, c]$, es decir,

$$M_X(s) = M_Y(s), \quad \forall s \in [-c, c].$$

Entonces, X e Y tienen la misma función de distribución:

$$F_X(x) = F_Y(x), \quad \forall x \in \mathbb{R}.$$

Demostración: Este resultado es un caso particular de la unicidad de la función generadora de momentos para variables aleatorias discretas con rango finito.

Supongamos que X e Y son variables aleatorias discretas que toman valores en $\{0, 1, 2, \ldots, n\}$ y que sus funciones generadoras de momentos son idénticas para todos los valores de s. Entonces, tenemos:

$$\sum_{x=0}^{n} e^{sx} P_X(x) = \sum_{y=0}^{n} e^{sy} P_Y(y).$$

Definiendo $s = e^t$ y los coeficientes $c_i = P_X(i) - P_Y(i)$ para $i = 0, 1, \ldots, n$, podemos reescribir la ecuación como:

$$\sum_{x=0}^{n} s^x P_X(x) - \sum_{y=0}^{n} s^y P_Y(y) = 0.$$

Esto nos lleva a:

$$\sum_{x=0}^{n} s^x (P_X(x) - P_Y(x)) = 0.$$

Dado que la expresión es un polinomio en s con coeficientes $c_0, c_1, \ldots, c_n$, y que un polinomio solo puede ser idénticamente cero para todo s si y solo si todos sus coeficientes son nulos, concluimos que $c_i = 0$ para todo i, es decir:

$$P_X(i) = P_Y(i), \quad \forall i = 0, 1, \ldots, n.$$

Esto implica que las funciones de probabilidad de X e Y son iguales, y por lo tanto tienen la misma distribución.

8.6.4 FGM de una Suma de Variables Independientes

Otra de las propiedades de la función generadora de momentos es su comportamiento bajo sumas de variables aleatorias independientes. Recordemos que la FGM encapsula la información de todos los momentos de una variable aleatoria en una sola función. Esto nos permite analizar cómo se comportan las distribuciones cuando sumamos variables aleatorias.

En particular, si dos variables aleatorias X_1 y X_2 son independientes, sus distribuciones conjuntas pueden factorizarse en el producto de sus distribuciones marginales. Esta propiedad se refleja en la FGM, lo que nos lleva al siguiente resultado.

Teorema 8.6 FGM de una Suma de Variables Independientes

Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes con funciones generadoras de momentos $M_{X_i}(s)$, entonces la FGM de su suma es el producto de sus FGM's individuales:

$$M_{X_1+\cdots+X_n}(s) = M_{X_1}(s) \cdot M_{X_2}(s) \cdots M_{X_n}(s).$$

Demostración: Por definición, la FGM de $Y = X_1 + X_2 + \cdots + X_n$ es:

$$M_Y(s) = \mathbb{E}[e^{sY}] = \mathbb{E}[e^{s(X_1+X_2+\cdots+X_n)}].$$

Usando la propiedad de los exponentes:

$$M_Y(s) = \mathbb{E}[e^{sX_1} e^{sX_2} \cdots e^{sX_n}].$$

Dado que las variables $X_1, X_2, \ldots, X_n$ son independientes, la esperanza de un producto es el producto de las esperanzas:

$$M_Y(s) = \mathbb{E}[e^{sX_1}] \cdot \mathbb{E}[e^{sX_2}] \ldots \mathbb{E}[e^{sX_n}].$$

Reconociendo que cada término en el producto es la FGM correspondiente de cada X_i, obtenemos:

$$M_Y(s) = M_{X_1}(s) \cdot M_{X_2}(s) \cdots M_{X_n}(s).$$

Ejemplo:

Sean X_1 y X_2 dos variables aleatorias binomiales independientes, tales que $X_1 \sim$

$\text{Bin}(n,p)$ y $X_2 \sim \text{Bin}(m,p)$. Queremos demostrar que la suma $Y = X_1 + X_2$ sigue una distribución binomial $Y \sim \text{Bin}(n+m,p)$, tal como se indico en la sección de "Suma de Variables Aleatorias Independientes", utilizando la función generadora de momentos. Entonces, por definición la FGM de una variable binomial $X \sim \text{Bin}(n,p)$, es:

$$M_X(s) = \mathbb{E}[e^{sX}] = \sum_{k=0}^{n} e^{sk}\mathbb{P}(X=k) = \sum_{k=0}^{n} e^{sk}\binom{n}{k}p^k(1-p)^{n-k},$$

Factorizamos y agrupamos los términos tal que, $e^{sk}p^k = (e^s p)^k$, además, por el teorema del binomio $(a+b)^n = \sum_{i=0}^{n} a^i b^{n-i}$, así se obtiene que:

$$M_X(s) = \sum_{k=0}^{n}\binom{n}{k}(e^s p)^k(1-p)^{n-k} = (e^s p + 1 - p)^n,$$

y existe para todo $s \in \mathbb{R}$, ya que el rango de X es finito.

Aplicando este resultado a nuestras variables, tenemos que para $X_1 \sim \text{Bin}(n,p)$, la FGM es:

$$M_{X_1}(s) = (1 - p + pe^s)^n,$$

y para $X_2 \sim \text{Bin}(m,p)$:

$$M_{X_2}(s) = (1 - p + pe^s)^m.$$

Dado que X_1 y X_2 son independientes, por el teorema de la FGM de una suma de variables independientes, la FGM de $Y = X_1 + X_2$ es:

$$M_Y(s) = M_{X_1}(s) \cdot M_{X_2}(s) = (1 - p + pe^s)^n \cdot (1 - p + pe^s)^m = (1 - p + pe^s)^{n+m}.$$

Esta expresión coincide exactamente con la FGM de una variable binomial $\text{Bin}(n+m,p)$, y por el teorema de unicidad de la FGM, si $M_Y(s) = M_{\text{Bin}(n+m,p)}(s)$ para todo s, entonces:

$$Y = X_1 + X_2 \sim \text{Bin}(n+m,p).$$

8.6.5 Funciones Generadoras de Momentos Conocidas

A continuación, A continuación, se listan las funciones generadoras de momentos (FGM) de algunas distribuciones conocidas. Para cada distribución, se especifican los parámetros, la expresión de la FGM y el dominio de s en el que está definida.

- **Binomial:** Sea $X \sim \text{Bin}(n,p)$, con $n \in \mathbb{N}$ y $0 < p < 1$, entonces su FGM esta dada por:
$$M_X(s) = (1 - p + pe^s)^n, \quad s \in \mathbb{R}.$$

- **Poisson:** Sea $X \sim \text{Poisson}(\lambda)$, con $\lambda > 0$, entonces su FGM esta dada por:
$$M_X(s) = e^{\lambda(e^s - 1)}, \quad s \in \mathbb{R}.$$

- **Geométrica:** Sea $X \sim \text{Geom}(p)$, con $0 < p < 1$, entonces su FGM esta dada por:
$$M_X(s) = \frac{pe^s}{1 - (1-p)e^s}, \quad s < -\ln(1-p).$$

- **Exponencial:** Sea $X \sim \text{Exp}(\lambda)$, con $\lambda > 0$, , entonces su FGM esta dada por:

$$M_X(s) = \frac{\lambda}{\lambda - s}, \quad s < \lambda.$$

- **Normal:** Sea $X \sim N(\mu, \sigma^2)$, con $\mu \in \mathbb{R}$ y $\sigma^2 > 0$, , entonces su FGM esta dada por:

$$M_X(s) = e^{\mu s + \frac{\sigma^2 s^2}{2}}, \quad s \in \mathbb{R}.$$

- **Uniforme:** Sea $X \sim \text{Unif}(a, b)$, con $a < b$, entonces su FGM esta dada por:

$$M_X(s) = \frac{e^{sb} - e^{sa}}{s(b-a)}, \quad s \in \mathbb{R} \text{ (con } s \neq 0).$$

- **Gamma:** Sea $X \sim \text{Gamma}(\alpha, \beta)$, con $\alpha > 0$ y $\beta > 0$, entonces su FGM esta dada por:

$$M_X(s) = \left(1 - \frac{s}{\beta}\right)^{-\alpha}, \quad s < \beta.$$

8.7 Función Característica

El análisis de las distribuciones de variables aleatorias requiere herramientas que permitan describir sus propiedades de manera única y completa, otra de estas herramientas, es la *función característica*. A diferencia de la función generadora de momentos, la función característica siempre existe.

La *función característica* de una variable aleatoria captura su comportamiento a través de la transformada de Fourier de su distribución. Esta función no depende de la existencia de momentos finitos, lo que la hace aplicable incluso a distribuciones con colas pesadas o momentos indefinidos.

Definición 8.6 (Función Característica) Sea X una variable aleatoria. La *función característica* de X se define como:

$$\phi_X(\omega) = \mathbb{E}[e^{i\omega X}],$$

donde $i = \sqrt{-1}$ es la unidad imaginaria y $\omega \in \mathbb{R}$. La función característica existe para toda variable aleatoria X, ya que $|e^{i\omega X}| = 1$, lo que asegura que la esperanza siempre esté bien definida.

Así mismo, para un vector aleatorio $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, la función característica se define como:

$$\phi_{\mathbf{X}}(\omega) = \mathbb{E}[e^{i\omega^\top \mathbf{X}}],$$

donde $\omega = (\omega_1, \omega_2, \ldots, \omega_n)^\top$ es un vector de parámetros reales, $i = \sqrt{-1}$ es la unidad imaginaria, $\omega^\top \mathbf{X} = \omega_1 X_1 + \omega_2 X_2 + \cdots + \omega_n X_n$ es el producto interno, y la esperanza se toma con respecto a la distribución conjunta de $\mathbf{X}$.

La función característica siempre existe para todo $\omega \in \mathbb{R}^n$, ya que $|e^{i\omega^\top \mathbf{X}}| = 1$, y proporciona una descripción completa de la distribución conjunta del vector aleatorio mediante su unicidad.

Ejemplo:

Consideremos una variable aleatoria $X \sim \text{Poisson}(\lambda)$ con parámetro $\lambda > 0$. La función característica $\phi_X(\omega)$ se calcula como:

$$\phi_X(\omega) = \mathbb{E}[e^{i\omega X}] = \sum_{k=0}^{\infty} e^{i\omega k} \cdot \mathbb{P}(X = k) = \sum_{k=0}^{\infty} e^{i\omega k} \frac{\lambda^k e^{-\lambda}}{k!}.$$

Factorizando $e^{-\lambda}$ fuera de la suma y reorganizando:

$$\phi_X(\omega) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^{i\omega}\lambda)^k}{k!}.$$

Reconocemos que la suma es la serie de Taylor de la función exponencial e^x evaluada en $x = e^{i\omega}\lambda$. Por lo tanto:

$$\phi_X(\omega) = e^{-\lambda} e^{e^{i\omega}\lambda} = e^{\lambda(e^{i\omega}-1)}.$$

8.7.1 Cálculo de Momentos a partir de la FGM

Al igual que la función generadora de momentos, la función característica permite obtener los momentos de una variable aleatoria (si existen) mediante sus derivadas. Esto se deriva de la expansión en serie de Taylor de $e^{i\omega X}$.

Proposición 8.2

Si la variable aleatoria X tiene momentos finitos hasta el orden k, entonces la función característica $\phi_X(\omega)$ es k veces diferenciable en $t = 0$, y:

$$\mathbb{E}[X^k] = \frac{1}{i^k} \frac{d^k}{d\omega^k} \phi_X(\omega)\Big|_{\omega=0}.$$

Demostración: Expandimos $e^{i\omega X}$ en serie de Taylor alrededor de $\omega = 0$:

$$e^{i\omega X} = 1 + i\omega X + \frac{(i\omega X)^2}{2!} + \frac{(i\omega X)^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{(i\omega X)^k}{k!}.$$

Tomando la esperanza:

$$\phi_X(\omega) = \mathbb{E}[e^{i\omega X}] = \sum_{k=0}^{\infty} \frac{(i\omega)^k}{k!} \mathbb{E}[X^k],$$

siempre que los momentos $\mathbb{E}[X^k]$ existan. Derivando k veces con respecto a t:

$$\frac{d^k}{d\omega^k} \phi_X(\omega) = \sum_{m=k}^{\infty} \frac{i^m m(m-1)\cdots(m-k+1)}{m!} \omega^{m-k} \mathbb{E}[X^m].$$

Evaluando en $\omega = 0$, todos los términos con ω^{m-k} (donde $m > k$) se anulan, y el término con $m = k$ da:

$$\frac{d^k}{d\omega^k} \phi_X(\omega)\Big|_{\omega=0} = i^k \mathbb{E}[X^k].$$

Por lo tanto:

$$\mathbb{E}[X^k] = \frac{1}{i^k} \frac{d^k}{d\omega^k} \phi_X(\omega)\Big|_{\omega=0}.$$

8.7.2 Unicidad de la Función Característica

Una de las propiedades más importantes de la función característica es que determina de manera única la distribución de una variable aleatoria, lo que la convierte en una herramienta fundamental en teoría de probabilidad.

Teorema 8.7 Unicidad de la Función Característica

Sean X y Y dos variables aleatorias. Si sus funciones características son idénticas, es decir:

$$\phi_X(\omega) = \phi_Y(\omega), \quad \forall \omega \in \mathbb{R},$$

entonces X e Y tienen la misma función de distribución:

$$F_X(x) = F_Y(x), \quad \forall x \in \mathbb{R}.$$

Demostración: La unicidad de la función característica se deriva de su relación con la transformada de Fourier y el teorema de inversión. Supongamos que $\phi_X(\omega) = \phi_Y(\omega)$ para todo $\omega \in \mathbb{R}$. Queremos mostrar que las funciones de distribución de X y Y, denotadas $F_X(x)$ y $F_Y(x)$, son idénticas, es decir, $F_X(x) = F_Y(x)$ para todo $x \in \mathbb{R}$.

La función característica $\phi_X(\omega) = \mathbb{E}[e^{i\omega X}]$ es la transformada de Fourier de la medida de probabilidad inducida por X. Si X tiene una función de densidad $f_X(x)$, entonces:

$$\phi_X(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} f_X(x)\, dx,$$

y de manera similar para Y con densidad $f_Y(y)$. Dado que $\phi_X(\omega) = \phi_Y(t)$, las transformadas de Fourier de las densidades (o de las medidas de probabilidad, en el caso general) son idénticas.

Ahora, aplicamos el *Teorema de inversión* visto en Cálculo Avanzado y Aplicaciones, en la sección de la transformada de fourier. Si F_X y F_Y son funciones de distribución continuas, el teorema nos permite recuperar F_X a partir de $\phi_X(\omega)$. Específicamente, para cualquier $a < b$ que sean puntos de continuidad de F_X y F_Y, se tiene:

$$F_X(b) - F_X(a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-i\omega a} - e^{-i\omega b}}{i\omega} \phi_X(\omega)\, d\omega.$$

De manera análoga:

$$F_Y(b) - F_Y(a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-i\omega a} - e^{-i\omega b}}{i\omega} \phi_Y(\omega)\, d\omega.$$

Dado que $\phi_X(\omega) = \phi_Y(\omega)$ para todo t, las integrales son idénticas, y por lo tanto:

$$F_X(b) - F_X(a) = F_Y(b) - F_Y(a),$$

para todo $a < b$ que sean puntos de continuidad de ambas funciones.

Para extender este resultado a todos los $x \in \mathbb{R}$, consideremos un Corolario del teorema de inversión, el cúal nos dice, que si F_X y F_Y son continuas y tienen la misma transformada de Fourier (es decir, $\phi_X(\omega) = \phi_Y(\omega)$), entonces $F_X = F_Y$.

Por lo tanto, $F_X(x) = F_Y(x)$ para todo $x \in \mathbb{R}$, lo que implica que X e Y tienen la misma distribución.

8.7.3 Función Característica de una Suma de Variables Independientes

La función característica es particularmente útil para estudiar sumas de variables aleatorias independientes, ya que su estructura exponencial permite factorizar la esperanza conjunta.

Teorema 8.8 Función Característica de una Suma

Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes con funciones características $\phi_{X_i}(\omega)$, entonces la función característica de su suma es:

$$\phi_{X_1+\cdots+X_n}(\omega) = \phi_{X_1}(\omega) \cdot \phi_{X_2}(\omega) \cdots \phi_{X_n}(\omega).$$

Demostración: Sea $Y = X_1 + X_2 + \cdots + X_n$. Entonces:

$$\phi_Y(\omega) = \mathbb{E}[e^{i\omega Y}] = \mathbb{E}[e^{i\omega(X_1+X_2+\cdots+X_n)}] = \mathbb{E}[e^{i\omega X_1} e^{i\omega X_2} \cdots e^{i\omega X_n}].$$

Dado que $X_1, X_2, \ldots, X_n$ son independientes, la esperanza del producto es el producto de las esperanzas:

$$\phi_Y(\omega) = \mathbb{E}[e^{i\omega X_1}] \cdot \mathbb{E}[e^{i\omega X_2}] \ldots \mathbb{E}[e^{i\omega X_n}] = \phi_{X_1}(t) \cdot \phi_{X_2}(\omega) \cdots \phi_{X_n}(\omega).$$

8.7.4 Funciones Características Conocidas

A continuación, se listan las funciones características de algunas distribuciones conocidas. Para cada distribución, se especifican los parámetros, la expresión de la función característica y el dominio de ω en el que está definida.

- **Binomial:** Sea $X \sim \text{Bin}(n, p)$, con $n \in \mathbb{N}$ y $0 < p < 1$, entonces su función característica está dada por:

$$\phi_X(\omega) = (1 - p + pe^{i\omega})^n, \quad \omega \in \mathbb{R}.$$

- **Poisson:** Sea $X \sim \text{Poisson}(\lambda)$, con $\lambda > 0$, entonces su función característica está dada por:

$$\phi_X(\omega) = e^{\lambda(e^{i\omega}-1)}, \quad \omega \in \mathbb{R}.$$

- **Geométrica:** Sea $X \sim \text{Geom}(p)$, con $0 < p < 1$, entonces su función característica está dada por:

$$\phi_X(\omega) = \frac{pe^{i\omega}}{1 - (1-p)e^{i\omega}}, \quad \omega \in \mathbb{R}.$$

- **Exponencial:** Sea $X \sim \text{Exp}(\lambda)$, con $\lambda > 0$, entonces su función característica está dada por:

$$\phi_X(\omega) = \frac{\lambda}{\lambda - i\omega}, \quad \omega \in \mathbb{R}.$$

- **Normal:** Sea $X \sim N(\mu, \sigma^2)$, con $\mu \in \mathbb{R}$ y $\sigma^2 > 0$, entonces su función característica está dada por:

$$\phi_X(\omega) = e^{i\mu\omega - \frac{\sigma^2\omega^2}{2}}, \quad \omega \in \mathbb{R}.$$

- **Uniforme:** Sea $X \sim \text{Unif}(a, b)$, con $a < b$, entonces su función característica está dada por:

$$\phi_X(\omega) = \frac{e^{i\omega b} - e^{i\omega a}}{i\omega(b-a)}, \quad \omega \in \mathbb{R} \text{ (con } \omega \neq 0).$$

- **Gamma:** Sea $X \sim \text{Gamma}(\alpha, \beta)$, con $\alpha > 0$ y $\beta > 0$, entonces su función característica está dada por:

$$\phi_X(\omega) = \left(1 - \frac{i\omega}{\beta}\right)^{-\alpha}, \quad \omega \in \mathbb{R}.$$

Capítulo 9

Convergencia & Teoremas Límite

En el estudio de fenómenos aleatorios, calcular probabilidades exactas puede ser un desafío, especialmente cuando las distribuciones subyacentes son complejas o desconocidas. Sin embargo, si conocemos ciertas características de las variables aleatorias, como su esperanza o varianza, es posible obtener cotas que limiten la probabilidad de ciertos eventos. Este tipo de análisis, nos permite manejar la incertidumbre de manera práctica, proporcionando límites superiores para eventos que podrían ser difíciles de evaluar directamente.

Las desigualdades, a su vez, sientan las bases para comprender el comportamiento de secuencias de variables aleatorias a medida que el número de observaciones aumenta. Este comportamiento se formaliza a través de los conceptos de convergencia, que describen cómo una secuencia de variables aleatorias se aproxima a un valor o distribución límite. La convergencia es esencial para analizar fenómenos a largo plazo, como la estabilidad de promedios en procesos repetitivos o la distribución de sumas acumuladas.

Finalmente, los conceptos de convergencia nos llevan a los teoremas límite, que son resultados fundamentales en la teoría de la probabilidad. Estos teoremas, como la Ley de los Grandes Números y el Teorema Central del Límite, describen patrones universales en el comportamiento de promedios y sumas de variables aleatorias, incluso cuando las distribuciones individuales son diversas. Su importancia radica en que proporcionan una base teórica para aproximar distribuciones complicadas mediante distribuciones más simples, como la normal.

9.1 Desigualdades

Las desigualdades probabilísticas nos permiten acotar la probabilidad de eventos raros, extremos o de nuestro interés, especialmente cuando no conocemos la distribución exacta de una variable aleatoria. Estas desigualdades son útiles en contextos donde necesitamos garantizar que ciertos eventos no ocurran con alta probabilidad, como en el control de calidad o la gestión de riesgos.

9.1.1 Cota de la Unión (Desigualdad de Boole)

La desigualdad de Boole, también conocida como cota de la unión, es un resultado fundamental que ya hemos explorado en secciones anteriores. Este resultado establece un límite superior para la probabilidad de que ocurra al menos uno de varios eventos, lo que resulta particularmente útil cuando calcular la probabilidad exacta de la unión es complicado debido a las intersecciones entre eventos.
Esta desigualdad surge de la observación de que la función de probabilidad $\mathbb{P}$ mapea los eventos al intervalo $[0, 1]$ y cumple con la propiedad de aditividad para eventos disjuntos. Sin embargo, cuando los eventos no son mutuamente excluyentes, la suma

Teorema 9.1 Desigualdad de Boole

Para una colección de eventos $A_1, A_2, \ldots, A_n$ en un espacio de probabilidad, la probabilidad de la unión de estos eventos satisface:

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

de sus probabilidades individuales puede sobrestimar la probabilidad de la unión, ya que incluye múltiples veces las intersecciones. Por ejemplo, para dos eventos A y B, sabemos que:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Dado que $\mathbb{P}(A \cap B) \geq 0$, se deduce que:

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

lo que es precisamente la desigualdad de Boole para dos eventos. La cota se extiende de manera natural a n eventos, proporcionando un límite superior simple y efectivo.

Luego, a partir de esta desigualdad, Bonferroni la generalizó, obteniendo una serie de cotas que alternan entre límites superiores e inferiores para la probabilidad de la unión. La idea detrás de esta generalización es refinar la cota de Boole ajustándola con términos que consideran las intersecciones de los eventos. Intuitivamente, si en la cota de Boole sumamos las probabilidades individuales $\mathbb{P}(A_i)$, estamos contando de más las intersecciones dobles $\mathbb{P}(A_i \cap A_j)$. Para corregir esto, podemos restar estas intersecciones, pero entonces subestimamos la probabilidad porque las intersecciones triples $\mathbb{P}(A_i \cap A_j \cap A_k)$ se restan demasiadas veces. Bonferroni propuso un enfoque que consiste en alternar entre sumar y restar términos de intersecciones de orden creciente, obteniendo una secuencia de cotas que se acercan cada vez más a la probabilidad exacta.

Por ejemplo, para n eventos $A_1, A_2, \ldots, A_n$, la primera cota (que es la desigualdad de Boole) es:

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

Si consideramos las intersecciones dobles, obtenemos una cota inferior:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j).$$

Añadiendo las intersecciones triples, volvemos a una cota superior:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k).$$

Este proceso puede continuar, alternando entre cotas superiores e inferiores, y cada nuevo término hace que la aproximación sea más precisa. Sin embargo, en la

práctica, calcular intersecciones de orden alto puede ser complicado, especialmente si los eventos son dependientes o si no tenemos información detallada sobre sus intersecciones.

Ejemplo:

Supongamos que en una línea de producción, hay tres máquinas M_1, M_2, M_3, y cada una tiene una probabilidad de fallar en un día dado: $\mathbb{P}(M_1 \text{ falla}) = 0{,}1$, $\mathbb{P}(M_2 \text{ falla}) = 0{,}15$, y $\mathbb{P}(M_3 \text{ falla}) = 0{,}2$. Queremos estimar la probabilidad de que al menos una máquina falle. Usando la desigualdad de Boole:

$$\mathbb{P}(\text{al menos una falla}) \leq \mathbb{P}(M_1 \text{ falla}) + \mathbb{P}(M_2 \text{ falla}) + \mathbb{P}(M_3 \text{ falla}) = 0{,}1 + 0{,}15 + 0{,}2 = 0{,}45.$$

Esto nos dice que la probabilidad de que al menos una máquina falle es a lo más 0.45. Si tuviéramos información sobre las intersecciones (por ejemplo, la probabilidad de que dos máquinas fallen simultáneamente), podríamos usar las cotas de Bonferroni para obtener una estimación más ajustada.

9.1.2 Desigualdad de Markov

La desigualdad de Markov proporciona un límite superior para la probabilidad de que una variable aleatoria no negativa tome valores grandes. Es especialmente útil cuando se dispone de información sobre el valor esperado de una variable.

Teorema 9.2 Desigualdad de Markov

Sea X una variable aleatoria no negativa, y sea $a > 0$. Entonces:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

La desigualdad se deduce a partir de que para todo $x \geq 0$ y $a > 0$, se cumple

$$x \geq a \cdot \mathbb{1}_{\{x \geq a\}}.$$

Esto debido a que si $x < a$, la indicatriz vale cero y $x \geq 0 = a \cdot 0$; si $x \geq a$, la indicatriz vale uno y se tiene $x \geq a = a \cdot 1$. Al extender esta desigualdad punto a punto a la variable aleatoria X, se obtiene:

$$X \geq a \cdot \mathbb{1}_{\{X \geq a\}}.$$

Aplicando la esperanza en ambos lados y recordando que la esperanza es monótona, resulta:

$$\mathbb{E}[X] \geq \mathbb{E}\left[a \cdot \mathbb{1}_{\{X \geq a\}}\right] = a \cdot \mathbb{E}\left[\mathbb{1}_{\{X \geq a\}}\right].$$

En donde, la última igualdad se obtiene por la linealidad de la esperanza, y el hecho de que la esperanza de una indicatriz es, por definición, la probabilidad del evento que representa:

$$\mathbb{E}[\mathbb{1}_{\{X \geq a\}}] = 0 \cdot \mathbb{P}(X \leq a) + 1 \cdot \mathbb{P}(X \geq a) = \mathbb{P}(X \geq a).$$

Por lo tanto,

$$\mathbb{E}[X] \geq a \cdot \mathbb{P}(X \geq a),$$

lo que, al despejar, conduce directamente a la desigualdad de Markov:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Ejemplo:

Supongamos que X representa el tiempo (en horas) que una persona pasa diariamente usando una aplicación móvil, y que se conoce que $\mathbb{E}[X] = 1{,}5$. Si se desea acotar la probabilidad de que una persona haya usado la aplicación por al menos 5 horas en un día, se puede aplicar directamente la desigualdad de Markov:

$$\mathbb{P}(X \geq 5) \leq \frac{1{,}5}{5} = 0{,}3.$$

Esto implica que, como máximo, un 30 % de los usuarios podrían haber superado las 5 horas de uso.

9.1.3 Desigualdad de Chebyshev

La desigualdad de Chebyshev permite estimar la probabilidad de que una variable aleatoria se desvíe significativamente de su media, en función de su varianza. A diferencia de la desigualdad de Markov, que se aplica a variables no negativas, esta desigualdad es aplicable a cualquier variable aleatoria con varianza finita.

Teorema 9.3 Desigualdad de Chebyshev

Sea X una variable aleatoria con esperanza finita $\mu = \mathbb{E}[X]$ y varianza finita $\sigma^2 = \mathrm{Var}(X)$. Entonces, para todo $k > 0$:

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

Esta desigualdad se deduce aplicando la desigualdad de Markov sobre la variable aleatoria no negativa $(X - \mu)^2$, que mide el cuadrado de la desviación respecto a la media:

$$\mathbb{P}(|X - \mu| \geq k) = \mathbb{P}((X - \mu)^2 \geq k^2).$$

Como $(X-\mu)^2 \geq k^2$ implica directamente que la variable $(X-\mu)^2$ supera el umbral k^2, se puede aplicar la desigualdad de Markov:

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

Por lo tanto,

$$\mathbb{P}(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2},$$

lo cual acota la probabilidad de que la variable se aleje de su valor esperado por más de k unidades.

Ejemplo:

Supongamos que X representa la duración (en minutos) de una llamada telefónica,

con media $\mu = 10$ y varianza $\sigma^2 = 4$. Si se desea acotar la probabilidad de que una llamada dure al menos 6 minutos más o menos que el promedio, es decir, $|X-10| \geq 6$, se puede aplicar la desigualdad de Chebyshev:

$$\mathbb{P}(|X - 10| \geq 6) \leq \frac{4}{6^2} = \frac{4}{36} = \frac{1}{9} \approx 0{,}111.$$

Esto indica que, como máximo, aproximadamente un 11.1 % de las llamadas difieren de la media en 6 minutos o más.

9.1.4 Cotas de Chernoff

Las cotas de Chernoff ofrecen límites exponencialmente decrecientes para la probabilidad de que una variable aleatoria se desvíe por encima o por debajo de cierto umbral. Estas cotas se derivan a partir de la desigualdad de Markov aplicada a la función de la variable aleatoria e^{sX}, y se expresan naturalmente en términos de la función generadora de momentos.

Teorema 9.4 Cotas de Chernoff

Sea X una variable aleatoria y $a \in \mathbb{R}$. Entonces, para cualquier $s \in \mathbb{R}$:

$$\mathbb{P}(X \geq a) \leq e^{-sa} \cdot M_X(s), \quad \text{para todo } s > 0,$$

$$\mathbb{P}(X \leq a) \leq e^{-sa} \cdot M_X(s), \quad \text{para todo } s < 0,$$

donde $M_X(s) = \mathbb{E}[e^{sX}]$ es la función generadora de momentos de X.

La deducción se basa en que e^{sX} es una variable aleatoria no negativa para todo $s \in \mathbb{R}$. Usando la desigualdad de Markov sobre e^{sX}, la cual es una función monótona, se obtiene:

$$\mathbb{P}(X \geq a) = \mathbb{P}(e^{sX} \geq e^{sa}) \leq \frac{\mathbb{E}[e^{sX}]}{e^{sa}} = e^{-sa} M_X(s), \quad \text{si } s > 0,$$

$$\mathbb{P}(X \leq a) = \mathbb{P}(e^{sX} \geq e^{sa}) \leq \frac{\mathbb{E}[e^{sX}]}{e^{sa}} = e^{-sa} M_X(s), \quad \text{si } s < 0.$$

Como la desigualdad es válida para todo s positivo o negativo (según el caso), se puede optimizar la cota minimizando respecto a s:

$$\mathbb{P}(X \geq a) \leq \min_{s>0} e^{-sa} M_X(s), \qquad \mathbb{P}(X \leq a) \leq \min_{s<0} e^{-sa} M_X(s).$$

Ejemplo:

Sea X una variable aleatoria con $M_X(s) = e^{s^2}$ (por ejemplo, una variable normal estándar centrada con varianza 2). Se desea acotar $\mathbb{P}(X \geq 4)$. Usando Chernoff para $s > 0$:

$$\mathbb{P}(X \geq 4) \leq \min_{s>0} e^{-4s} e^{s^2} = \min_{s>0} e^{s^2 - 4s}.$$

La función $s^2 - 4s$ se minimiza en $s = 2$, por lo que:

$$\mathbb{P}(X \geq 4) \leq e^{2^2 - 4 \cdot 2} = e^{-4}.$$

Es decir, la probabilidad de que X exceda 4 es menor que $e^{-4} \approx 0{,}018$.

9.1.5 Comparación entre desigualdades

Las tres desigualdades consideradas —Markov, Chebyshev y Chernoff— ofrecen cotas superiores para la probabilidad de que una variable aleatoria se aleje de cierto valor central. Sin embargo, difieren en sus hipótesis, generalidad y precisión.

Cada una de estas desigualdades se puede interpretar como un refinamiento progresivo. La desigualdad de Chebyshev puede verse como una aplicación de la desigualdad de Markov sobre la variable cuadrática $(X-\mu)^2$, mientras que la cota de Chernoff aplica Markov sobre e^{sX}, optimizando sobre s para obtener una mejor cota.

Supongamos que una variable aleatoria X tiene media $\mu = 0$ y varianza $\sigma^2 = 1$, y queremos acotar $\mathbb{P}(X \geq 5)$. Supongamos además que X cumple que $M_X(s) \leq e^{s^2/2}$. Entonces:

- **Markov:** Aplicando sobre X^2,

$$\mathbb{P}(X \geq 5) \leq \mathbb{P}(X^2 \geq 25) \leq \frac{\mathbb{E}[X^2]}{25} = \frac{1}{25} = 0{,}04.$$

- **Chebyshev:**

$$\mathbb{P}(|X| \geq 5) \leq \frac{1}{25} \quad \Rightarrow \quad \mathbb{P}(X \geq 5) \leq \frac{1}{25} = 0{,}04.$$

- **Chernoff:** Usando $M_X(s) \leq e^{s^2/2}$, optimizamos

$$\mathbb{P}(X \geq 5) \leq \min_{s>0} e^{-5s} e^{s^2/2}.$$

  El mínimo ocurre en $s = 5$, y da:

$$\mathbb{P}(X \geq 5) \leq e^{-25/2} \approx 3{,}73 \times 10^{-6}.$$

Esto evidencia cómo Chernoff ofrece una cota mucho más ajustada para desvíos grandes, a costa de requerir más información sobre la estructura de X a través de su función generadora de momentos. En particular, se puede destacar lo siguiente:

- **Chernoff vs. Chebyshev:** Las cotas de Chernoff dominan estrictamente a la desigualdad de Chebyshev siempre que la función generadora de momentos esté definida, ya que logran un decaimiento exponencial en lugar de uno polinomial.
- **Chernoff vs. Markov:** Aunque Chernoff suele ser más precisa, en algunos casos puntuales, cuando se busca acotar $\mathbb{P}(X \geq a)$ y a es pequeño (cercano a la media), la desigualdad de Markov puede proporcionar una cota igual o incluso más ajustada si se aplica directamente sobre X, sin centrado ni transformación.
- **Chebyshev vs. Markov:** La desigualdad de Chebyshev mejora la de Markov cuando se conoce la varianza de X y se estudia la desviación respecto a su media. No obstante, si el valor a no está centrado en la esperanza (es decir, si se estudia $\mathbb{P}(X \geq a)$ sin restar μ), entonces aplicar Markov directamente a X puede ser más efectivo que usar Chebyshev sobre $|X - \mu|$.

La elección de una u otra cota depende del tipo de información disponible sobre la variable y del rango de desviaciones que se desea estimar.

9.1.6 Desigualdad de Cauchy–Schwarz

La desigualdad de Cauchy–Schwarz es probable que haya sido vista anteriormente en contextos de álgebra lineal o cálculo, aplicada a productos internos en $\mathbb{R}^n$ o a integrales. En el contexto de las probabilídades, esta desigualdad conserva la misma estructura y se expresa mediante esperanzas de productos de variables aleatorias.

Teorema 9.5 Desigualdad de Cauchy–Schwarz

Sean X y Y variables aleatorias reales con $\mathbb{E}[X^2] < \infty$ y $\mathbb{E}[Y^2] < \infty$. Entonces:

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2] \cdot \mathbb{E}[Y^2]}.$$

Esta desigualdad puede interpretarse como una generalización del hecho de que el valor esperado de un producto no puede ser "demasiado grande" comparado con las esperanzas cuadráticas individuales. En particular, si las variables aleatorias X y Y están fuertemente correlacionadas en magnitud (aunque no necesariamente en signo), entonces el valor absoluto de $\mathbb{E}[XY]$ se aproxima al producto de las normas cuadráticas.

Demostración: Consideremos la variable aleatoria $X - \lambda Y$, cuya esperanza cuadrática es siempre no negativa:

$$\mathbb{E}[(X - \lambda Y)^2] \geq 0, \quad \text{para todo } \lambda \in \mathbb{R}.$$

Expandiendo:

$$\mathbb{E}[X^2] - 2\lambda\mathbb{E}[XY] + \lambda^2\mathbb{E}[Y^2] \geq 0.$$

Ahora, esto se cumple para cualquier $\lambda \in \mathbb{R}$, así que elegimos un valor conveniente:

$$\lambda = \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}.$$

Sustituyendo este valor:

$$\begin{aligned} &\mathbb{E}[(X - \lambda Y)^2] \geq 0, \\ \iff &\mathbb{E}[X^2] - 2 \cdot \frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]} + \frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]^2} \cdot \mathbb{E}[Y^2] \geq 0, \\ \iff &\mathbb{E}[X^2] - \frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]} \geq 0. \end{aligned}$$

Multiplicando por $\mathbb{E}[Y^2]$ en ambos lados se obtiene:

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

aplicando la raíz cuadrática:

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2] \cdot \mathbb{E}[Y^2]}.$$

Ejemplo:

Supongamos que X es una variable aleatoria con $\mathbb{E}[X^2] = 4$ y Y una variable aleatoria con $\mathbb{E}[Y^2] = 9$. Entonces, cualquier valor posible de $\mathbb{E}[XY]$ debe satisfacer:

$$|\mathbb{E}[XY]| \leq \sqrt{4 \cdot 9} = 6.$$

Esto acota el valor absoluto del producto esperado sin requerir conocer la distribución conjunta de X y Y.
La desigualdad de Cauchy–Schwarz es especialmente útil para establecer otras desigualdades fundamentales, como Jensen o la de la correlación, y se utiliza también en la deducción de propiedades de la covarianza o la varianza de sumas.

9.1.7 Desigualdad de Jensen

La desigualdad de Jensen relaciona la esperanza de una transformación convexa de una variable aleatoria con la transformación convexa de su esperanza. Antes de enunciarla formalmente, recordemos brevemente el concepto de convexidad.

Definición 9.1 (Función convexa) Sea $g : I \subseteq \mathbb{R} \to \mathbb{R}$ una función definida en un intervalo I. Se dice que g es convexa si para todo $x, y \in I$ y todo $\lambda \in [0, 1]$, se cumple:

$$g(\lambda x + (1 - \lambda)y) \leq \lambda g(x) + (1 - \lambda)g(y).$$

Además, si g es dos veces derivable en I, entonces g es convexa si y solo si su segunda derivada es no negativa en todo el intervalo:

$$g''(x) \geq 0 \quad \text{para todo } x \in I.$$

Geométricamente, esta definición significa que el gráfico de una función convexa se encuentra por debajo del segmento recto que une dos de sus puntos. Ejemplos clásicos de funciones convexas son $g(x) = x^2$, $g(x) = e^x$ y $g(x) = -\log x$ (en su dominio correspondiente).

Teorema 9.6 Desigualdad de Jensen

Sea X una variable aleatoria tal que $\mathbb{E}[X]$ está bien definida y finita, y sea g una función convexa tal que $g(X)$ también tiene esperanza finita. Entonces:

$$g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)].$$

Esta desigualdad formaliza la intuición de que aplicar una transformación convexa a una media da un resultado menor o igual que la media de la transformación aplicada punto a punto. El sentido de la desigualdad se invierte si g es cóncava.
Demostración: Sea $g : I \to \mathbb{R}$ una función convexa definida en un intervalo $I \subseteq \mathbb{R}$, y sea X una variable aleatoria tal que $X(\omega) \in I$ para todo ω, con $\mathbb{E}[X] \in I$ y $\mathbb{E}[g(X)] < \infty$.

Por convexidad de g, existe para todo $x_0 \in I$ una recta $h(x) = ax + b$ tal que:

$$g(x) \geq h(x) \quad \text{para todo } x \in I, \quad \text{y} \quad g(x_0) = h(x_0).$$

Dicho h se interpreta como una tangente inferior al gráfico de g en el punto x_0. Elegimos $x_0 = \mathbb{E}[X]$ y consideramos la función afín $h(x) = ax + b$ que satisface:

$$g(x) \geq h(x), \quad \text{con igualdad en } x = \mathbb{E}[X].$$

Como $g(X) \geq h(X)$ punto a punto, se tiene que:

$$\mathbb{E}[g(X)] \geq \mathbb{E}[h(X)].$$

Dado que h es afín, se puede extraer la constante y usar linealidad de la esperanza:

$$\mathbb{E}[h(X)] = \mathbb{E}[aX + b] = a\mathbb{E}[X] + b = h(\mathbb{E}[X]).$$

Pero por construcción, $h(\mathbb{E}[X]) = g(\mathbb{E}[X])$, por lo tanto:

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X]),$$

que es precisamente la desigualdad de Jensen.

Ejemplo:
Sea X una variable aleatoria con $\mathbb{E}[X] = 0$ y $\mathbb{E}[X^2] = 1$. Tomando $g(x) = x^2$, que es convexa, se tiene:

$$g(\mathbb{E}[X]) = g(0) = 0, \qquad \mathbb{E}[g(X)] = \mathbb{E}[X^2] = 1.$$

Así, se verifica:

$$0 = g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)] = 1,$$

tal como afirma la desigualdad de Jensen.

9.2 Convergencia de Variables Aleatorias

En muchos contextos, es natural preguntarse si una secuencia de variables aleatorias $X_1, X_2, \ldots$ se aproxima, en cierto sentido, a una variable aleatoria X a medida que el índice crece. Por ejemplo, cuando se construyen estimaciones sucesivas de una cantidad desconocida, se espera que estas se acerquen al valor verdadero. En esta sección se formaliza qué significa que $X_n \to X$, introduciendo distintas nociones de convergencia: casi segura, en probabilidad, en media, y en distribución. Cada una refleja una forma distinta de interpretar el concepto de "aproximarse" o "hacerse más cercano".

9.2.1 Convergencia de Sucesiones Numéricas

Antes de estudiar la convergencia de variables aleatorias, conviene revisar primero el concepto de convergencia de sucesiones de números reales. Esto permitirá establecer una analogía clara entre ambos contextos, ya que muchas ideas sobre "aproximación" provienen de esta noción básica.

Definición 9.2 (Convergencia de una sucesión numérica) Sea $\{a_n\}_{n\in\mathbb{N}}$ una sucesión de números reales. Se dice que a_n converge a un número real L si, para todo $\epsilon > 0$, existe un número natural N tal que

$$|a_n - L| < \epsilon, \quad \text{para todo } n > N.$$

En notación matemática, esto se escribe como:

$$\lim_{n\to\infty} a_n = L.$$

Intuitivamente, esto significa que los términos de la sucesión eventualmente se mantienen tan cerca del valor límite L como se desee, siempre que n sea suficientemente grande.

Ejemplo:

Consideremos la sucesión $\{a_n\}$ definida por:

$$a_n = \frac{n}{n+1}, \quad n = 1, 2, 3, \ldots$$

Observamos que a medida que n crece, el valor de a_n se acerca cada vez más a 1, ya que el numerador y el denominador se aproximan entre sí. En efecto,

$$\lim_{n\to\infty} \frac{n}{n+1} = 1.$$

Este tipo de comportamiento es la base sobre la cual se construye la teoría de convergencia en espacios probabilísticos. Las secuencias de variables aleatorias pueden entenderse como versiones aleatorizadas de estas sucesiones, y las nociones de convergencia que veremos más adelante generalizan esta definición clásica.

9.2.2 Secuencias de Variables Aleatorias

En probabilidad, una *secuencia de variables aleatorias* $\{X_n\}_{n=1}^{\infty}$ es un conjunto de funciones definidas sobre un mismo espacio muestral Ω, con una medida de probabilidad $\mathbb{P}$. Cada función X_n es una variable aleatoria que asigna un valor real a cada elemento $\omega \in \Omega$:

$$X_n(\omega) = x_{n\omega}, \quad \text{para todo } \omega \in \Omega.$$

En este contexto, Ω representa todos los posibles resultados de un experimento aleatorio, y cada X_n puede interpretarse como una observación sucesiva de un mismo fenómeno. Por ejemplo, si Ω describe los resultados del lanzamiento de un dado, entonces X_1 podría ser el resultado del primer lanzamiento, X_2 del segundo, y así sucesivamente. La secuencia completa $\{X_n\}$ modela un proceso aleatorio a lo largo del tiempo o sobre múltiples réplicas del mismo experimento.
Formalmente, cada variable aleatoria X_n es una función medible $X_n : \Omega \to \mathbb{R}$, y una secuencia $\{X_n\}$ puede entenderse como una secuencia de funciones del tipo:

$$X_n : \Omega \longrightarrow \mathbb{R}, \quad n = 1, 2, 3, \ldots$$

Este marco permite analizar el comportamiento conjunto de las variables a medida que n crece, lo que motiva el estudio de diferentes nociones de *convergencia de variables aleatorias*.

9.2.3 Convergencia en Distribución

La convergencia en distribución es la forma más débil de convergencia entre secuencias de variables aleatorias. Intuitivamente, esta noción describe que las distribuciones de las variables aleatorias X_n se aproximan cada vez más a la distribución de una variable aleatoria X a medida que $n \to \infty$. Esta convergencia no implica que los valores de X_n se acerquen a los de X, sino que las probabilidades acumuladas asociadas a X_n tienden a coincidir con las de X.

Definición 9.3 (Convergencia en distribución) Sea $\{X_n\}_{n\in\mathbb{N}}$ una secuencia de variables aleatorias y X otra variable aleatoria. Se dice que X_n converge en distribución a X, y se escribe

$$X_n \xrightarrow{d} X,$$

si se cumple que

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \quad \text{para todo } x \in \mathbb{R} \text{ donde } F_X \text{ es continua,}$$

donde F_{X_n} y F_X son las funciones de distribución acumulada de X_n y X, respectivamente.

Esta definición establece que la probabilidad acumulada de X_n en cualquier punto de continuidad de F_X converge a la probabilidad acumulada de X en ese mismo punto. Por tanto, lo que converge es la forma de la distribución, no los valores de las variables.

Ejemplo:

Sea $\{X_n\}_{n\in\mathbb{N}}$ una secuencia de variables aleatorias con función de distribución acumulada dada por:

$$F_{X_n}(x) = \begin{cases} 1 - \left(1 - \frac{1}{n}\right)^{nx}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Queremos demostrar que $X_n \xrightarrow{d} X$, donde $X \sim$ Exponencial(1), es decir, $F_X(x) = 1 - e^{-x}$ para $x \geq 0$, y $F_X(x) = 0$ para $x < 0$.
Estudiamos el límite de $F_{X_n}(x)$ para distintos valores de x:

- Para $x \leq 0$, se tiene directamente:

$$F_{X_n}(x) = 0 = F_X(x).$$

- Para $x > 0$, observamos:

$$\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} \left[1 - \left(1 - \frac{1}{n}\right)^{nx}\right] = 1 - \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^{nx}.$$

  Sabemos que $\lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^{n} = \frac{1}{e}$, por lo tanto:

$$\lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^{nx} = e^{-x}.$$

  Así, concluimos que:

$$\lim_{n\to\infty} F_{X_n}(x) = 1 - e^{-x} = F_X(x).$$

Por lo tanto, se tiene que:

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \quad \text{para todo } x \in \mathbb{R} \text{ donde } F_X \text{ es continua,}$$

y así se concluye que:

$$X_n \xrightarrow{d} X, \quad \text{con } X \sim \text{Exponencial}(1).$$

Teorema 9.7 Convergencia de variables discretas

Sea $\{X_n\}$ una secuencia de variables aleatorias discretas no negativas con valores en subconjuntos de $\mathbb{N}$, es decir,

$$R_X \subset \{0, 1, 2, \dots\}, \quad R_{X_n} \subset \{0, 1, 2, \dots\}.$$

Entonces,

$$X_n \xrightarrow{d} X \quad \text{si y solo si} \quad \lim_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k), \quad \text{para todo } k \in \mathbb{N}.$$

Demostración:
(⇒) Supongamos que $X_n \xrightarrow{d} X$. Entonces, por definición, se tiene:

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \quad \text{para todo } x \in \mathbb{R} \text{ donde } F_X \text{ es continua.}$$

Como X toma valores en $\mathbb{N}$, su función de distribución F_X es discontinua exactamente en los enteros $k \in \mathbb{N}$, y continua en los reales que no pertenecen a $\mathbb{N}$. Sea $k \in \mathbb{N}$. Entonces F_X es continua en $k-1$, por lo que:

$$\lim_{n\to\infty} F_{X_n}(k-1) = F_X(k-1), \quad \text{y} \quad \lim_{n\to\infty} F_{X_n}(k) = F_X(k).$$

Luego, usando que:

$$\mathbb{P}(X_n = k) = F_{X_n}(k) - F_{X_n}(k-1),$$

tomando límites se obtiene:

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = F_X(k) - F_X(k-1) = \mathbb{P}(X = k),$$

para todo $k \in \mathbb{N}$.

(⇐) Supongamos ahora que

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k), \quad \text{para todo } k \in \mathbb{N}.$$

Entonces, para cualquier $x \in \mathbb{R}$, se tiene:

$$F_{X_n}(x) = \sum_{j\le x} \mathbb{P}(X_n = j) \to \sum_{j\le x} \mathbb{P}(X = j) = F_X(x).$$

Esto se tiene ya que la suma tiene un número finito de términos, y se puede intercambiar el límite con la suma. Por lo tanto:

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \quad \text{para todo } x \in \mathbb{R},$$

y así se concluye que $X_n \xrightarrow{d} X$.

Teorema 9.8 Conv. en distribución a través de FGM

Si $\{X_n\}$ es una secuencia de variables aleatorias tales que sus funciones generadoras de momentos existen y satisfacen:

$$\lim_{n\to\infty} M_{X_n}(s) = M_X(s), \quad \text{para todo } s \in (-c, c) \text{ con } c > 0,$$

entonces:

$$X_n \xrightarrow{d} X.$$

Demostración: Este resultado es una consecuencia del *Teorema de unicidad para funciones generadoras de momentos*. Si dos variables aleatorias tienen la misma función generadora de momentos en un entorno abierto del origen, entonces tienen la misma distribución. Si $M_{X_n}(s) \to M_X(s)$ puntualmente en un intervalo alrededor del cero, entonces la distribución de X_n converge a la de X en el sentido de distribución.

Ejemplo:

Sea $X_n \sim \text{Binomial}(n, \frac{1}{n})$. Su función generadora de momentos es

$$M_{X_n}(s) = \left(1 - \frac{1}{n} + \frac{1}{n}e^s\right)^n.$$

Entonces,

$$\lim_{n\to\infty} M_{X_n}(s) = e^{e^s - 1},$$

que es la función generadora de momentos de una distribución de Poisson(1), por lo tanto,

$$X_n \xrightarrow{d} \text{Poisson}(1).$$

9.2.4 Convergencia en Probabilidad

La convergencia en probabilidad es una forma más fuerte que la convergencia en distribución, pues convergencia en probabilidad implica convergencia en distribución, lo contrario no es necesariamente cierto. Este tipo de convergencia asegura que, para una secuencia de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$, la probabilidad de que X_n se aleje de X por al menos un valor $\epsilon > 0$ tiende a cero conforme $n \to \infty$. En otras palabras, X_n se aproxima a X con alta probabilidad.

Definición 9.4 (Convergencia en probabilidad) Sea $\{X_n\}_{n\in\mathbb{N}}$ una secuencia de variables aleatorias y X una variable aleatoria. Se dice que X_n converge en probabilidad a X, y se denota:

$$X_n \xrightarrow{\mathbb{P}} X,$$

si para todo $\epsilon > 0$ se cumple:

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0.$$

Esta definición implica que los eventos en los que X_n difiere significativamente de X se vuelven cada vez menos probables a medida que n crece. La convergencia en

probabilidad garantiza que las observaciones estén "concentradas" en torno a X con alta probabilidad.

Ejemplo:

Sea $X_n \sim \text{Exponencial}(n)$ y sea $X \equiv 0$ (la variable aleatoria constante nula). Queremos demostrar que $X_n \xrightarrow{\mathbb{P}} 0$.

Notamos que $X_n \geq 0$ para todo n, por lo que $|X_n - 0| = X_n$. Entonces, para todo $\epsilon > 0$,

$$\mathbb{P}(|X_n - 0| \geq \epsilon) = \mathbb{P}(X_n \geq \epsilon) = \int_{\epsilon}^{\infty} ne^{-nx}dx = e^{-n\epsilon}.$$

Por lo tanto,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - 0| \geq \epsilon) = \lim_{n\to\infty} e^{-n\epsilon} = 0, \quad \forall \epsilon > 0,$$

y concluimos que $X_n \xrightarrow{\mathbb{P}} 0$.

Teorema 9.9 Convergencia hacia una constante

Si $X_n \xrightarrow{d} c$ donde $c \in \mathbb{R}$ es constante, entonces también se cumple:

$$X_n \xrightarrow{\mathbb{P}} c.$$

Demostración: Sea $c \in \mathbb{R}$ una constante. Dado que $X_n \xrightarrow{d} c$, se tiene que para todo $x \in \mathbb{R}$ donde la función de distribución F_X es continua, se cumple:

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

donde $F_X(x) = \mathbb{1}_{\{x \geq c\}}$ es la función de distribución de la constante $X = c$.
Para probar que $X_n \xrightarrow{\mathbb{P}} c$, consideramos la definición:

$$\lim_{n\to\infty} \mathbb{P}(|X_n - c| \geq \epsilon) = 0, \quad \forall \epsilon > 0.$$

Fijado $\epsilon > 0$, descomponemos el evento:

$$\begin{aligned}\mathbb{P}(|X_n - c| \geq \epsilon) &= \mathbb{P}(X_n \leq c - \epsilon) + \mathbb{P}(X_n \geq c + \epsilon) \\ &= F_{X_n}(c - \epsilon) + \left[1 - F_{X_n}(c + \epsilon)\right].\end{aligned}$$

Usando la convergencia puntual de las funciones de distribución:

$$\lim_{n\to\infty} F_{X_n}(c - \epsilon) = F_X(c - \epsilon) = 0, \quad \lim_{n\to\infty} F_{X_n}(c + \epsilon) = F_X(c + \epsilon) = 1.$$

Por lo tanto:

$$\lim_{n\to\infty} \mathbb{P}(|X_n - c| \geq \epsilon) = 0 + (1 - 1) = 0,$$

lo cual es precisamente la convergencia en probabilidad hacia c.
Este resultado establece un caso particular donde la convergencia en distribución sí implica convergencia en probabilidad, aunque en general esto no ocurre.

Teorema 9.10 Conv. en probabilidad implica conv. en distribución

Si $X_n \xrightarrow{\mathbb{P}} X$, entonces $X_n \xrightarrow{d} X$.

Demostración Queremos probar que si X_n converge en probabilidad a X, entonces también converge en distribución. Es decir, si

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0, \quad \forall \epsilon > 0,$$

entonces

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \quad \text{para todo } x \in \mathbb{R} \text{ donde } F_X \text{ es continua.}$$

Fijamos un punto x de continuidad de F_X, y sea $\delta > 0$ arbitrario. Consideramos:

$$\begin{aligned} F_{X_n}(x) &= \mathbb{P}(X_n \leq x) \\ &= \mathbb{P}(X_n \leq x, X \leq x + \epsilon) + \mathbb{P}(X_n \leq x, X > x + \epsilon) \\ &\leq \mathbb{P}(X \leq x + \epsilon) + \mathbb{P}(|X_n - X| \geq \epsilon). \end{aligned}$$

Del mismo modo, también se puede acotar por abajo:

$$\begin{aligned} F_{X_n}(x) &= \mathbb{P}(X_n \leq x) \\ &\geq \mathbb{P}(X \leq x - \epsilon) - \mathbb{P}(|X_n - X| \geq \epsilon). \end{aligned}$$

Entonces se obtiene la desigualdad doble:

$$\mathbb{P}(X \leq x - \epsilon) - \mathbb{P}(|X_n - X| \geq \epsilon) \leq F_{X_n}(x) \leq \mathbb{P}(X \leq x + \epsilon) + \mathbb{P}(|X_n - X| \geq \epsilon).$$

Tomando límite cuando $n \to \infty$:

$$F_X(x - \epsilon) \leq \liminf_{n\to\infty} F_{X_n}(x) \leq \limsup_{n\to\infty} F_{X_n}(x) \leq F_X(x + \epsilon).$$

Finalmente, como F_X es continua en x, se tiene que:

$$F_X(x - \epsilon) \to F_X(x), \quad \text{y} \quad F_X(x + \epsilon) \to F_X(x) \quad \text{cuando } \epsilon \to 0,$$

y por lo tanto, al hacer $\epsilon \to 0$, se obtiene:

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x).$$

Existen contraejemplos donde $X_n \xrightarrow{d} X$, pero no $X_n \xrightarrow{\mathbb{P}} X$ si X es una variable aleatoria no constante, por ejemplo, sea $X_n \sim \text{Bernoulli}\left(\frac{1}{2}\right)$ para todo n, y sea $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$, independiente de la secuencia $\{X_n\}$. Entonces $X_n \xrightarrow{d} X$, ya que todas tienen la misma distribución, pero:

$$|X_n - X| \sim \text{Bernoulli}(1/2) \Rightarrow \mathbb{P}(|X_n - X| \geq \epsilon) = \frac{1}{2}, \quad \forall \epsilon \in (0, 1),$$

por lo que la convergencia en probabilidad no se cumple.

9.2.5 Convergencia Casi Segura

La convergencia casi segura es uno de los tipos más fuertes de convergencia entre secuencias de variables aleatorias. En este caso, una secuencia $\{X_n\}_{n\in\mathbb{N}}$ converge casi seguramente a una variable aleatoria X si, para casi todo $\omega \in \Omega$, la sucesión $X_n(\omega)$ converge a $X(\omega)$ como secuencia de números reales. Es decir, salvo en un conjunto de probabilidad cero, los valores de X_n terminan acercándose y permaneciendo arbitrariamente cerca de X a medida que $n \to \infty$.

Definición 9.5 (Convergencia casi segura) Sea $\{X_n\}$ una secuencia de variables aleatorias y X otra variable aleatoria en el mismo espacio de probabilidad. Se dice que X_n converge *casi seguramente* a X si:

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

Se denota por:

$$X_n \xrightarrow{\text{c.s.}} X \quad \text{o bien} \quad X_n \xrightarrow{\text{a.s.}} X.$$

Esta definición garantiza que la secuencia converge punto a punto, excepto quizás en un conjunto nulo. A diferencia de la convergencia en probabilidad, aquí se requiere convergencia de la trayectoria para casi todo ω.

Ejemplo:

Sea $S = \{\text{C}, \text{S}\}$ un espacio muestral finito, y definimos la secuencia:

$$X_n(\omega) = \begin{cases} \frac{n}{n+1}, & \text{si } \omega = \text{C}, \\ (-1)^n, & \text{si } \omega = \text{S}. \end{cases}$$

Entonces, para $\omega = \text{C}$, se tiene $X_n(\omega) \to 1$; mientras que para $\omega = \text{S}$ la secuencia oscila y no converge. Si asignamos $\mathbb{P}(\{\text{C}\}) = 1$, entonces:

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = 1\right\}\right) = 1,$$

por lo tanto, $X_n \xrightarrow{\text{a.s.}} 1$.

Teorema 9.11 Condición suficiente para convergencia casi segura

Sea $\{X_n\}$ una secuencia de variables aleatorias. Si para todo $\epsilon > 0$ se cumple:

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \epsilon) < \infty,$$

entonces:

$$X_n \xrightarrow{\text{a.s.}} X.$$

La demostración se escapa de los contenidos del curso, pues se basa en el lema de Borel–Cantelli. Para todo $\epsilon > 0$, definimos:

$$A_n = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\}.$$

Si la suma de las probabilidades de estos eventos es finita, entonces, por Borel–Cantelli:

$$\mathbb{P}(A_n \text{ ocurre infinitas veces}) = 0,$$

lo que significa que $|X_n - X| > \epsilon$ sólo ocurre un número finito de veces, casi seguramente. Como esto vale para todo $\epsilon > 0$, se deduce que $X_n \to X$ punto a punto casi seguramente.

Ejemplo:
Sea X_n definido por:

$$X_n = \begin{cases} \frac{-1}{n}, & \text{con probabilidad } \frac{1}{2}, \\ \frac{1}{n}, & \text{con probabilidad } \frac{1}{2}. \end{cases}$$

Entonces $X_n \xrightarrow{\mathbb{P}} 0$. Además, para todo $\epsilon > 0$, se tiene:

$$\mathbb{P}(|X_n - 0| > \epsilon) = \begin{cases} 1, & \text{si } \epsilon < \frac{1}{n}, \\ 0, & \text{si } \epsilon \geq \frac{1}{n}. \end{cases}$$

Por tanto, para ϵ fijo, hay finitos n tales que $\frac{1}{n} > \epsilon$, y se concluye que la suma $\sum_n \mathbb{P}(|X_n - 0| > \epsilon) < \infty$, por lo que:

$$X_n \xrightarrow{\text{a.s.}} 0.$$

Teorema 9.12 Caract. alternativa de convergencia casi segura

Sea $\{X_n\}$ una secuencia de variables aleatorias, y sea X una variable aleatoria. Para $\epsilon > 0$, definimos:

$$A_m = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \epsilon, \quad \forall n \geq m\}.$$

Entonces:

$$X_n \xrightarrow{\text{a.s.}} X \quad \text{si y sólo si} \quad \lim_{m \to \infty} \mathbb{P}(A_m) = 1, \quad \forall \epsilon > 0.$$

Este resultado ofrece una interpretación equivalente que a menudo resulta más útil para pruebas directas. Al igual que el teorema anterior, la demostración de este teorema se escapa de los contenidos del curso, sin embargo es visto en Modelos Estocásticos en Sistemas de Ingeniería.

Teorema 9.13 Conv. casi segura implica conv. en probabilidad

$$\text{Si } X_n \xrightarrow{\text{a.s.}} X, \text{ entonces } X_n \xrightarrow{\mathbb{P}} X.$$

9.3 Teoremas Límite

En esta sección se presentan dos resultados fundamentales en teoría de la probabilidad: la *ley de los grandes números* (LLN, por sus siglas en inglés) y el *teorema central del límite* (CLT). Ambos teoremas describen el comportamiento asintótico de secuencias de variables aleatorias.

La ley de los grandes números establece que el promedio de un número creciente de variables aleatorias independientes e idénticamente distribuidas converge al valor esperado. Por otro lado, el teorema central del límite afirma que, bajo condiciones adecuadas, la suma normalizada de muchas variables aleatorias tiende en distribución a una variable normal. Estos resultados proporcionan la base teórica para justificar inferencias estadísticas y aproximaciones en contextos reales donde se trabaja con grandes muestras.

9.3.1 Ley Débil de los Grandes Números (WLLN)

La Ley Débil de los Grandes Números establece que el promedio de una gran cantidad de variables aleatorias independientes e idénticamente distribuidas con esperanza finita converge en probabilidad al valor esperado. Esta ley garantiza que, a medida que el tamaño de la muestra crece, el promedio observado se aproxima al valor teórico esperado.

Teorema 9.14 Ley Débil de los Grandes Números

Sea $\{X_i\}_{i=1}^{\infty}$ una secuencia de variables aleatorias independientes e idénticamente distribuidas (i.i.d.) tales que $\mathbb{E}[X_i] = \mu < \infty$ y $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Denotando la media muestral por

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

se cumple que:

$$\overline{X}_n \xrightarrow{\mathbb{P}} \mu, \quad \text{es decir,} \quad \lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| \geq \epsilon) = 0, \quad \forall \epsilon > 0.$$

Demostración: Como las variables son i.i.d., todas tienen la misma varianza $\mathrm{Var}(X_i) = \sigma^2 < \infty$. La media muestral $\overline{X}_n$ tiene esperanza:

$$\mathbb{E}[\overline{X}_n] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n} n\mu = \mu,$$

y su varianza es:

$$\mathrm{Var}(\overline{X}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Aplicando la desigualdad de Chebyshev:

$$\mathbb{P}(|\overline{X}_n - \mu| \geq \epsilon) \leq \frac{\mathrm{Var}(\overline{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}.$$

Dado que el lado derecho tiende a cero cuando $n \to \infty$, se concluye que:

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| \geq \epsilon) = 0, \quad \forall \epsilon > 0,$$

lo cual prueba que $\overline{X}_n \xrightarrow{\mathbb{P}} \mu$.

Ejemplo:

Supongamos que lanzamos una moneda justa n veces y definimos $X_i = 1$ si la i-ésima tirada resulta cara, y $X_i = 0$ si resulta cruz. Entonces $X_i \sim$ Bernoulli(1/2), por lo que:

$$\mathbb{E}[X_i] = \frac{1}{2}, \quad \text{Var}(X_i) = \frac{1}{4}.$$

La media muestral representa la proporción de caras observadas:

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

La Ley Débil de los Grandes Números garantiza que:

$$\overline{X}_n \xrightarrow{\mathbb{P}} \frac{1}{2},$$

lo que significa que, a medida que n crece, la proporción observada de caras se aproxima en probabilidad a 0,5.

9.3.2 Ley Fuerte de los Grandes Números (SLLN)

La Ley Fuerte de los Grandes Números establece que, bajo ciertas condiciones, la media muestral de una secuencia de variables aleatorias independientes e idénticamente distribuidas converge casi seguramente a la esperanza. Esto significa que, con probabilidad uno, los valores promedio se aproximan al valor esperado a medida que el tamaño de la muestra tiende a infinito.

Teorema 9.15 Ley Fuerte de los Grandes Números

Sea $\{X_i\}_{i=1}^{\infty}$ una secuencia de variables aleatorias independientes e idénticamente distribuidas (i.i.d.) con esperanza finita $\mathbb{E}[X_i] = \mu < \infty$ y varianza finita $\text{Var}(X_i) = \sigma^2 < \infty$. Definiendo la media muestral como

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

se cumple que:

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n\to\infty} \overline{X}_n(\omega) = \mu\right\}\right) = 1,$$

es decir, $\overline{X}_n \xrightarrow{\text{c.s.}} \mu$.

La demostración de este resultado se basa en elementos que exceden el alcance del presente curso, tales como el criterio de convergencia casi segura de series de variables independientes centradas. Por lo tanto, se omite su desarrollo formal.

Ejemplo:

Supongamos nuevamente $X_i \sim \text{Bernoulli}(p)$ con $0 < p < 1$. Entonces $\mathbb{E}[X_i] = p$ y $\text{Var}(X_i) = p(1-p) < \infty$. La media muestral $\overline{X}_n$ representa la proporción de éxitos observados en n repeticiones. La Ley Fuerte de los Grandes Números garantiza que:

$$\overline{X}_n \xrightarrow{\text{a.s.}} p.$$

Esto implica que, con probabilidad uno, la proporción de éxitos observados converge al valor esperado p.

9.3.3 Teorema Central del Límite (TCL)

El Teorema Central del Límite es uno de los resultados fundamentales en teoría de la probabilidad. Este teorema establece que, bajo condiciones generales, la suma (o promedio) de un número suficientemente grande de variables aleatorias independientes e idénticamente distribuidas con media y varianza finitas se aproxima a una distribución normal, independientemente de la distribución original de dichas variables.

Teorema 9.16 Teorema Central del Límite

Sean $\{X_i\}_{i=1}^n$ variables aleatorias i.i.d. con esperanza $\mathbb{E}[X_i] = \mu < \infty$ y varianza $\text{Var}(X_i) = \sigma^2 < \infty$. Definimos la media muestral:

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

y la variable aleatoria estandarizada:

$$Z_n = \frac{(\overline{X}_n - \mu)}{\frac{\sigma}{\sqrt{n}}} = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n} X_i - n\mu.$$

Entonces, Z_n converge en distribución a una variable normal estándar $N(0,1)$:

$$Z_n \xrightarrow{d} Z, \quad \text{donde } Z \sim \mathcal{N}(0,1),$$

lo que equivale a:

$$\lim_{n\to\infty} \mathbb{P}(Z_n \leq x) = \Phi(x), \quad \forall x \in \mathbb{R},$$

donde $\Phi(x)$ es la función de distribución acumulada de la normal estándar.

La justificación del TCL se basa en la idea de que, aunque las X_i no tengan distribución normal, la acumulación de efectos aleatorios individuales tiende a producir un comportamiento global aproximadamente gaussiano. Esta propiedad es consecuencia del comportamiento de la función característica de la suma estandarizada, que converge puntualmente a la de una normal estándar.

La normalización es clave: al centrar las variables en su media y escalarlas por su desviación estándar σ y por $\sqrt{n}$, se obtiene una variable con esperanza cero y

varianza unitaria:

$$\mathbb{E}[Z_n] = 0, \qquad \mathrm{Var}(Z_n) = 1.$$

Aplicación práctica del TCL

Para utilizar el Teorema Central del Límite en un problema aplicado, se siguen los siguientes pasos:

1. Identificar la suma o el promedio muestral de interés, por ejemplo: $Y_n = \sum_{i=1}^{n} X_i$.
2. Calcular su esperanza y varianza:

$$\mathbb{E}[Y_n] = n\mu, \qquad \mathrm{Var}(Y_n) = n\sigma^2.$$

3. Estandarizar la variable:

$$Z_n = \frac{Y_n - \mathbb{E}[Y_n]}{\sqrt{\mathrm{Var}(Y_n)}}.$$

4. Usar que $Z_n \xrightarrow{d} \mathcal{N}(0,1)$ para aproximar probabilidades:

$$\mathbb{P}(a \leq Y_n \leq b) \approx \mathbb{P}\left(\frac{a - n\mu}{\sqrt{n}\sigma} \leq Z \leq \frac{b - n\mu}{\sqrt{n}\sigma}\right).$$

Ejemplo:

Sea $X_i \sim \text{Bernoulli}(p)$, con $0 < p < 1$, y sean i.i.d. Entonces $\mu = p$, $\sigma^2 = p(1-p)$. Si definimos $S_n = \sum_{i=1}^{n} X_i$, entonces:

$$Z_n = \frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} \mathcal{N}(0,1).$$

Esto permite aproximar la distribución binomial con una normal estándar para valores grandes de n.