

Introducción a los Modelos Probabilísticos (PSG-01M47)

Apuntes de Clase

Alejandro Verri Kozlowski

Table of contents

Resumen	7
1 Espacios de Probabilidad	8
1.1 Introducción	8
1.2 Sigma-Álgebras y Preliminares de Teoría de la Medida	8
1.3 Espacio de Probabilidad (Definición Formal)	8
1.4 Variables Aleatorias	9
1.4.1 Medibilidad e Imagen Inversa	9
1.5 Distribuciones de Probabilidad de Variables Aleatorias	9
1.5.1 Función de Distribución Acumulada (CDF)	9
1.5.2 Distribuciones Discretas y Continuas	10
1.5.3 Distribuciones Mixtas	10
1.6 Esperanza Matemática	10
1.6.1 Linealidad y Convergencia Monótona	10
1.7 Varianza y Desviación Estándar	11
1.8 Probabilidad Condicional	11
1.9 Teoremas Fundamentales	11
1.9.1 Ley de Probabilidad Total	11
1.9.2 Teorema de Bayes	11
1.10 Ejercicios	12
2 Técnicas de Conteo en Probabilidad y Combinatoria	13
2.1 Introducción	13
2.2 Regla del Producto	13
2.2.1 Definición Formal	13
2.2.2 Demostración	13
2.3 Permutaciones	14
2.3.1 Definición	14
2.3.2 Demostración	14
2.4 k -Permutaciones (anteriormente “Variaciones”)	15
2.4.1 Terminología y Definición	15
2.4.2 Demostración	15
2.4.3 Restricciones	15
2.4.4 Distinción Conceptual	15
2.5 Combinaciones	15
2.5.1 Definición	15
2.5.2 Soporte: Probabilidad Hipergeométrica	16
2.6 Permutaciones con Elementos Repetidos	16
2.6.1 Definición	16

2.6.2	Demostración	16
2.7	Modelo de Maxwell–Boltzmann	17
2.7.1	Contexto Físico y Supuestos	17
2.7.2	Fórmula y Derivación	17
2.8	Modelo de Bose–Einstein	17
2.8.1	Contexto Físico y Supuestos	17
2.8.2	Fórmula y Demostración	17
2.9	Distribuciones Derivadas de Maxwell–Boltzmann: Binomial y Poisson	18
2.9.1	Derivación de la Distribución Binomial	18
2.9.2	Límite de Poisson	18
2.10	Distribuciones Derivadas de Bose–Einstein: Geométrica	18
2.10.1	Distribución Geométrica de Ocupación	18
2.10.2	Derivación	18
2.11	Ejercicios	19
3	Probabilidad Condicional e Independencia	20
3.1	Probabilidad Condicional	20
3.1.1	Extensión: Probabilidad Condicional con Respecto a una σ -Álgebra . . .	20
3.1.2	Medida de Probabilidad Condicional	20
3.1.3	Demostración: Propiedades de la Probabilidad Condicional	21
3.2	Ley de Probabilidad Total	21
3.2.1	Demostración	21
3.3	Teorema de Bayes	21
3.3.1	Demostración	22
3.3.2	Interpretación Bayesiana	22
3.4	Independencia de Eventos y Variables Aleatorias	22
3.4.1	Independencia de Eventos	22
3.4.2	Independencia de Variables Aleatorias	22
3.4.3	Independencia con Respecto a σ -Álgebras	23
3.5	Ejercicios	23
4	Variables Aleatorias	24
4.1	Definición y Propiedades	24
4.1.1	Propiedades de la CDF	24
4.2	Variables Aleatorias Discretas	24
4.3	Variables Aleatorias Continuas	25
4.4	Transformaciones de Variables Aleatorias	25
4.5	Vectores Aleatorios	26
4.6	Independencia de Variables Aleatorias	27
4.7	Esperanza y Varianza	27
4.8	Simulación de Variables Aleatorias	27
4.9	Ejercicios	28
5	Variables Aleatorias y Vectores: Transformaciones y Aplicaciones	29
5.1	Distribuciones de Mezcla	29
5.2	Transformaciones de Variables Aleatorias	30
5.2.1	Teorema de Transformación	31

5.3	Simulación vía Transformación Inversa	31
5.4	Vectores Aleatorios	31
5.5	Independencia	32
5.6	Transformaciones de Vectores Aleatorios	32
5.6.1	Transformaciones Lineales	32
5.6.2	Transformaciones No Lineales	33
5.7	Aplicaciones en Análisis Multivariado	34
5.7.1	Análisis de Componentes Principales (PCA)	34
5.7.2	Análisis de Correlación Canónica	34
5.7.3	Análisis Factorial	34
5.8	Ejercicios	34
6	Distribuciones Condicionales y Predicción	35
6.1	Distribuciones Condicionales	35
6.1.1	Propiedades	35
6.2	Teoría de Predicción	36
6.2.1	Propiedades	36
6.3	Predicción Lineal	37
6.3.1	Caso Especial: Normal Bivariado	37
6.4	Aplicaciones	37
6.4.1	Análisis de Regresión	37
6.4.2	Pronóstico de Series de Tiempo	38
6.4.3	Aprendizaje Automático	38
6.5	Ejercicios	38
7	Ley de los Grandes Números y Teorema Central del Límite	39
7.1	Ley de los Grandes Números	39
7.2	Teorema Central del Límite	40
7.3	Aproximaciones del Teorema Central del Límite	41
8	El Proceso de Bernoulli: Definición, Propiedades y Aplicaciones	43
8.1	Definición Formal de Procesos Estocásticos	43
8.2	El Proceso de Bernoulli	43
8.2.1	Construcción Rigurosa	43
8.2.2	Distribución Conjunta	44
8.3	Variables Aleatorias Asociadas	44
8.3.1	Número Total de Éxitos en n Ensayos	44
8.3.2	Tiempo de Espera para el Primer Éxito	44
8.3.3	Tiempo de Espera para el k -ésimo Éxito	45
8.4	Propiedades Clave	45
8.5	Conexiones con Otras Distribuciones	46
8.5.1	Aproximación de Poisson a la Binomial	46
8.5.2	Generalización Multinomial	46
8.6	Ejercicios	47

9	Procesos de Poisson	48
9.1	Definición y Propiedades Fundamentales	48
9.1.1	Demostración Formal de la Ley de Distribución de Poisson	48
9.1.2	Valores Enteros	48
9.2	Tiempos entre Llegadas y la Ley Exponencial	49
9.2.1	Demostración de la Propiedad de Falta de Memoria	49
9.3	Proceso de Poisson Compuesto	49
9.4	Aplicaciones	50
9.4.1	Teoría de Colas	50
9.4.2	Teoría de Confiabilidad	50
9.4.3	Seguros y Riesgo	50
9.5	Propiedades Avanzadas	50
9.5.1	Teorema de Superposición	50
9.5.2	Teorema de Adelgazamiento	51
9.6	Ejercicios	51
10	Procesos de Markov	52
10.1	Introducción Conceptual	52
10.2	Matrices de Transición	52
10.3	Clasificación de Estados y Recurrencia	53
10.4	Irreducibilidad y Periodicidad	53
10.5	Distribuciones Estacionarias	53
10.5.1	Teorema de Existencia y Unicidad	54
10.6	Ergodicidad y Convergencia	54
10.7	Ecuación de Chapman–Kolmogorov	54
10.8	Reversibilidad	55
10.9	Ejercicios	56
11	Procesos de Poisson Compuestos	57
11.1	Introducción Conceptual	57
11.2	Definición Formal y Propiedades Fundamentales	57
11.2.1	Incrementos Independientes y Estacionaridad	58
11.3	Momentos del Proceso de Poisson Compuesto	58
11.3.1	Esperanza	58
11.3.2	Varianza	58
11.4	Representación Distribucional y Métodos de Transformadas	59
11.5	Ejercicios	60
12	Serie de Tiempo	61
12.1	Definiciones Fundamentales	61
12.2	Descomposición de Series de Tiempo	61
12.3	Estacionaridad	62
12.4	Ruido Blanco	62
12.4.1	Demostración: El Ruido Blanco es Débilmente Estacionario	63
12.5	Métodos de Preprocesamiento	63
12.6	Caminata Aleatoria	63
12.7	Modelos ARIMA	64

12.8	Proceso Autoregresivo (AR)	64
12.8.1	Demostración: Condición de Estacionaridad	64
12.9	Proceso de Promedio Móvil (MA)	65
12.9.1	Demostración: Autocovarianza de $MA(q)$	65
12.9.2	Ejemplo: Ajuste de Modelo $MA(1)$	65
12.10	Extensiones de Modelos ARIMA	65
12.11	Ejercicios	66
13	Modelos ARIMA: Fundamentos, Teoría y Aplicación	67
13.1	Estacionaridad	67
13.2	Ruido Blanco	67
13.3	Caminatas Aleatorias y Raíces Unitarias	68
13.4	Preprocesamiento y Suavizado	68
13.5	Diferenciación para Estacionaridad	69
13.6	Descomposición Aditiva y Multiplicativa	69
13.7	El Modelo ARIMA: Definición y Propiedades	70
13.8	Modelos Autoregresivos (AR)	70
13.9	Modelos de Media Móvil (MA)	71
13.10	Procesos Integrados	71
13.11	Ejercicios	71

Resumen

Las siguientes páginas corresponden a las notas de clase del curso de [Introducción a los Modelos Probabilísticos \(PSG-01M47\)](#) dictado por la Dra. Verónica Estela Pastor, durante el primer semestre de 2025. Los contenidos se han organizado de acuerdo al orden cronológico de las clases.

1 Espacios de Probabilidad

1.1 Introducción

Un espacio de probabilidad proporciona la base matemática formal para modelar fenómenos aleatorios. Esta estructura permite el razonamiento riguroso sobre la incertidumbre al definir un marco consistente para eventos, resultados y sus probabilidades asociadas. Este capítulo presenta definiciones formales, propiedades esenciales y resultados fundamentales que subyacen a la teoría de probabilidades moderna.

1.2 Sigma-Álgebras y Preliminares de Teoría de la Medida

Sea Ω un conjunto no vacío llamado **espacio muestral**, que representa todos los posibles resultados de un experimento aleatorio. Una **sigma-álgebra** \mathcal{F} en Ω es una colección de subconjuntos de Ω (llamados **eventos**) que satisface:

1. $\Omega \in \mathcal{F}$,
2. Si $A \in \mathcal{F}$, entonces $A^c \in \mathcal{F}$,
3. Si $A_n * n = 1^\infty \subset \mathcal{F}$, entonces $\bigcup * n = 1^\infty A_n \in \mathcal{F}$.

La clausura bajo intersecciones numerables se sigue de las leyes de De Morgan y las propiedades 2–3.

1.3 Espacio de Probabilidad (Definición Formal)

Un **espacio de probabilidad** es una terna (Ω, \mathcal{F}, P) , donde:

- Ω es el espacio muestral,
- \mathcal{F} es una sigma-álgebra de subconjuntos de Ω ,
- $P : \mathcal{F} \rightarrow [0, 1]$ es una **medida de probabilidad**, que satisface:
 1. $P(\Omega) = 1$,
 2. $P(\emptyset) = 0$,
 3. $P(A) \geq 0$ para todo $A \in \mathcal{F}$,
 4. Para cualquier sucesión numerable de eventos disjuntos por pares $A_{n=1}^\infty \subset \mathcal{F}$,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Este sistema axiomático es fundamental y generaliza las asignaciones intuitivas de probabilidad a espacios complejos, posiblemente no numerables [Billingsley1995].

1.4 Variables Aleatorias

Una **variable aleatoria** es una función medible $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, donde $\mathcal{B}(\mathbb{R})$ denota la sigma-álgebra de Borel de \mathbb{R} . Es decir, para todo conjunto de Borel $B \subseteq \mathbb{R}$,

$$X^{-1}(B) \in \mathcal{F}.$$

Esta propiedad (medibilidad) asegura que la probabilidad de cualquier evento concerniente a X esté bien definida [Durrett2019].

1.4.1 Medibilidad e Imagen Inversa

Una función X es **medible** si, para todo conjunto de Borel B , la imagen inversa $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ pertenece a \mathcal{F} .

1.5 Distribuciones de Probabilidad de Variables Aleatorias

La **distribución** de una variable aleatoria X se define como la medida imagen $P_X(B) = P(X^{-1}(B))$ para $B \in \mathcal{B}(\mathbb{R})$.

1.5.1 Función de Distribución Acumulada (CDF)

La **función de distribución acumulada** de X es

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

1.5.2 Distribuciones Discretas y Continuas

- **Discreta:** Si X toma a lo sumo valores numerables, definimos la **función de masa de probabilidad (pmf)**

$$p_X(x) = P(X = x), \quad \sum_{x \in \text{Range}(X)} p_X(x) = 1.$$

- **Continua:** Si existe una función f_X tal que

$$P(X \in B) = \int_B f_X(x) dx$$

para todos los conjuntos de Borel B , entonces f_X es la **función de densidad de probabilidad (pdf)** de X . Debe satisfacer $f_X(x) \geq 0$ y

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

1.5.3 Distribuciones Mixtas

Las variables aleatorias también pueden tener **distribuciones mixtas** que no son ni puramente discretas ni puramente continuas. Tales casos surgen, por ejemplo, en distribuciones con átomos y una componente absolutamente continua [Billingsley1995].

1.6 Esperanza Matemática

Sea X una variable aleatoria en (Ω, \mathcal{F}, P) .

- La **esperanza matemática** (valor esperado) se define como

$$E[X] = \begin{cases} \sum_{x \in \text{Range}(X)} x p_X(x), & \text{si } X \text{ es discreta,} \\ \int_{-\infty}^{\infty} x f_X(x) dx, & \text{si } X \text{ es continua.} \end{cases}$$

siempre que $E[|X|] < \infty$.

1.6.1 Linealidad y Convergencia Monótona

La esperanza es lineal: Para X, Y integrables y constantes a, b ,

$$E[aX + bY] = aE[X] + bE[Y].$$

Si $0 \leq X_n \uparrow X$, entonces $E[X_n] \uparrow E[X]$ (Teorema de Convergencia Monótona) [Durrett2019].

1.7 Varianza y Desviación Estándar

La **varianza** de X (con $E[X^2] < \infty$) es

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

La **desviación estándar** es $\sigma_X = \sqrt{\text{Var}(X)}$.

1.8 Probabilidad Condicional

Para $A, B \in \mathcal{F}$ con $P(B) > 0$, la **probabilidad condicional** de A dado B es

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

1.9 Teoremas Fundamentales

1.9.1 Ley de Probabilidad Total

Sea $A_{i=1}^n$ una partición finita o numerable de Ω con $P(A_i) > 0$ para todo i . Para cualquier $B \in \mathcal{F}$,

$$P(B) = \sum_i P(B|A_i)P(A_i).$$

Demostración. Como los A_i son disjuntos y $\bigcup_i A_i = \Omega$,

$$P(B) = P(B \cap \Omega) = P\left(B \cap \bigcup_i A_i\right) = \sum_i P(B \cap A_i) = \sum_i P(B|A_i)P(A_i).$$

1.9.2 Teorema de Bayes

Para eventos $A, B \in \mathcal{F}$ con $P(B) > 0$,

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

Demostración. Por la definición de probabilidad condicional,

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

Ejemplo Discreto: Suma de Dos Dados

Sea $\Omega = (i, j) : i, j \in 1, \dots, 6$ y definamos $X(i, j) = i + j$. Entonces $\text{Range}(X) = 2, 3, \dots, 12$.

Para $x \in \text{Range}(X)$,

$$p_X(x) = P(X = x) = \frac{\# \text{ de pares con suma } x}{36}.$$

Para $x = 7$, hay 6 pares; así $p_X(7) = 6/36 = 1/6$.

Calcular la esperanza:

$$E[X] = \sum_{x=2}^{12} x \cdot p_X(x) = 7.$$

Calcular la varianza:

$$\text{Var}(X) = \sum_{x=2}^{12} (x - 7)^2 p_X(x) = \frac{35}{6}.$$

Ejemplo Continuo: Distribución Exponencial

Sea X con pdf $f_X(x) = \lambda e^{-\lambda x}$ para $x \geq 0$ ($\lambda > 0$).

- $F_X(x) = 1 - e^{-\lambda x}$ para $x \geq 0$,
- $E[X] = \int_0^\infty x \lambda e^{-\lambda x} dx = 1/\lambda$,
- $\text{Var}(X) = \int_0^\infty (x - 1/\lambda)^2 \lambda e^{-\lambda x} dx = 1/\lambda^2$.

1.10 Ejercicios

1. **Construcción de Espacio de Probabilidad:** Dado $\Omega = [0, 1]$, sea \mathcal{F} la sigma-álgebra de Borel, y defina P como la medida de Lebesgue. Demuestre que (Ω, \mathcal{F}, P) es un espacio de probabilidad.
2. **Verificación de Sigma-Álgebra:** Demuestre que la intersección de una colección numerable de sigma-álgebras en Ω es en sí misma una sigma-álgebra.
3. **Aplicación del Teorema:** Para A_1, A_2, \dots, A_n mutuamente excluyentes con $\sum_{i=1}^n P(A_i) = 1$ y un evento B con $P(B) > 0$, derive $P(A_j|B)$ usando el teorema de Bayes.
4. **Cálculo Avanzado:** Sea $X \sim \text{Exp}(\lambda)$. Demuestre que $E[X^k] = k!/\lambda^k$ para entero $k \geq 1$.

2 Técnicas de Conteo en Probabilidad y Combinatoria

2.1 Introducción

Las técnicas de conteo son herramientas fundamentales en la teoría de probabilidad y combinatoria, proporcionando métodos rigurosos para enumerar el número de formas en que pueden ocurrir eventos. Este capítulo desarrolla principios clave—reglas de producto, permutaciones (incluyendo elementos repetidos), k -permutaciones y combinaciones—enfaticando definiciones precisas, demostraciones formales y los fundamentos combinatorios de las distribuciones básicas de probabilidad. Toda la notación se introduce según sea necesario, y la terminología sigue convenciones matemáticas internacionales.

2.2 Regla del Producto

2.2.1 Definición Formal

Sean A_1, A_2, \dots, A_n conjuntos finitos que representan los posibles resultados de n experimentos independientes. La regla del producto establece que el número de tuplas ordenadas (a_1, a_2, \dots, a_n) con $a_i \in A_i$ está dado por:

$$|A_1 \times A_2 \times \dots \times A_n| = |A_1| \cdot |A_2| \cdots |A_n|$$

donde $|A|$ denota la cardinalidad del conjunto A [Billingsley1995].

2.2.2 Demostración

Puesto que los experimentos son independientes, para cada elección $a_1 \in A_1$, hay $|A_2|$ opciones para a_2 , y así sucesivamente. Por inducción:

- Caso base $n = 1$: $|A_1|$.
- Paso inductivo: Asumamos que es verdadero para $n = k$. Para $n = k + 1$,

$$|A_1 \times \dots \times A_k \times A_{k+1}| = (|A_1| \cdots |A_k|) \cdot |A_{k+1}|.$$

Por tanto, la regla se cumple para todo $n \in \mathbb{N}$.

Ejemplo: Supongamos que una urna contiene 6 bolas rojas y 4 negras (10 en total). Se extraen dos bolas en secuencia **sin reemplazo**. Calculemos el número total de resultados ordenados:

Sea A_1 = conjunto de bolas para la primera extracción ($|A_1| = 10$); A_2 = conjunto de bolas para la segunda extracción ($|A_2| = 9$ puesto que una ha sido removida). El número total de pares ordenados es $10 \times 9 = 90$.

Supongamos que queremos el número de formas en que ambas bolas sean rojas. Hay 6 opciones para la primera bola roja y 5 para la segunda (puesto que no hay reemplazo). Por tanto, $6 \times 5 = 30$ resultados favorables ordenados.

La probabilidad de que ambas sean rojas:

$$P(\text{ambas rojas}) = \frac{30}{90} = \frac{1}{3}$$

Ejemplo Avanzado: Conexión con Espacios de Probabilidad: Si en cambio las bolas fueran extraídas **con reemplazo** (cada extracción independiente), el número total de pares ordenados sería $10 \times 10 = 100$, ilustrando el efecto de la independencia.

2.3 Permutaciones

2.3.1 Definición

Una **permutación** de un conjunto finito S con $|S| = n$ es una biyección $\pi : S \rightarrow S$, o equivalentemente, un arreglo ordenado de todos los n elementos [Stanley2012]. El número de permutaciones es:

$$n! = n \cdot (n - 1) \cdots 2 \cdot 1$$

2.3.2 Demostración

Cada posición en el arreglo ordenado tiene un elemento único asignado, con n opciones para la primera, $n - 1$ para la segunda, ..., 1 para la última.

Ejemplo. Arreglar 5 libros distintos en un estante. El número de arreglos es $5! = 120$.

Ejemplo Avanzado: Paridad de Permutaciones: Para $n \geq 2$, la mitad de todas las permutaciones son pares y la mitad son impares, un resultado crucial en la teoría de determinantes y estructuras algebraicas [Durrett2019].

2.4 k -Permutaciones (anteriormente “Variaciones”)

2.4.1 Terminología y Definición

Una **k -permutación** (inglés estándar) de un conjunto de n elementos es una selección ordenada de k elementos distintos ($0 \leq k \leq n$). El número es:

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

2.4.2 Demostración

Para la primera posición, n opciones; segunda, $n - 1$; hasta k posiciones:

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

Ejemplo: Seleccionar y ordenar 3 de 5 libros: $P(5, 3) = 60$.

2.4.3 Restricciones

Si $k > n$, $P(n, k) = 0$ puesto que la selección sin repetición es imposible.

2.4.4 Distinción Conceptual

- Permutaciones: Todos los n elementos, el orden importa.
- k -Permutaciones: Subconjunto de k elementos, el orden importa.
- Combinaciones: Subconjunto de k elementos, el orden **no** importa.

2.5 Combinaciones

2.5.1 Definición

Una **combinación** es una selección no ordenada de k elementos de un conjunto de n elementos:

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

La notación $\binom{n}{k}$ es estándar.

2.5.2 Soporte: Probabilidad Hipergeométrica

Al muestrear k elementos sin reemplazo de una población de n objetos, de los cuales m son “especiales” (ej., defectuosos), la probabilidad de que exactamente s sean especiales está dada por la **distribución hipergeométrica**:

$$P(s \text{ especiales}) = \frac{\binom{m}{s} \binom{n-m}{k-s}}{\binom{n}{k}}$$

[@Feller1970]

Ejemplo: De 50 componentes (4 defectuosos), seleccionar 10 al azar. La probabilidad de que **al menos uno** sea defectuoso:

$$P(\geq 1 \text{ defectuoso}) = 1 - \frac{\binom{46}{10}}{\binom{50}{10}}$$

Ejemplo (Escala Pequeña) De 6 objetos (2 defectuosos), elegir 3:

- Formas totales: $\binom{6}{3} = 20$
- Formas sin defectuosos: $\binom{4}{3} = 4$
- $P(\text{al menos uno defectuoso}) = 1 - \frac{4}{20} = 0.8$

2.6 Permutaciones con Elementos Repetidos

2.6.1 Definición

Dado un multiconjunto de n objetos donde n_1 son del tipo 1, n_2 del tipo 2, ..., n_k del tipo k (con $n_1 + \dots + n_k = n$), el número de permutaciones distintas es:

$$\frac{n!}{n_1! n_2! \dots n_k!}$$

2.6.2 Demostración

Hay $n!$ formas de permutar todos los objetos si fueran distinguibles. Para cada grupo de elementos indistinguibles, dividir por $n_i!$ corrige el sobreconteo debido a arreglos repetidos [@Stanley2012].

Ejemplo: Número de anagramas distintos de “BANANA” (3 A’s, 2 N’s, 1 B):

$$\frac{6!}{3!2!1!} = \frac{720}{12} = 60$$

2.7 Modelo de Maxwell–Boltzmann

2.7.1 Contexto Físico y Supuestos

El modelo de Maxwell–Boltzmann describe el número de formas de distribuir r partículas **distinguibles** en n cajas **distinguibles** (“urnas”), sin restricción en la ocupación [Feller1970]. Cada partícula se asigna independientemente a una caja.

2.7.2 Fórmula y Derivación

Para cada una de r partículas, hay n opciones de caja:

$$N = n^r$$

Ejemplo. Distribuir 3 bolas etiquetadas en 2 cajas: $2^3 = 8$ configuraciones.

2.8 Modelo de Bose–Einstein

2.8.1 Contexto Físico y Supuestos

El modelo de Bose–Einstein aborda la distribución de r partículas **indistinguibles** en n cajas **distinguibles**, permitiendo cualquier número de partículas por caja [Feller1970].

2.8.2 Fórmula y Demostración

Las configuraciones corresponden a soluciones de $x_1 + x_2 + \cdots + x_n = r$ con $x_i \geq 0$, contadas por:

$$N = \binom{r + n - 1}{r}$$

Esto se sigue del teorema de “estrellas y barras”.

Ejemplo: Colocar 4 bolas indistinguibles en 3 cajas:

$$\binom{4 + 3 - 1}{4} = \binom{6}{4} = 15$$

2.9 Distribuciones Derivadas de Maxwell–Boltzmann: Binomial y Poisson

2.9.1 Derivación de la Distribución Binomial

Si cada una de r partículas distinguibles entra independientemente a una de n urnas con probabilidad igual $1/n$, la probabilidad de que una urna fija contenga exactamente k partículas:

$$p(k) = \binom{r}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{r-k}$$

Esta es la **distribución binomial** con parámetros r (ensayos), $p = 1/n$ (probabilidad de éxito).

2.9.2 Límite de Poisson

Cuando $n \rightarrow \infty$, $r \rightarrow \infty$ con $\lambda = r/n$ fijo,

$$p(k) \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}$$

Esta es la **distribución de Poisson** [Billingsley1995].

Ejemplo. Supongamos $r = 10$, $n = 100$, urna fija: $p(0) = \left(1 - \frac{1}{100}\right)^{10} \approx 0.904$.

2.10 Distribuciones Derivadas de Bose–Einstein: Geométrica

2.10.1 Distribución Geométrica de Ocupación

En el modelo de Bose–Einstein, en el límite $n \rightarrow \infty$, $r/n \rightarrow \lambda$, la probabilidad de que una caja fija contenga exactamente k partículas indistinguibles es:

$$p(k) = \frac{\lambda^k}{(1 + \lambda)^{k+1}}$$

2.10.2 Derivación

El resultado se sigue de la combinatoria analítica considerando el límite de ocupación multinomial [Feller1970].

Ejemplo. Sea $\lambda = 2$, $k = 3$:

$$p(3) = \frac{2^3}{(1 + 2)^4} = \frac{8}{81} \approx 0.0988$$

2.11 Ejercicios

1. **Regla del Producto** Sean A , B y C conjuntos con $|A| = 3$, $|B| = 4$, $|C| = 2$. ¿Cuántas triples ordenadas (a, b, c) hay con $a \in A$, $b \in B$, $c \in C$? *Respuesta:* $3 \times 4 \times 2 = 24$
2. **Permutaciones con Elementos Repetidos** ¿Cuántos arreglos únicos hay de las letras en “STATISTICS”?
3. **[Maxwell–Boltzmann/Binomial]** Distribuir r bolas etiquetadas en n cajas. Derivar la probabilidad de que una caja fija contenga exactamente k bolas y analizar la distribución límite cuando $n \rightarrow \infty$ con λ fijo.
4. **[Bose–Einstein/Geométrica]** Demostrar la fórmula para $p(k)$ en el límite geométrico del modelo de Bose–Einstein y calcular $p(0)$ para $\lambda = 1$.

3 Probabilidad Condicional e Independencia

3.1 Probabilidad Condicional

Sea (Ω, \mathcal{A}, P) un espacio de probabilidad. Para eventos $A, B \in \mathcal{A}$ con $P(A) > 0$, la **probabilidad condicional** de B dado A se define por

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

Esta fórmula expresa cómo la probabilidad de B cambia cuando se restringe a la ocurrencia de A .

3.1.1 Extensión: Probabilidad Condicional con Respecto a una σ -Álgebra

Dada una sub- σ -álgebra $\mathcal{G} \subseteq \mathcal{A}$, la **probabilidad condicional de B dado \mathcal{G}** es una variable aleatoria \mathcal{G} -medible $P(B | \mathcal{G})$ tal que, para todo $G \in \mathcal{G}$,

$$\int_G P(B | \mathcal{G}) dP = P(B \cap G).$$

Esta formulación generaliza la probabilidad condicional a contextos que involucran variables aleatorias y estructuras de información [@Billingsley1995].

3.1.2 Medida de Probabilidad Condicional

Para A fijo con $P(A) > 0$, definamos la **medida de probabilidad condicional** $P_A(\cdot)$ en (Ω, \mathcal{A}) por

$$P_A(B) = P(B | A).$$

P_A satisface:

- No negatividad: $P_A(B) \geq 0$ para todo $B \in \mathcal{A}$,
- Normalización: $P_A(\Omega) = 1$,
- Aditividad numerable: $P_A(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} P_A(B_n)$ para cualquier secuencia de conjuntos disjuntos $B_n \in \mathcal{A}$.

3.1.2.1 Existencia (Teorema de Radon–Nikodym)

Para una medida de probabilidad P y $A \in \mathcal{A}$ con $P(A) > 0$, el teorema de Radon–Nikodym asegura la existencia de una medida de probabilidad condicional y, más generalmente, esperanzas condicionales dada una σ -álgebra [@Durrett2019].

3.1.3 Demostración: Propiedades de la Probabilidad Condicional

La *no negatividad* y la *normalización* se siguen directamente de la definición. La *aditividad numerable* se verifica: Sea (B_n) disjunta. Entonces

$$P\left(\bigcup_{n=1}^{\infty} B_n \mid A\right) = \frac{P\left(\bigcup_{n=1}^{\infty} (B_n \cap A)\right)}{P(A)} = \frac{\sum_{n=1}^{\infty} P(B_n \cap A)}{P(A)} = \sum_{n=1}^{\infty} \frac{P(B_n \cap A)}{P(A)} = \sum_{n=1}^{\infty} P(B_n \mid A).$$

3.2 Ley de Probabilidad Total

Sea $A_{i=1}^n$ una partición de Ω (es decir, $A_i \in \mathcal{A}$, $A_i \cap A_j = \emptyset$ para $i \neq j$, y $\bigcup_{i=1}^n A_i = \Omega$) con $P(A_i) > 0$ para todo i . Para cualquier $B \in \mathcal{A}$,

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

3.2.1 Demostración

Por la definición de una partición,

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

3.3 Teorema de Bayes

Para $A_{i=1}^n$ una partición de Ω con $P(A_i) > 0$, y B tal que $P(B) > 0$,

$$P(A_j \mid B) = \frac{P(B \mid A_j)P(A_j)}{\sum_{i=1}^n P(B \mid A_i)P(A_i)}.$$

3.3.1 Demostración

Por la definición de probabilidad condicional y la ley de probabilidad total,

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

3.3.2 Interpretación Bayesiana

El teorema de Bayes permite la actualización de una **probabilidad a priori** $P(A_j)$ a una **probabilidad a posteriori** $P(A_j | B)$ después de observar B . En estadística bayesiana, este mecanismo sustenta la inferencia estadística, con **distribuciones conjugadas a priori** frecuentemente utilizadas para asegurar tractabilidad analítica [@BernardoSmith1994].

3.4 Independencia de Eventos y Variables Aleatorias

3.4.1 Independencia de Eventos

Los eventos A_1, \dots, A_n son **mutuamente independientes** si para todo subconjunto $i_1, \dots, i_k \subseteq 1, \dots, n$,

$$P\left(\bigcap_{j=1}^k A_{i_j}\right) = \prod_{j=1}^k P(A_{i_j}).$$

Independencia por pares significa $P(A_i \cap A_j) = P(A_i)P(A_j)$ para todo $i \neq j$, pero esto **no** garantiza independencia mutua.

3.4.2 Independencia de Variables Aleatorias

Las variables aleatorias X_1, \dots, X_n son **independientes** si para todos los conjuntos de Borel $B_1, \dots, B_n \subseteq \mathbb{R}$,

$$P(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{j=1}^n P(X_j \in B_j).$$

3.4.3 Independencia con Respecto a σ -Álgebras

Las sub- σ -álgebras $\mathcal{G}_1, \dots, \mathcal{G}_n \subseteq \mathcal{A}$ son **independientes** si para todo $G_j \in \mathcal{G}_j$,

$$P\left(\bigcap_{j=1}^n G_j\right) = \prod_{j=1}^n P(G_j).$$

Ejemplo: Independencia por Pares pero No Mutua. Sea (Ω, \mathcal{A}, P) el espacio generado por tres lanzamientos independientes de moneda justa. Definamos los siguientes eventos:

- A : El primer lanzamiento es cara.
- B : El segundo lanzamiento es cara.
- C : El número de caras es par.

Cada par (A, B) , (A, C) , y (B, C) es independiente, pero A, B, C no son mutuamente independientes.

Ejemplo: Independencia en un Espacio de Probabilidad Continuo. Sea (Ω, \mathcal{A}, P) igual a $([0, 1], \mathcal{B}, \lambda)$, donde λ es la medida de Lebesgue. Definamos los eventos:

- $A = [0.1, 0.2)$ (el primer dígito decimal es 1).
- $B = \bigcup_{k=0}^8 [0.01 + 0.1k, 0.02 + 0.1k)$ (el segundo dígito decimal es 1).

Calculamos:

- $P(A) = 0.1$ (longitud del intervalo A),
- $P(B) = 0.1$ (suma de longitudes de intervalos en B),
- $P(A \cap B) = 0.01$ (medida de la intersección).

Por tanto,

$$P(A \cap B) = 0.01 = 0.1 \cdot 0.1 = P(A)P(B),$$

verificando la independencia. Más generalmente, la independencia en espacios continuos puede justificarse rigurosamente usando medidas producto e integración [@Billingsley1995].

3.5 Ejercicios

1. **Probabilidad Condicional como Variable Aleatoria** Sea X una variable aleatoria en (Ω, \mathcal{A}, P) y $\mathcal{G} \subseteq \mathcal{A}$ una sub- σ -álgebra. Demostrar que existe una función \mathcal{G} -medible f tal que, para cualquier $G \in \mathcal{G}$, $\int_G X dP = \int_G f dP$. (*Sugerencia: Teorema de Radon-Nikodym.*)
2. **Independencia Mutua vs. por Pares** Construir tres eventos que sean independientes por pares pero no mutuamente independientes. Demostrar la distinción explícitamente.

4 Variables Aleatorias

4.1 Definición y Propiedades

Sea (Ω, \mathcal{F}, P) un espacio de probabilidad, donde Ω es el espacio muestral, \mathcal{F} es una σ -álgebra de subconjuntos de Ω , y P es una medida de probabilidad en \mathcal{F} . Una **variable aleatoria** es una función medible $X : \Omega \rightarrow \mathbb{R}$, lo que significa que para todo $x \in \mathbb{R}$, el conjunto $\omega \in \Omega : X(\omega) \leq x$ pertenece a \mathcal{F} [Billingsley1995].

La **función de distribución acumulativa (CDF)** de X se define por

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

4.1.1 Propiedades de la CDF

1. F_X es monótonamente no decreciente: Si $x_1 < x_2$, entonces $F_X(x_1) \leq F_X(x_2)$.
2. F_X es continua por la derecha: $\lim_{h \rightarrow 0^+} F_X(x+h) = F_X(x)$ para todo x .
3. Límites: $\lim_{x \rightarrow -\infty} F_X(x) = 0$ y $\lim_{x \rightarrow \infty} F_X(x) = 1$.

Demostración: La monotonidad y la continuidad por la derecha se siguen directamente de la definición de la medida de probabilidad y las propiedades de las σ -álgebras [Billingsley1995, p.14]. Los límites en $\pm\infty$ se siguen de la aditividad numerable de P y el agotamiento de la recta real.

4.2 Variables Aleatorias Discretas

Una variable aleatoria X es **discreta** si su rango es un conjunto finito o numerable $S \subset \mathbb{R}$. La **función de masa de probabilidad (pmf)** es

$$p_X(x) = P(X = x), \quad x \in S$$

con $\sum_{x \in S} p_X(x) = 1$.

Ejemplo: Suma de Dos Dados Justos: Sea X la suma de los resultados de lanzar dos dados justos independientes. Los valores posibles son $x \in 2, 3, \dots, 12$. Cada resultado (i, j) con $i, j \in 1, \dots, 6$ es igualmente probable. El número de pares (i, j) que dan suma x es $n(x) = 6 - |7 - x|$, por lo que

$$p_X(x) = \frac{n(x)}{36}, \quad n(x) = 6 - |7 - x|, \quad x \in 2, 3, \dots, 12.$$

Derivación: Para $x = 7$, $n(7) = 6$; para $x = 2$, $n(2) = 1$; etc.

4.3 Variables Aleatorias Continuas

Una variable aleatoria X es **continua** si existe una **función de densidad de probabilidad (pdf)** $f_X : \mathbb{R} \rightarrow [0, \infty)$ tal que, para todo $a < b$,

$$P(a < X \leq b) = \int_a^b f_X(x) dx.$$

La pdf satisface

1. $f_X(x) \geq 0$ para todo $x \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} f_X(x) dx = 1$

Observación: La existencia de una pdf requiere que F_X sea absolutamente continua [Durrett2019, Sec. 1.3].

Ejemplo: Distribución Normal Estándar: Sea X con distribución normal estándar:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}$$

Entonces, para cualquier $a < b$,

$$P(a < X \leq b) = \int_a^b f_X(x) dx.$$

4.4 Transformaciones de Variables Aleatorias

Dada una variable aleatoria X con pdf f_X y una función $h : \mathbb{R} \rightarrow \mathbb{R}$ medible, invertible y diferenciable, definamos $Y = h(X)$. La pdf de Y es

$$f_Y(y) = f_X(h^{-1}(y)) \left| \frac{d}{dy} h^{-1}(y) \right|, \quad y \in h(\mathbb{R})$$

Supuestos: h es estrictamente monótona y diferenciable con inversa h^{-1} , y f_X es conocida.

**** Ejemplo: Normal Estándar al Cuadrado:**** Sea $X \sim N(0, 1)$, $Y = X^2$. Entonces para $y \geq 0$,

- $h^{-1}(y) = \pm\sqrt{y}$,
- $f_X(h^{-1}(y)) = \frac{1}{\sqrt{2\pi}} e^{-y/2}$ para ambas raíces,

- $\left| \frac{d}{dy} h^{-1}(y) \right| = \frac{1}{2\sqrt{y}}.$

Por tanto,

$$f_Y(y) = f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi y}} e^{-y/2}, \quad y > 0,$$

que es la distribución chi-cuadrada con un grado de libertad [Billingsley1995, Sec. 10].

4.5 Vectores Aleatorios

Un **vector aleatorio** $\mathbf{X} = (X_1, \dots, X_n)$ es una función medible $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$. La **función de distribución conjunta** es

$$F_{\mathbf{X}}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$$

Si existe una pdf conjunta $f_{\mathbf{X}}$, entonces

$$P((X_1, \dots, X_n) \in B) = \int_B f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \cdots dx_n$$

para todos los conjuntos de Borel $B \subset \mathbb{R}^n$.

Marginales y Condicionales: La pdf marginal de X_1 es

$$f_{X_1}(x_1) = \int_{\mathbb{R}^{n-1}} f_{\mathbf{X}}(x_1, x_2, \dots, x_n) dx_2 \cdots dx_n$$

Las distribuciones condicionales se definen de manera análoga [Durrett2019, Sec. 1.5].

Ejemplo: Normal Bivariada. Sea (X_1, X_2) conjuntamente normal con media cero, varianzas 1, y covarianza ρ . La pdf conjunta es

$$f_{\mathbf{X}}(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\right)$$

4.6 Independencia de Variables Aleatorias

Las variables aleatorias X e Y son **independientes** si, para todo $x, y \in \mathbb{R}$,

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$$

Equivalentemente, $F_{(X,Y)}(x, y) = F_X(x)F_Y(y)$ para todo x, y .

Criterios Equivalentes: La independencia se cumple si y solo si $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ para todos los conjuntos de Borel $A, B \subseteq \mathbb{R}$ [Billingsley1995, Thm. 12.1].

Contraejemplo: Sea X uniforme en $[-1, 1]$, $Y = X$. $P(X \leq x, Y \leq y) = P(X \leq \min x, y)$ no es igual a $P(X \leq x)P(Y \leq y)$ a menos que x o y sea $-\infty$ o ∞ .

4.7 Esperanza y Varianza

Sea X una variable aleatoria.

- **Esperanza:**

- Discreta: $E[X] = \sum_x xp_X(x)$, siempre que $\sum_x |x|p_X(x) < \infty$.
- Continua: $E[X] = \int_{-\infty}^{\infty} xf_X(x) dx$ si $\int |x|f_X(x) dx < \infty$.

- **Varianza:**

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

Propiedades: Para cualquier $a, b \in \mathbb{R}$,

- $E[aX + b] = aE[X] + b$
- $\text{Var}(aX + b) = a^2 \text{Var}(X)$

Demostraciones: Ver [Durrett2019, Prop. 1.5.1].

4.8 Simulación de Variables Aleatorias

El **método de transformación inversa** es una técnica fundamental de simulación: Si $U \sim \text{Uniforme}(0, 1)$ y F_X es continua y estrictamente creciente, entonces $X = F_X^{-1}(U)$ tiene función de distribución F_X .

Simulación avanzada:

- **Método de aceptación-rechazo:** Generar Y con densidad g , aceptar con probabilidad $f_X(Y)/[cg(Y)]$, donde c satisface $f_X(y) \leq cg(y)$ para todo y [Devroye1986].
- **Método de Box–Muller:** Para X normal estándar, generar U_1, U_2 independientes, uniformes en $(0, 1)$. Entonces $X = \sqrt{-2 \log U_1} \cos(2\pi U_2)$.

4.9 Ejercicios

1. Demostrar que para cualquier CDF F , la función es continua por la derecha y satisface los límites establecidos en $\pm\infty$.
2. Sea X uniforme en $[0, 1]$, y $Y = -\ln(1 - X)$. Mostrar que Y es exponencial con parámetro 1.
3. Dadas las variables aleatorias X e Y con pdf conjunta $f(x, y) = 6xy$ para $0 < x < 1$, $0 < y < 1$, $x + y < 1$, calcular $P(X < Y)$ y verificar si X e Y son independientes.

5 Variables Aleatorias y Vectores: Transformaciones y Aplicaciones

5.1 Distribuciones de Mezcla

Una **distribución de mezcla** describe la ley de probabilidad de una variable aleatoria X que depende de cuál evento ocurre entre una partición A_1, A_2, \dots, A_n del espacio de probabilidad subyacente (Ω, \mathcal{A}, P) [Billingsley1995]. Para cada i , sea $F_{X|A_i}(x)$ la función de distribución acumulativa (CDF) condicional de X dado A_i , y $P(A_i)$ la probabilidad de A_i . La ley de probabilidad total produce:

$$F_X(x) = \sum_{i=1}^n F_{X|A_i}(x)P(A_i).$$

Demostración. Sea $F_X(x) = P(X \leq x)$. Puesto que los A_i son mutuamente exclusivos y exhaustivos,

$$P(X \leq x) = \sum_{i=1}^n P(X \leq x, A_i) = \sum_{i=1}^n P(X \leq x|A_i)P(A_i).$$

Por definición, $P(X \leq x|A_i) = F_{X|A_i}(x)$. ■

Ejemplo Avanzado. Supongamos que X es una **mezcla gaussiana finita**:

- Con probabilidad p_1 , $X \sim N(\mu_1, \sigma_1^2)$
- Con probabilidad $p_2 = 1 - p_1$, $X \sim N(\mu_2, \sigma_2^2)$

Entonces

$$f_X(x) = p_1 \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} + p_2 \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}.$$

5.2 Transformaciones de Variables Aleatorias

Sea X una variable aleatoria de valor real con densidad $f_X(x)$, y $h : \mathbb{R} \rightarrow \mathbb{R}$ una **biyección diferenciable** (es decir, h es invertible y h^{-1} es diferenciable). Definamos $Y = h(X)$. La distribución de Y se determina como sigue.

Definición (Fórmula de Cambio de Variable). Si h es estrictamente monótona y diferenciable, entonces para y en el rango de h ,

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

Demostración. Sea h estrictamente creciente. La CDF de Y es $F_Y(y) = P(Y \leq y) = P(X \leq h^{-1}(y)) = F_X(h^{-1}(y))$. Diferenciando ambos lados da

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

Si h es decreciente, $\frac{d}{dy} h^{-1}(y)$ es negativo; por tanto, usamos el valor absoluto [CasellaBerger2002].

Soporte para Casos No Monótonos. Si h no es monótona, $h^{-1}(y)$ es multivaluada. Entonces,

$$f_Y(y) = \sum_{x \in h^{-1}(y)} \frac{f_X(x)}{|h'(x)|}.$$

Ejemplo. Sea $Z \sim N(0, 1)$. Encontrar la distribución de $X = Z^2$.

- $h(z) = z^2$, por lo que X tiene soporte en $x > 0$.
- $h^{-1}(x) = \pm\sqrt{x}$.
- $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$.

Por tanto,

$$f_X(x) = \frac{f_Z(\sqrt{x})}{|2\sqrt{x}|} + \frac{f_Z(-\sqrt{x})}{|2\sqrt{x}|} = \frac{1}{\sqrt{2\pi}} \frac{e^{-x/2}}{2\sqrt{x}}, \quad x > 0.$$

X sigue una distribución chi-cuadrada con 1 grado de libertad.

5.2.1 Teorema de Transformación

Teorema. Sea X con CDF F_X e $Y = h(X)$, donde h es estrictamente creciente y continuamente diferenciable. Entonces,

$$F_Y(y) = F_X(h^{-1}(y)), \quad f_Y(y) = f_X(h^{-1}(y)) \frac{1}{h'(h^{-1}(y))}, \quad \text{para } y \in h(\mathbb{R}).$$

Demostración. Inmediato por la regla de la cadena y argumentos de cambio de variable anteriores [Durrett2019].

Nota Terminológica. Usar “transformación diferenciable” (no “transformación continua”) según la terminología estándar.

5.3 Simulación vía Transformación Inversa

Supongamos que F es la CDF de X , estrictamente creciente y continua. Si $U \sim \text{Uniforme}(0, 1)$, entonces

$$X = F^{-1}(U)$$

tiene CDF F [Devroye1986].

Soporte y Advertencias:

- F debe ser invertible; para F discreta o no invertible, este método requiere adaptación.
- En la práctica, F^{-1} puede calcularse numéricamente; se requiere cuidado para la precisión.

5.4 Vectores Aleatorios

Un **vector aleatorio** $\bar{X} = (X_1, \dots, X_k)$ es una función medible de (Ω, \mathcal{A}, P) a $(\mathbb{R}^k, \mathcal{B}^k)$, donde \mathcal{B}^k es la σ -álgebra de Borel en \mathbb{R}^k . Su CDF conjunta es:

$$F_{\bar{X}}(\bar{x}) = P(X_1 \leq x_1, \dots, X_k \leq x_k).$$

Caso Discreto y Continuo. Si (X, Y) es discreto:

$$p_{X,Y}(x, y) = P(X = x, Y = y).$$

Marginales: $p_X(x) = \sum_y p_{X,Y}(x, y)$, $p_Y(y) = \sum_x p_{X,Y}(x, y)$.

Si continuo:

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y).$$

Coherencia. La función indicadora $\mathbf{1}$, denota 1 si su argumento es verdadero, 0 en caso contrario. La función $\Phi(x)$ es la CDF normal estándar.

5.5 Independencia

Definición. Las variables aleatorias X e Y son independientes si para todo x, y ,

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y).$$

Equivalentemente, su densidad conjunta (si existe) se factoriza:

$$f_{X,Y}(x,y) = f_X(x)f_Y(y).$$

5.6 Transformaciones de Vectores Aleatorios

5.6.1 Transformaciones Lineales

Sea $\bar{X} = (X_1, \dots, X_n)^T$ con vector de medias μ y matriz de covarianza Σ . Para $A \in \mathbb{R}^{m \times n}$, definamos $\bar{Y} = A\bar{X}$.

Teorema.

- $\mathbb{E}[\bar{Y}] = A\mathbb{E}[\bar{X}]$
- $\text{Cov}(\bar{Y}) = A\text{Cov}(\bar{X})A^T$

Demostración. Por la linealidad de la esperanza, $\mathbb{E}[\bar{Y}] = \mathbb{E}[A\bar{X}] = A\mathbb{E}[\bar{X}]$.

Para la covarianza,

$$\text{Cov}(\bar{Y}) = \mathbb{E}[(\bar{Y} - \mathbb{E}[\bar{Y}])(\bar{Y} - \mathbb{E}[\bar{Y}])^T] = A\mathbb{E}[(\bar{X} - \mu)(\bar{X} - \mu)^T]A^T = A\Sigma A^T.$$

Ejemplo. Sea $\bar{X} \sim N_2(\mu, \Sigma)$ y

$$\bar{Y} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \bar{X}.$$

Si $\mu = (\mu_1, \mu_2)^T$, y $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, entonces

$$\mathbb{E}[\bar{Y}] = \begin{pmatrix} \mu_1 + \mu_2 \\ \mu_1 - \mu_2 \end{pmatrix},$$

$$\text{Cov}(\bar{Y}) = \begin{pmatrix} \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2 & \sigma_1^2 - \sigma_2^2 \\ \sigma_1^2 - \sigma_2^2 & \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2 \end{pmatrix}.$$

5.6.2 Transformaciones No Lineales

Supongamos que $\bar{Y} = h(\bar{X})$ es una biyección diferenciable de \mathbb{R}^n a \mathbb{R}^n . Sea $J = \det\left(\frac{\partial h}{\partial \bar{x}}\right)$ el determinante jacobiano.

Teorema (Cambio de Variables Multivariado). Sea $f_{\bar{X}}$ la densidad de \bar{X} , y supongamos que h y h^{-1} son diferenciables, con J no cero en todas partes. Entonces la densidad de \bar{Y} es

$$f_{\bar{Y}}(\bar{y}) = f_{\bar{X}}(h^{-1}(\bar{y})) |\det Dh^{-1}(\bar{y})|,$$

donde Dh^{-1} es el jacobiano de h^{-1} [Billingsley1995].

Demostración. Ver [Billingsley1995, §16] para justificación rigurosa. La fórmula surge de la regla de sustitución en integrales multidimensionales.

Ejemplo (Avanzado). Sean X_1, X_2 independientes, cada una $\sim \text{Exp}(\lambda)$. Definamos:

$$Y_1 = X_1 + X_2, \quad Y_2 = \frac{X_1}{X_1 + X_2}.$$

Calcular la densidad conjunta $f_{Y_1, Y_2}(y_1, y_2)$.

- La transformación inversa:
 - $X_1 = y_1 y_2$
 - $X_2 = y_1(1 - y_2)$
- Determinante jacobiano:

$$J = \begin{vmatrix} \frac{\partial X_1}{\partial y_1} & \frac{\partial X_1}{\partial y_2} \\ \frac{\partial X_2}{\partial y_1} & \frac{\partial X_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} y_2 & y_1 \\ 1 - y_2 & -y_1 \end{vmatrix} = -y_1$$

- La densidad conjunta:

$$f_{Y_1, Y_2}(y_1, y_2) = f_{X_1, X_2}(y_1 y_2, y_1(1 - y_2)) \cdot |J| = \lambda^2 y_1 e^{-\lambda y_1} \mathbf{1}_{y_1 > 0, 0 < y_2 < 1}.$$

Ejemplo Avanzado. Considerar el mapeo de coordenadas polares a cartesianas: $X = R \cos \Theta$, $Y = R \sin \Theta$, con R, Θ independientes, $R > 0$, $\Theta \in (0, 2\pi)$. El jacobiano es r , dando la densidad conjunta $f_{X, Y}(x, y) = f_{R, \Theta}(\sqrt{x^2 + y^2}, \arctan(y/x)) \cdot \frac{1}{\sqrt{x^2 + y^2}}$.

5.7 Aplicaciones en Análisis Multivariado

5.7.1 Análisis de Componentes Principales (PCA)

PCA transforma variables aleatorias correlacionadas en componentes no correlacionados mediante una transformación lineal ortogonal [Jolliffe2016]. Sea \bar{X} con matriz de covarianza Σ . El primer componente principal es la combinación lineal que maximiza la varianza, es decir, maximizar $\text{Var}(a^T \bar{X})$ bajo $\|a\| = 1$. Los vectores propios de Σ proporcionan la transformación; los valores propios correspondientes cuantifican la varianza explicada.

Ejemplo Numérico. Sea $\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. Valores propios: 3, 1. Primer componente principal: $a_1 = (1/\sqrt{2}, 1/\sqrt{2})$, varianza 3.

5.7.2 Análisis de Correlación Canónica

Dados dos vectores aleatorios (\bar{X}, \bar{Y}) , CCA busca combinaciones lineales que maximizan la correlación [MardiaKentBibby1979]. Sean a, b que maximizan $\text{corr}(a^T \bar{X}, b^T \bar{Y})$. Esto requiere tanto matrices de covarianza como de covarianza cruzada.

5.7.3 Análisis Factorial

El análisis factorial modela la matriz de covarianza Σ como $\Sigma = \Lambda\Lambda^T + \Psi$, donde Λ captura factores comunes y Ψ es diagonal (varianza específica). Asume normalidad e independencia de factores [MardiaKentBibby1979].

5.8 Ejercicios

1. **Distribución de Mezcla.** Sea X con distribución de mezcla: con probabilidad p , $X \sim N(0, 1)$; con $1 - p$, $X \sim N(\mu, 1)$.
 - (a) Escribir la FDP y CDF de X .
 - (b) Mostrar todos los pasos.
2. **Demostración de Transformación.** Sea X con FDP $f_X(x)$, h estrictamente creciente y diferenciable. Demostrar la fórmula para $f_Y(y)$.
3. **Cambio de Variables Multivariado.** Sean X_1, X_2 independientes, $\sim \text{Exp}(1)$. Encontrar la densidad conjunta de $(Y_1, Y_2) = (X_1 + X_2, X_1/X_2)$.
4. **Cálculo de PCA.** Dada $\Sigma = \begin{pmatrix} 4 & 2 \\ 2 & 3 \end{pmatrix}$, encontrar los componentes principales y la proporción de varianza total explicada.

6 Distribuciones Condicionales y Predicción

6.1 Distribuciones Condicionales

Sea (Ω, \mathcal{F}, P) un espacio de probabilidad. Consideremos un vector aleatorio (X, Y) definido en este espacio, donde $X : \Omega \rightarrow \mathbb{R}$ e $Y : \Omega \rightarrow \mathbb{R}$ son variables aleatorias. La **distribución condicional** de Y dado X es una familia de medidas de probabilidad $P_{Y|X=x} : x \in \mathbb{R}$ tal que, para cualquier conjunto de Borel $B \subseteq \mathbb{R}$,

$$P_{Y|X=x}(B) = P(Y \in B \mid X = x), \quad \text{siempre que } P(X = x) > 0.$$

Más generalmente, para variables aleatorias arbitrarias y σ -álgebras, la distribución condicional se define vía la derivada de Radon–Nikodym como la versión regular única (salvo conjuntos P -nulos) $P_{Y|X}(\cdot \mid X)$ [Billingsley1995].

Para variables aleatorias **discretas**,

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)}, \quad \text{siempre que } P(X = x) > 0.$$

Para variables aleatorias **absolutamente continuas**, con densidad conjunta $f_{X,Y}$ y marginal f_X ,

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}, \quad \text{donde } f_X(x) > 0.$$

6.1.1 Propiedades

- **Propiedad de Torre:** $\mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}[Y]$.
- **Desigualdad de Jensen:** Para cualquier función convexa φ , $\varphi(\mathbb{E}[Y|X]) \leq \mathbb{E}[\varphi(Y)|X]$ [Durrett2019].

Ejemplo.

Sea $(X, Y) \sim N_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$ un vector aleatorio normal bivariado, donde $|\rho| < 1$. La distribución condicional de Y dado $X = x$ es normal:

$$Y|X = x \sim N\left(\mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1 - \rho^2)\right).$$

Demostración: Sea $\Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$. Las propiedades estándar de la normal multivariada producen la media y varianza condicional como arriba; ver [Billingsley1995, p. 186].

6.2 Teoría de Predicción

Sea (X, Y) como arriba. Supongamos que se busca una función $g(X)$ para predecir Y tal que el **error cuadrático medio** $\mathbb{E}[(Y - g(X))^2]$ se minimice. El predictor óptimo en este sentido es la **esperanza condicional**:

$$\hat{Y} = \mathbb{E}[Y | X].$$

6.2.1 Propiedades

Sea $\mathcal{G} = \sigma(X)$. Entonces:

1. **Insesgadez:** $\mathbb{E}[\hat{Y}] = \mathbb{E}[Y]$.
2. **Ortogonalidad:** $\mathbb{E}[(Y - \hat{Y})h(X)] = 0$ para cualquier función integrable h medible con respecto a \mathcal{G} .
3. **Varianza Mínima:** Para cualquier otro predictor $g(X)$,

$$\mathbb{E}[(Y - \hat{Y})^2] \leq \mathbb{E}[(Y - g(X))^2].$$

Demostraciones:

- *Insesgadez:* Por la propiedad de torre, $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$.
- *Ortogonalidad:* Para cualquier h medible, $\mathbb{E}[(Y - \mathbb{E}[Y|X])h(X)] = 0$ [Durrett2019, Theorem 5.5.3].
- *Varianza Mínima:* Para cualquier $g(X)$,

$$\mathbb{E}[(Y - g(X))^2] = \mathbb{E}[(Y - \mathbb{E}[Y|X])^2] + \mathbb{E}[(\mathbb{E}[Y|X] - g(X))^2],$$

por lo que el mínimo se alcanza en $g = \mathbb{E}[Y|X]$.

6.3 Predicción Lineal

Supongamos que restringimos la atención a **predictores lineales** de la forma $g(X) = a + bX$. Los coeficientes a, b que minimizan el error cuadrático medio son:

$$b^* = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}, \quad a^* = \mathbb{E}[Y] - b^* \mathbb{E}[X],$$

siempre que $\text{Var}(X) > 0$. Por tanto, el **mejor predictor lineal** es

$$\hat{Y}_L = \mathbb{E}[Y] + \frac{\text{Cov}(X, Y)}{\text{Var}(X)}(X - \mathbb{E}[X]).$$

Demostración: Sea $g(X) = a + bX$; expandir $\mathbb{E}[(Y - (a + bX))^2]$, establecer las derivadas con respecto a a y b igual a cero, y resolver para los a y b que minimizan.

6.3.1 Caso Especial: Normal Bivariado

Si $(X, Y) \sim N_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, entonces $\mathbb{E}[Y|X]$ es afín en X , por lo que el mejor predictor lineal iguala la esperanza condicional:

$$\mathbb{E}[Y|X] = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(X - \mu_X).$$

Por tanto, el mejor predictor lineal y el mejor predictor general (sin restricciones) coinciden para la normal bivariada [LehmannCasella1998].

6.4 Aplicaciones

6.4.1 Análisis de Regresión

La regresión estudia la relación entre una variable dependiente Y y variables independientes X , con la esperanza condicional $\mathbb{E}[Y|X]$ como objetivo de estimación. **Ejemplo:** Supongamos $Y = \beta_0 + \beta_1 X + \varepsilon$ con $\varepsilon \sim N(0, \sigma^2)$, independiente de X . Dadas las observaciones $(X_1, Y_1), \dots, (X_n, Y_n)$, el estimador de mínimos cuadrados ordinarios $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ proporciona predicciones para Y dado X . Los intervalos de predicción se siguen de la distribución normal condicional [Hamilton1994].

6.4.2 Pronóstico de Series de Tiempo

Para una serie de tiempo estacionaria X_t , la predicción de X_{t+h} dado $\mathcal{F}_t = \sigma(X_1, \dots, X_t)$ es $\mathbb{E}[X_{t+h}|\mathcal{F}_t]$. **Ejemplo:** En el modelo AR(1), $X_{t+1} = \phi X_t + \varepsilon_{t+1}$ con $\varepsilon_{t+1} \sim N(0, \sigma^2)$,

$$\mathbb{E}[X_{t+1}|X_t] = \phi X_t.$$

La varianza del error de pronóstico es $\text{Var}(X_{t+1} - \phi X_t) = \sigma^2$. Los intervalos de predicción se derivan de la distribución normal condicional [Hamilton1994].

6.4.3 Aprendizaje Automático

En aprendizaje supervisado, los algoritmos buscan aproximar $\mathbb{E}[Y|X]$ usando datos. **Ejemplo:** Una red neuronal de alimentación hacia adelante $f_\theta(X)$ se entrena minimizando el error cuadrático medio empírico sobre muestras (X_i, Y_i) . Después del entrenamiento, $f_\theta(X)$ aproxima la esperanza condicional $\mathbb{E}[Y|X]$. Los métodos de ensamble, como los bosques aleatorios, combinan múltiples predictores para mejorar la precisión de estimación [HastieTibshiraniFriedman2009].

6.5 Ejercicios

1. **Esperanza Condicional (Demostración):** Demostrar que $\mathbb{E}[(Y - \mathbb{E}[Y|X])h(X)] = 0$ para cualquier función integrable $h(X)$.
2. **Predictor Lineal (Cálculo):** Sean X e Y con distribución conjunta: $\mathbb{E}[X] = 2$, $\mathbb{E}[Y] = 5$, $\text{Cov}(X, Y) = 3$, $\text{Var}(X) = 4$. Calcular el mejor predictor lineal de Y dado X .
3. **Series de Tiempo (Aplicación):** Para el modelo AR(1) $X_{t+1} = 0.7X_t + \varepsilon_{t+1}$, $\varepsilon_{t+1} \sim N(0, 1)$, calcular $\mathbb{E}[X_{t+2}|X_t]$.
4. **Aprendizaje Automático (Interpretación):** Supongamos que una red neuronal produce $\hat{Y} = f_\theta(X)$. Explicar bajo qué condiciones $f_\theta(X)$ aproxima $\mathbb{E}[Y|X]$ y cómo esto se relaciona con la optimalidad de predicción.

7 Ley de los Grandes Números y Teorema Central del Límite

7.1 Ley de los Grandes Números

La Ley de los Grandes Números (LGN) es un resultado fundamental en la teoría de probabilidades, que captura la estabilización de las medias muestrales alrededor del valor esperado a medida que aumenta el número de ensayos. Para proceder rigurosamente, clarificamos la terminología subyacente y establecemos definiciones precisas de los modos de convergencia.

Definición (Variables Aleatorias Independientes e Idénticamente Distribuidas) Una sucesión $X_{i \in \mathbb{N}}$ es independiente e idénticamente distribuida (iid) si:

1. Cada X_i es una variable aleatoria definida en el mismo espacio de probabilidad (Ω, \mathcal{F}, P) .
2. $P(X_i \leq x) = F(x)$ para todo i y todo $x \in \mathbb{R}$ (distribución idéntica).
3. Para cualquier subconjunto finito i_1, \dots, i_k , las variables X_{i_1}, \dots, X_{i_k} son independientes.

Definición (Convergencia en Probabilidad) Una sucesión de variables aleatorias Y_n converge en probabilidad a Y si para todo $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|Y_n - Y| > \varepsilon) = 0.$$

Definición (Convergencia Casi Segura) Y_n converge casi seguramente (c.s.) a Y si

$$P\left(\lim_{n \rightarrow \infty} Y_n = Y\right) = 1.$$

Sea $X_{i \in \mathbb{N}}$ variables aleatorias iid con media $\mu = E[X_1]$ y varianza finita $\sigma^2 = \text{Var}(X_1) < \infty$. La media muestral es

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

Teorema (Ley Débil de los Grandes Números) \bar{X}_n converge en probabilidad a μ :

$$\bar{X}_n \xrightarrow{p} \mu.$$

Demostración. Por la desigualdad de Chebyshev:

- $E[\bar{X}_n] = \mu$
- $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$ Así, para $\varepsilon > 0$,

$$P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0 \text{ cuando } n \rightarrow \infty.$$

Por tanto, \bar{X}_n converge en probabilidad a μ . ■

Teorema (Ley Fuerte de los Grandes Números) \bar{X}_n converge casi seguramente a μ :

$$P\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right) = 1.$$

Esbozo de Demostración. La demostración, debida a Kolmogorov, aprovecha el lema de Borel-Cantelli y las propiedades de secuencias iid con varianza finita. La convergencia casi segura es estrictamente más fuerte que la convergencia en probabilidad; véase [Billingsley1995, §22] para una exposición detallada. ■

Ejemplo Sea X_i el resultado del i -ésimo lanzamiento de una moneda equilibrada ($X_i = 1$ para cara, 0 para cruz). Entonces $E[X_i] = 0.5$, $\text{Var}(X_i) = 0.25$. Por la Ley Fuerte de los Grandes Números,

$$P\left(\lim_{n \rightarrow \infty} \bar{X}_n = 0.5\right) = 1,$$

donde \bar{X}_n es la fracción de caras en n lanzamientos.

7.2 Teorema Central del Límite

El Teorema Central del Límite (TCL) describe la emergencia universal de la distribución normal en sumas de variables aleatorias iid, independientemente de la distribución original (bajo condiciones suaves).

Definición (Estandarización) Dadas variables aleatorias iid $X_{i \in \mathbb{N}}$ con media μ y varianza $\sigma^2 > 0$, definimos

$$Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}.$$

Teorema (Teorema Central del Límite) Cuando $n \rightarrow \infty$, Z_n converge en distribución a una variable aleatoria normal estándar:

$$Z_n \xrightarrow{d} N(0, 1),$$

es decir,

$$\lim_{n \rightarrow \infty} P(Z_n \leq z) = \Phi(z),$$

donde $\Phi(z)$ es la función de distribución acumulada de $N(0, 1)$.

Demostración (Enfoque de Función Característica). Sea $\varphi_X(t) = E[e^{itX}]$ la función característica de X_i . La función característica de Z_n es

$$\varphi_{Z_n}(t) = \left[\varphi_X \left(\frac{t}{\sigma\sqrt{n}} \right) \right]^n e^{-it \frac{n\mu}{\sigma\sqrt{n}}}.$$

Una expansión de Taylor y la independencia producen

$$\lim_{n \rightarrow \infty} \varphi_{Z_n}(t) = e^{-t^2/2},$$

que es la función característica de $N(0, 1)$. Así, por el teorema de continuidad de Lévy, $Z_n \xrightarrow{d} N(0, 1)$. Para una demostración completa, véase [Durrett2019, §2.5]. ■

Ejemplo Supongamos que X_i son iid con $E[X_i] = 2$, $\text{Var}(X_i) = 9$. Para $n = 100$,

$$Z_{100} = \frac{\sum_{i=1}^{100} X_i - 200}{30}.$$

Según el TCL, Z_{100} está aproximadamente distribuida como $N(0, 1)$, incluso si las X_i no son normales. Una simulación con X_i no normales (por ejemplo, Bernoulli o Poisson) mostrará empíricamente la convergencia a la curva de campana a medida que n aumenta.

7.3 Aproximaciones del Teorema Central del Límite

El TCL proporciona justificación rigurosa para aproximaciones normales a distribuciones discretas bajo regímenes de parámetros apropiados. Los siguientes resultados se usan ampliamente en inferencia estadística y probabilidad combinatoria.

Aproximación Binomial-a-Normal Sea $X \sim \text{Bin}(n, p)$. Si n es grande, $np > 5$, y $n(1-p) > 5$:

$$X \approx N(np, np(1-p)).$$

Justificación: Estos criterios aseguran que la asimetría es suficientemente pequeña y que la distribución discreta está bien aproximada por la normal continua [Feller1968, §7.1].

Aproximación Poisson-a-Normal Si $X \sim \text{Poisson}(\lambda)$, entonces para λ grande (por ejemplo, $\lambda > 10$),

$$X \approx N(\lambda, \lambda).$$

Justificación: A medida que λ aumenta, la distribución de Poisson se vuelve cada vez más simétrica y similar a la normal [Durrett2019, §2.7].

Ejercicio

1. *Aproximación Normal:* Sea $X \sim \text{Bin}(200, 0.1)$.
 - (a) Use el TCL para aproximar $P(15 \leq X \leq 25)$.
 - (b) Compare el resultado con la probabilidad binomial exacta.

8 El Proceso de Bernoulli: Definición, Propiedades y Aplicaciones

8.1 Definición Formal de Procesos Estocásticos

Sea (Ω, \mathcal{F}, P) un espacio de probabilidad, donde Ω es el espacio muestral, \mathcal{F} es una σ -álgebra de eventos, y P es una medida de probabilidad. Un **proceso estocástico** es una colección $X_t : t \in T$ de variables aleatorias $X_t : \Omega \rightarrow S$, indexadas por un conjunto T (frecuentemente \mathbb{N} o \mathbb{R}_+), donde S es un espacio medible. Para cada $t \in T$ fijo, X_t es una función medible, y para cada $\omega \in \Omega$ fijo, el mapeo $t \mapsto X_t(\omega)$ se llama una *trayectoria muestral* o *trayectoria* del proceso [Billingsley1995; Durrett2019].

8.2 El Proceso de Bernoulli

Un **proceso de Bernoulli** es un proceso estocástico prototípico de tiempo discreto que modela secuencias de ensayos binarios independientes (éxito/fracaso).

8.2.1 Construcción Rigurosa

Sea (Ω, \mathcal{F}, P) un espacio de probabilidad. Un proceso de Bernoulli con parámetro $p \in (0, 1)$ es una secuencia infinita de variables aleatorias independientes e idénticamente distribuidas $X_{n \in \mathbb{N}}$, donde para cada n ,

$$P(X_n = 1) = p, \quad P(X_n = 0) = 1 - p,$$

y todas las X_n son independientes.

- **Espacio muestral:** $\Omega = 0, 1^{\mathbb{N}}$, cada $\omega \in \Omega$ es una secuencia binaria infinita.
- **Sigma-álgebra:** \mathcal{F} es la σ -álgebra producto en Ω .
- **Medida de probabilidad:** P es la medida producto que asigna $P(X_n = 1) = p$, $P(X_n = 0) = 1 - p$ independientemente para todo n .

8.2.2 Distribución Conjunta

Para cualquier $n \in \mathbb{N}$ y cualquier $(x_1, \dots, x_n) \in 0, 1^n$,

$$P(X_1 = x_1, \dots, X_n = x_n) = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}.$$

Demostración: Por independencia y distribución idéntica, $P(X_i = x_i) = p$ si $x_i = 1$, $1-p$ si $x_i = 0$; el producto sobre i produce el resultado [Durrett2019].

8.3 Variables Aleatorias Asociadas

8.3.1 Número Total de Éxitos en n Ensayos

Definamos

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

Entonces S_n es el número total de éxitos en los primeros n ensayos.

Teorema: $S_n \sim \text{Binomial}(n, p)$.

Demostración: S_n es la suma de n variables aleatorias Bernoulli(p) independientes. Para $k = 0, 1, \dots, n$,

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Esto se sigue de la enumeración combinatoria de secuencias con k éxitos y $n-k$ fracasos [Billingsley1995].

8.3.2 Tiempo de Espera para el Primer Éxito

Sea

$$T = \min\{n \geq 1 : X_n = 1\}$$

el tiempo de espera hasta el primer éxito.

Proposición: T es una variable aleatoria geométrica con parámetro p :

$$P(T = n) = (1-p)^{n-1} p, \quad n = 1, 2, \dots$$

Demostración: Los primeros $n - 1$ ensayos son fracasos ($X_1 = 0, \dots, X_{n-1} = 0$), el n -ésimo ensayo es un éxito ($X_n = 1$). Por independencia,

$$P(T = n) = (1 - p)^{n-1}p.$$

[@Durrett2019]

Propiedad de Falta de Memoria: Para $m, n \in \mathbb{N}$,

$$P(T > m + n \mid T > n) = P(T > m).$$

Demostración: Por la definición de independencia,

$$P(T > n) = (1 - p)^n.$$

Por tanto,

$$P(T > m + n \mid T > n) = \frac{P(T > m + n)}{P(T > n)} = \frac{(1 - p)^{m+n}}{(1 - p)^n} = (1 - p)^m = P(T > m).$$

8.3.3 Tiempo de Espera para el k -ésimo Éxito

Sea T_k el ensayo en el que ocurre el k -ésimo éxito.

Teorema: T_k sigue la distribución binomial negativa:

$$P(T_k = t) = \binom{t-1}{k-1} p^k (1-p)^{t-k}, \quad t = k, k+1, \dots$$

Demostración: Esta es la probabilidad de que en $t - 1$ ensayos, haya exactamente $k - 1$ éxitos (en cualquier orden), y el t -ésimo ensayo sea un éxito. Ver [@Billingsley1995].

Además, T_k es la suma de k variables aleatorias geométrica(p) independientes [@Durrett2019].

8.4 Propiedades Clave

- **Independencia:** La secuencia X_n es independiente por construcción.
- **Estacionaridad del Proceso:** Los incrementos $S_{n+m} - S_n$ son binomiales con parámetros (m, p) y son independientes de S_n , reflejando la falta de memoria en el proceso.
- **Propiedad de Falta de Memoria:** El tiempo de espera geométrico T para el primer éxito satisface $P(T > m + n \mid T > n) = P(T > m)$.
- **Terminología Estándar:** La propiedad se llama universalmente la *propiedad de falta de memoria*, no “memoria corta” [@Durrett2019].

8.5 Conexiones con Otras Distribuciones

8.5.1 Aproximación de Poisson a la Binomial

Cuando n es grande y p es pequeño con $\lambda = np$ fijo, la distribución binomial se aproxima a la distribución de Poisson:

$$\lim_{n \rightarrow \infty, p \rightarrow 0, np = \lambda} \binom{n}{k} p^k (1-p)^{n-k} = \frac{e^{-\lambda} \lambda^k}{k!}.$$

Esta es la *ley de eventos raros* [Billingsley1995, §19].

Nota: Las cotas de error y tasas de convergencia se pueden encontrar en [LeCam1960].

8.5.2 Generalización Multinomial

Si cada ensayo admite $r > 2$ resultados posibles con probabilidades constantes p_1, \dots, p_r , el vector (N_1, \dots, N_r) , donde N_i cuenta las ocurrencias del resultado i en n ensayos, sigue la distribución multinomial:

$$P(N_1 = n_1, \dots, N_r = n_r) = \frac{n!}{n_1! \dots n_r!} p_1^{n_1} \dots p_r^{n_r}, \quad \sum_{i=1}^r n_i = n.$$

Esta es una generalización directa del modelo binomial [Durrett2019].

Ejemplo: Supongamos que una línea de producción produce artículos con probabilidad de defecto independiente $p = 0.02$. Si se verifican $n = 100$ artículos:

- La probabilidad de exactamente $k = 3$ artículos defectuosos es

$$P(S_{100} = 3) = \binom{100}{3} (0.02)^3 (0.98)^{97} \approx 0.180.$$

- La probabilidad de que el primer defecto aparezca en el artículo 15:

$$P(T = 15) = (0.98)^{14} (0.02) \approx 0.0148.$$

8.6 Ejercicios

1. (**Construcción Rigurosa**) Construir explícitamente un espacio de probabilidad y definir las variables aleatorias coordinadas para modelar un proceso de Bernoulli con parámetro p .
2. (**Binomial Negativa**) Demostrar que la suma de k variables aleatorias geométrica(p) independientes tiene la distribución binomial negativa con parámetros (k, p) .
3. (**Aproximación de Poisson**) Para $n = 500$, $p = 0.006$, calcular $P(S_n = 3)$ exactamente y aproximar con Poisson; discutir la precisión.

9 Procesos de Poisson

9.1 Definición y Propiedades Fundamentales

Un **proceso de Poisson** $N(t) : t \geq 0$ con tasa $\lambda > 0$ es un proceso estocástico con valores en $\mathbb{Z}_{\geq 0}$ que satisface los siguientes axiomas [Kingman1993; Ross2019]:

1. **Valor inicial:** $N(0) = 0$.
2. **Incrementos independientes:** Para cualquier $0 \leq t_0 < t_1 < \dots < t_k$, las variables aleatorias $N(t_1) - N(t_0), \dots, N(t_k) - N(t_{k-1})$ son independientes.
3. **Incrementos estacionarios:** Para todo $s, t \geq 0$, la distribución de $N(t + s) - N(s)$ depende solo de t .
4. **Distribución de incrementos:** Para $h > 0$ pequeño,

$$P(N(h) = 1) = \lambda h + o(h), \quad P(N(h) \geq 2) = o(h)$$

donde $o(h)/h \rightarrow 0$ cuando $h \rightarrow 0$.

Observación. Estas condiciones implican que para todo $t \geq 0$,

$$P(N(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n = 0, 1, 2, \dots$$

y que $N(t)$ tiene incrementos estacionarios e independientes.

9.1.1 Demostración Formal de la Ley de Distribución de Poisson

Esbozo: Sea $N(0) = 0$, y dividamos $[0, t]$ en n subintervalos de longitud $h = t/n$. La probabilidad de que cada subintervalo contenga a lo más un evento y que k intervalos contengan exactamente un evento puede mostrarse (usando el límite cuando $n \rightarrow \infty$ y las propiedades anteriores) que converge a la ley de Poisson. Ver [Billingsley1995, Ch. 6] para una demostración completa.

9.1.2 Valores Enteros

Por construcción, $N(t)$ cuenta el número de eventos hasta el tiempo t y es siempre un entero no negativo.

9.2 Tiempos entre Llegadas y la Ley Exponencial

Sea T_1 el tiempo hasta el primer evento. Para $t \geq 0$:

$$P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t}.$$

Por tanto, el tiempo de espera T_1 está distribuido exponencialmente con parámetro λ :

$$f_{T_1}(t) = \lambda e^{-\lambda t}, \quad t \geq 0.$$

Esta propiedad se extiende a todos los tiempos entre llegadas $T_{k+1} - T_k$, que son variables aleatorias exponenciales independientes e idénticamente distribuidas con media $1/\lambda$.

9.2.1 Demostración de la Propiedad de Falta de Memoria

Para $s, t \geq 0$,

$$P(T_1 > s + t \mid T_1 > s) = \frac{P(T_1 > s + t)}{P(T_1 > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(T_1 > t),$$

demostrando que la distribución exponencial carece de memoria [Durrett2019].

9.3 Proceso de Poisson Compuesto

Sea $Y_{i=1}^{\infty}$ una secuencia de variables aleatorias independientes e idénticamente distribuidas, independiente del proceso de Poisson $N(t)$. El **proceso de Poisson compuesto** se define por

$$X(t) = \sum_{i=1}^{N(t)} Y_i, \quad t \geq 0.$$

Supuestos:

- $E|Y_1| < \infty$ para cálculos de media/varianza.

Propiedades:

- $E[X(t)] = \lambda t E[Y_1]$
- $\text{Var}(X(t)) = \lambda t E[Y_1^2]$

Soporte: Para demostraciones y propiedades adicionales, ver [Kingman1993], [Billingsley1995, Sec. 22].

9.4 Aplicaciones

9.4.1 Teoría de Colas

En la **cola M/M/1**, las llegadas se modelan como un proceso de Poisson con tasa λ y los tiempos de servicio están distribuidos exponencialmente. El proceso modela los conteos de llegadas $N(t)$, el tiempo de espera para el próximo cliente, y las probabilidades de estado del sistema. Ver [GrossHarris1998].

Ejemplo: Supongamos que las llegadas ocurren a tasa $\lambda = 3$ por minuto. La probabilidad de que exactamente n llegadas ocurran en $t = 2$ minutos es

$$P(N(2) = n) = \frac{(6)^n}{n!} e^{-6}.$$

Ejemplo: Distribución del Tiempo de Espera. Sea S_n el tiempo de espera hasta la n -ésima llegada. La suma de n variables aleatorias exponenciales i.i.d. está distribuida gamma:

$$P(S_n \leq t) = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad t \geq 0.$$

9.4.2 Teoría de Confiabilidad

Para **procesos de Poisson homogéneos**, las fallas de componentes independientes en sistemas complejos se modelan como eventos de Poisson. Los procesos de Poisson no homogéneos, con función de tasa $\lambda(t)$, se requieren para tasas de falla dependientes de la edad [Ross2019, Ch. 5].

9.4.3 Seguros y Riesgo

En matemáticas de seguros, las **llegadas de reclamos** se modelan como un proceso de Poisson; el tamaño total de reclamos sobre $[0, t]$ es una variable aleatoria de Poisson compuesta.

9.5 Propiedades Avanzadas

9.5.1 Teorema de Superposición

Enunciado: Si $N_1(t)$ y $N_2(t)$ son procesos de Poisson independientes con tasas λ_1 y λ_2 , entonces $N(t) = N_1(t) + N_2(t)$ es un proceso de Poisson con tasa $\lambda_1 + \lambda_2$ [Kingman1993].

Demostración: Se sigue de la convolución de distribuciones de Poisson independientes y la preservación de independencia y estacionariedad.

9.5.2 Teorema de Adelgazamiento

Enunciado: Si cada evento de un proceso de Poisson de tasa λ se retiene independientemente con probabilidad p (es decir, cada evento se asigna una marca Bernoulli(p) independiente), los eventos retenidos forman un proceso de Poisson con tasa λp .

Demostración: Para $h \rightarrow 0$, la probabilidad de que un evento ocurra y sea retenido en $(t, t+h]$ es $\lambda h \cdot p + o(h)$; todas las propiedades del proceso de Poisson se preservan. Ver [Kingman1993, Sec. 2.3].

Ejemplo: Confiabilidad del Sistema. Supongamos que las fallas de componentes de un sistema siguen un proceso de Poisson con tasa $\lambda = 0.1$ fallas/hora.

1. **Probabilidad de 2 fallas en 24 horas:**

$$P(N(24) = 2) = \frac{(0.1 \times 24)^2}{2!} e^{-0.1 \times 24} = \frac{2.4^2}{2} e^{-2.4} \approx 0.261$$

2. **Tiempo esperado hasta la primera falla:**

$$E[T_1] = 1/\lambda = 10 \text{ horas}$$

3. **Ejemplo de Poisson Compuesto:** Si el costo Y_i de cada falla es i.i.d. con $E[Y_1] = 500$, el costo total esperado en $t = 24$ horas es:

$$E[X(24)] = E[N(24)] E[Y_1] = (0.1 \times 24) \times 500 = USD 1,200$$

9.6 Ejercicios

1. **Derivación de la Ley de Poisson:** Demostrar que el único proceso de valores enteros con incrementos estacionarios e independientes, y $P(N(h) = 1) = \lambda h + o(h)$ cuando $h \rightarrow 0$, es el proceso de Poisson homogéneo.
2. **Simulación de Superposición:** Simular dos procesos de Poisson independientes y verificar empíricamente el teorema de superposición.
3. **Varianza de Poisson Compuesto:** Mostrar que $\text{Var}(X(t)) = \lambda t E[Y_1^2]$ para un proceso de Poisson compuesto.

10 Procesos de Markov

10.1 Introducción Conceptual

Un **proceso estocástico** $X_{t \geq 0}$ es una colección de variables aleatorias indexadas por tiempo t y definidas en un espacio de probabilidad común (Ω, \mathcal{F}, P) . El **proceso de Markov** se distingue por la **propiedad de Markov**, que formaliza el concepto de falta de memoria: la evolución futura del proceso depende solo del estado presente, no de la trayectoria tomada para llegar allí.

Definición (Propiedad de Markov): Un proceso $X_{t \geq 0}$ con espacio de estados S es un proceso de Markov si, para todos los tiempos $0 \leq t_1 < t_2 < \dots < t_n < t$ y estados $i_1, \dots, i_n, i, j \in S$,

$$P(X_{t+s} = j \mid X_t = i, X_{t_n} = i_n, \dots, X_{t_1} = i_1) = P(X_{t+s} = j \mid X_t = i).$$

Un **proceso de Markov homogéneo** (o **cadena de Markov homogénea en el tiempo** si el espacio de estados es discreto) es aquel donde las probabilidades de transición dependen solo del tiempo transcurrido, no del tiempo absoluto:

$$p_{ij}(s) = P(X_{t+s} = j \mid X_t = i), \quad \forall t \geq 0.$$

10.2 Matrices de Transición

Para un espacio de estados finito $S = 1, 2, \dots, m$, el comportamiento de transición se describe por una **matriz de transición** $P(s) = (p_{ij}(s))$ donde

$$p_{ij}(s) = P(X_{t+s} = j \mid X_t = i).$$

Convenciones comunes:

- P típicamente denota la matriz de transición de un paso: $P = P(1)$.
- P^n denota la matriz de transición de n pasos, donde $(P^n)_{ij} = P(X_n = j \mid X_0 = i)$.

Propiedades:

- **No negatividad:** $p_{ij}(s) \geq 0$ para todo i, j, s .
- **Sumas de fila:** $\sum_{j=1}^m p_{ij}(s) = 1$ para todo i .
- **Distribución inicial:** Para una distribución inicial dada μ sobre S , la distribución en el paso n es μP^n .

10.3 Clasificación de Estados y Recurrencia

Los estados de una cadena de Markov pueden clasificarse como:

- **Estado recurrente:** El estado i es recurrente si, partiendo de i , el proceso regresa a i con probabilidad uno. Formalmente, sea $f_{ii} = P(\text{regresar alguna vez a } i \mid X_0 = i)$; i es recurrente si $f_{ii} = 1$.
- **Estado transitorio:** El estado i es transitorio si $f_{ii} < 1$; es decir, hay una probabilidad positiva de nunca regresar a i .
- **Estado absorbente:** El estado i es absorbente si $p_{ii} = 1$ y $p_{ij} = 0$ para $j \neq i$; una vez ingresado, el proceso permanece en i para siempre.

Definición (Recurrencia Positiva): Un estado recurrente i es **positivamente recurrente** si el tiempo esperado de retorno $m_i = \mathbb{E}[\text{primer tiempo de retorno a } i \mid X_0 = i]$ es finito.

- La existencia de una distribución estacionaria está íntimamente conectada con la presencia de estados positivamente recurrentes [Durrett2019].

10.4 Irreducibilidad y Periodicidad

Definición (Irreducibilidad): Una cadena de Markov es **irreducible** si, para cualquier par de estados $i, j \in S$, existe n tal que $p_{ij}(n) > 0$. Es decir, cada estado es accesible desde cualquier otro estado en un número finito de pasos.

Definición (Periodicidad): Un estado i tiene período d si d es el máximo común divisor de todos los $n \geq 1$ tales que $p_{ii}(n) > 0$. La cadena es **aperiódica** si cada estado tiene período 1.

10.5 Distribuciones Estacionarias

Una **distribución estacionaria** (o medida invariante) $\pi = (\pi_1, \dots, \pi_m)$ es un vector de probabilidad que satisface

$$\pi P = \pi,$$

donde $\pi_j \geq 0$ y $\sum_j \pi_j = 1$. Intuitivamente, si la distribución inicial es π , entonces la distribución permanece sin cambios en todos los tiempos futuros.

10.5.1 Teorema de Existencia y Unicidad

Teorema (Perron–Frobenius; Estacionaridad y Recurrencia Positiva): Para una cadena de Markov irreducible con espacio de estados finito:

- Existe una distribución estacionaria única π si y solo si todos los estados son positivamente recurrentes.
- La distribución estacionaria satisface $\pi_j = \frac{1}{\mathbb{E}[\text{tiempo de retorno a } j | X_0 = j]}$.

Esbozo de demostración: Ver [Durrett2019, Ch. 1]; la demostración se basa en irreducibilidad, clasificación de recurrencia, y propiedades de valores propios de matrices estocásticas.

10.6 Ergodicidad y Convergencia

Definición (Ergodicidad): Una cadena de Markov es **ergódica** si es irreducible, aperiódica y positivamente recurrente. En este caso, para todo $i, j \in S$,

$$\lim_{n \rightarrow \infty} p_{ij}(n) = \pi_j,$$

donde π es la distribución estacionaria única [Billingsley1995].

10.7 Ecuación de Chapman–Kolmogorov

Teorema (Ecuación de Chapman–Kolmogorov): Para todo $m, n \geq 0$ y $i, j \in S$,

$$p_{ij}(m+n) = \sum_{k \in S} p_{ik}(m)p_{kj}(n).$$

Demostración: Por la ley de probabilidad total y la propiedad de Markov,

$$p_{ij}(m+n) = P(X_{m+n} = j \mid X_0 = i) = \sum_{k \in S} P(X_{m+n} = j, X_m = k \mid X_0 = i) = \sum_{k \in S} p_{ik}(m)p_{kj}(n).$$

10.8 Reversibilidad

Una distribución ν es **reversible** para una cadena de Markov con matriz de transición P si satisface las **ecuaciones de balance detallado**:

$$\nu_i p_{ij} = \nu_j p_{ji}, \quad \forall i, j \in S.$$

Si tal ν existe, es estacionaria, pero no toda distribución estacionaria es reversible.

Observación: Para verificar reversibilidad, resolver las ecuaciones de balance detallado. Para métodos generales, ver [LevinPeres2017].

Ejemplo Avanzado

Ejemplo: Considerar una cadena de Markov con espacio de estados $S = 1, 2, 3, 4$ y matriz de transición

$$P = \begin{pmatrix} 0.1 & 0.3 & 0.4 & 0.2 \\ 0.3 & 0.2 & 0.4 & 0.1 \\ 0.2 & 0.3 & 0.4 & 0.1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

- El estado 4 es absorbente.
- ¿Es la cadena irreducible? No; el estado 4 no puede abandonarse, por lo que la cadena no es irreducible.

Probabilidad de Absorción en Tres Pasos: Calcular $P(X_3 = 4 \mid X_0 = 1)$.

Primero, calcular $P^3 = P \cdot P \cdot P$ y extraer la entrada $(1, 4)$:

1. $P^1 = P$.
2. $P^2 = P \cdot P$.
3. $P^3 = P^2 \cdot P$.

Alternativamente, enumerar todas las trayectorias de tres pasos de 1 a 4:

Sea $A = 1, 2, 3$ (estados no absorbentes).

- Trayectoria 1: $1 \rightarrow i \rightarrow j \rightarrow 4$, $i, j \in A$.
- $P(X_3 = 4 \mid X_0 = 1) = \sum_{i \in A} p_{1i} \sum_{j \in A} p_{ij} p_{j4}$.

Explícitamente,

$$\begin{aligned}
P(X_3 = 4 \mid X_0 = 1) &= \sum_{i=1}^3 p_{1i} \sum_{j=1}^3 p_{ij} p_{j4} \\
&= p_{11} \left(\sum_{j=1}^3 p_{1j} p_{j4} \right) + p_{12} \left(\sum_{j=1}^3 p_{2j} p_{j4} \right) + p_{13} \left(\sum_{j=1}^3 p_{3j} p_{j4} \right) \\
&= 0.1 \cdot (0.2 \cdot 1 + 0.3 \cdot 0.1 + 0.4 \cdot 0.1) \\
&\quad + 0.3 \cdot (0.4 \cdot 0.1 + 0.2 \cdot 0.1 + 0.4 \cdot 0.1) \\
&\quad + 0.4 \cdot (0.3 \cdot 0.1 + 0.4 \cdot 0.1 + 0.3 \cdot 0.1) \\
&= 0.1 \cdot (0.2 + 0.03 + 0.04) + 0.3 \cdot (0.04 + 0.02 + 0.04) + 0.4 \cdot (0.03 + 0.04 + 0.03) \\
&= 0.1 \cdot 0.27 + 0.3 \cdot 0.10 + 0.4 \cdot 0.10 \\
&= 0.027 + 0.03 + 0.04 \\
&= 0.097.
\end{aligned}$$

Por tanto, la probabilidad es 0.097.

10.9 Ejercicios

1. **Demostración:** Mostrar que para una cadena de Markov finita e irreducible, la recurrencia positiva de un estado implica que todos los estados son positivamente recurrentes.
2. **Cálculo:** Para el ejemplo anterior, calcular la distribución estacionaria para la cadena restringida a los estados no absorbentes 1, 2, 3.
3. **Clasificación:** Para una cadena de Markov finita dada, determinar los períodos de todos los estados y clasificar cada uno como recurrente, transitorio o absorbente.

11 Procesos de Poisson Compuestos

11.1 Introducción Conceptual

En aplicaciones avanzadas de probabilidad—como ciencias actuariales, matemáticas financieras, ingeniería de confiabilidad, y gestión cuantitativa de riesgos—es frecuentemente necesario modelar no solo el número de eventos aleatorios a lo largo del tiempo, sino también la magnitud aleatoria asociada con cada evento. Por ejemplo, en matemáticas de seguros, cada reclamo se caracteriza por su tamaño aleatorio de reclamo (o severidad), y en confiabilidad, cada evento de falla incurre un costo de reparación aleatorio. Para capturar esta doble aleatoriedad, el **proceso de Poisson compuesto** es fundamental.

Un proceso de Poisson compuesto combina dos fuentes de incertidumbre: el número de eventos dentro de un intervalo de tiempo, modelado por un proceso de Poisson, y el tamaño o severidad aleatoria de cada evento, modelado por una secuencia independiente de variables aleatorias idénticamente distribuidas.

11.2 Definición Formal y Propiedades Fundamentales

Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad. Definamos los siguientes objetos:

- Sea $N(t) : t \geq 0$ un proceso de Poisson de tasa $\lambda > 0$, es decir, $N(t)$ cuenta el número de eventos que ocurren en $[0, t]$ con $N(0) = 0$, incrementos estacionarios e independientes, y $N(t) - N(s) \sim \text{Poisson}(\lambda(t - s))$ para $0 \leq s < t$ [Billingsley1995].
- Sea $Y_{i=1}^\infty$ una secuencia de variables aleatorias independientes e idénticamente distribuidas (i.i.d.), independiente de $N(t)$, con $E[|Y_1|] < \infty$ y $E[Y_1^2] < \infty$.

El **proceso de Poisson compuesto** $X(t) : t \geq 0$ se define por

$$X(t) = \sum_{i=1}^{N(t)} Y_i, \quad t \geq 0,$$

con la convención $X(t) = 0$ si $N(t) = 0$. Aquí,

- $N(t)$ es el número de eventos hasta el tiempo t ,
- Y_i es el tamaño del reclamo (o severidad del evento) para el i -ésimo evento.

11.2.1 Incrementos Independientes y Estacionaridad

Proposición: Para cualquier $0 \leq s < t$, el incremento $X(t) - X(s)$ es independiente del pasado $X(u) : u \leq s$, y está distribuido como un proceso de Poisson compuesto con parámetros $\lambda(t - s)$ y la misma distribución Y_i .

Demostración: Sea $M = N(t) - N(s)$, que es independiente de $N(s)$ y $M \sim \text{Poisson}(\lambda(t - s))$. La colección $Y_{N(s)+1}, \dots, Y_{N(t)}$ son independientes tanto de $N(s)$ como de todas las Y_i anteriores, debido a la independencia. Por tanto,

$$X(t) - X(s) = \sum_{i=N(s)+1}^{N(t)} Y_i$$

es una suma de M variables i.i.d., independiente del proceso antes de s , satisfaciendo la estructura de Poisson compuesto [Sato1999]. ■

11.3 Momentos del Proceso de Poisson Compuesto

11.3.1 Esperanza

El valor esperado de $X(t)$ se sigue por linealidad de la esperanza e independencia:

$$E[X(t)] = E\left(\sum_{i=1}^{N(t)} Y_i\right) = E[N(t)]E[Y_1] = \lambda t \cdot E[Y_1]$$

puesto que $N(t) \sim \text{Poisson}(\lambda t)$ y las Y_i son i.i.d. [Durrett2019].

11.3.2 Varianza

Aplicando la ley de varianza total y la segunda identidad de Wald:

$$\text{Var}(X(t)) = E[\text{Var}(X(t) \mid N(t))] + \text{Var}(E[X(t) \mid N(t)]).$$

Dado $N(t) = n$, $X(t) \mid N(t) = n$ es una suma de n Y_i i.i.d., por lo que

$$E[X(t) \mid N(t) = n] = nE[Y_1], \quad \text{Var}(X(t) \mid N(t) = n) = n\text{Var}(Y_1).$$

Por tanto,

$$\begin{aligned} E[\text{Var}(X(t) \mid N(t))] &= E[N(t)] \cdot \text{Var}(Y_1) = \lambda t \cdot \text{Var}(Y_1), \\ \text{Var}(E[X(t) \mid N(t)]) &= \text{Var}(N(t) \cdot E[Y_1]) = \text{Var}(N(t)) \cdot (E[Y_1])^2 = \lambda t (E[Y_1])^2. \end{aligned}$$

Por lo tanto,

$$\text{Var}(X(t)) = \lambda t (\text{Var}(Y_1) + (E[Y_1])^2) = \lambda t E[Y_1^2].$$

donde $E[Y_1^2] = \text{Var}(Y_1) + (E[Y_1])^2$.

11.4 Representación Distribucional y Métodos de Transformadas

La ley de probabilidad de $X(t)$ es la de una suma aleatoria:

$$X(t) = \sum_{i=1}^{N(t)} Y_i$$

con $N(t) \sim \text{Poisson}(\lambda t)$. La función de distribución es

$$P(X(t) \leq x) = \sum_{n=0}^{\infty} P(N(t) = n) \cdot P\left(\sum_{i=1}^n Y_i \leq x\right),$$

donde el término $n = 0$ se interpreta como $P(0 \leq x) = 1$ si $x \geq 0$, 0 en caso contrario.

El cálculo directo es intratable para Y_i general, pero herramientas analíticas están disponibles:

- La **función característica** (transformada de Fourier) de $X(t)$ es

$$\phi_{X(t)}(u) = E[e^{iuX(t)}] = \exp(\lambda t(\phi_Y(u) - 1)),$$

donde $\phi_Y(u) = E[e^{iuY_1}]$ es la función característica de Y_1 .

- Similarmente, la **transformada de Laplace** es

$$E[e^{-sX(t)}] = \exp(\lambda t(E[e^{-sY_1}] - 1)),$$

que permite métodos de inversión numérica y aproximación de punto de silla [Sato1999; Asmussen2000]. La simulación Monte Carlo también se usa frecuentemente para Y_i complejas.

Ejemplo: Supongamos que los reclamos llegan a una aseguradora como un proceso de Poisson de tasa $\lambda = 3$ por semana. Los tamaños de reclamos Y_i son i.i.d. con distribución de Pareto: para $y \geq y_m > 0$,

$$F_Y(y) = 1 - \left(\frac{y_m}{y}\right)^\alpha, \quad \alpha > 1.$$

Sea $y_m = 1000$, $\alpha = 2$. Calcular la media y varianza de reclamos agregados en una semana, y comentar sobre el impacto de colas pesadas.

- La media es

$$E[Y_1] = \frac{\alpha y_m}{\alpha - 1} = \frac{2 \times 1000}{2 - 1} = 2000.$$

- El segundo momento es

$$E[Y_1^2] = \begin{cases} \frac{\alpha y_m^2}{\alpha - 2}, & \alpha > 2 \\ \infty, & 1 < \alpha \leq 2 \end{cases}$$

Puesto que $\alpha = 2$, $E[Y_1^2] = \infty$; por tanto, el agregado semanal tiene media finita $E[X(1)] = 6000$, pero **varianza infinita**. Esto demuestra el impacto de reclamos de cola pesada: la media permanece bien definida, pero medidas de riesgo como la varianza no lo están.

11.5 Ejercicios

1. **Simulación y Análisis de Poisson Compuesto:** Sea $\lambda = 2$ eventos por hora, y $Y_i \sim \text{Exp}(1/500)$ independientemente.
 - (a) Simular $X(1)$ (reclamos agregados en una hora) 10,000 veces y estimar la media y varianza.
 - (b) Comparar sus estimaciones empíricas con las fórmulas analíticas.
 - (c) Discutir cómo cambiarían los resultados si las Y_i tuvieran cola pesada.

12 Series de Tiempo

12.1 Definiciones Fundamentales

Una **serie de tiempo** es una colección de variables aleatorias $Y_t : t \in T$, donde $T \subseteq \mathbb{Z}$, representando observaciones secuenciales indexadas por tiempo discreto. Formalmente, estas variables están definidas en un espacio de probabilidad (Ω, \mathcal{F}, P) , con cada $Y_t : \Omega \rightarrow \mathbb{R}$ medible. El objetivo central del análisis de series de tiempo es modelar, inferir y predecir la estructura temporal y dinámica incrustada en tales procesos estocásticos, aprovechando la dependencia temporal típicamente ausente en contextos estadísticos clásicos [Hamilton1994].

12.2 Descomposición de Series de Tiempo

Un paso fundamental en el modelado de series de tiempo es la descomposición del proceso observado en componentes interpretables, cada uno capturando diferentes fuentes de variación. El valor observado Y_t se expresa típicamente como una combinación de:

- **Tendencia (T_t):** Un componente estocástico o determinístico que representa la progresión a largo plazo del proceso. T_t puede modelarse como una función determinística (ej., lineal, cuadrática) o como un proceso aleatorio para capturar cambios persistentes.
- **Estacionalidad (S_t):** Captura efectos cíclicos o periódicos con frecuencias conocidas y fijas (ej., anual o semanal), que también pueden modelarse como procesos determinísticos o estocásticos.
- **Componente Irregular (I_t):** Representa fluctuaciones aleatorias impredecibles no explicadas por la tendencia o estacionalidad, modelado como un proceso estocástico.

Formalmente, se usan dos esquemas de descomposición estándar [Hyndman2021]:

- **Modelo aditivo:** $Y_t = T_t + S_t + I_t$
- **Modelo multiplicativo:** $Y_t = T_t \times S_t \times I_t$

La forma aditiva es apropiada cuando la varianza de Y_t es aproximadamente constante en el tiempo, mientras que la forma multiplicativa es preferible cuando la varianza aumenta con el nivel. Técnicas avanzadas de descomposición como STL (descomposición Estacional-Tendencia usando Loess) y X-13ARIMA-SEATS se emplean comúnmente en la práctica [Cleveland1990].

12.3 Estacionaridad

La **estacionaridad** es una propiedad central en la teoría y aplicación de series de tiempo. Hay dos formas principales:

- **Estacionaridad estricta (fuerte):** Un proceso Y_t es estrictamente estacionario si la distribución conjunta de cualquier conjunto finito $(Y_{t_1}, \dots, Y_{t_n})$ es invariante a desplazamientos en el tiempo. Es decir, para todo n , t_1, \dots, t_n , y h ,

$$F_{Y_{t_1}, \dots, Y_{t_n}}(y_1, \dots, y_n) = F_{Y_{t_1+h}, \dots, Y_{t_n+h}}(y_1, \dots, y_n).$$

- **Estacionaridad débil (de segundo orden):** Y_t es débilmente estacionario si:
 1. $E[Y_t] = \mu$ es constante para todo t ,
 2. $Var(Y_t) = \sigma^2$ es constante para todo t ,
 3. $Cov(Y_t, Y_{t+h}) = \gamma(h)$ depende solo del rezago h , no de t .

En la práctica, la estacionaridad débil es mucho más fácil de verificar y es suficiente para la mayoría de enfoques de modelado lineal. La ergodicidad, que asegura que los promedios temporales converjan a promedios de ensamble, es una condición más estricta pero no se requiere para la mayoría de inferencias estadísticas [Billingsley1995].

Ejemplo: Estacionaridad Estricta vs. Débil.

Sea Y_t una secuencia donde $Y_t = Z$ para todo t y $Z \sim \mathcal{N}(0, 1)$. Este proceso es estrictamente estacionario puesto que todas las distribuciones de dimensión finita son invariantes bajo desplazamiento temporal. En contraste, un proceso con $Y_t = e^{i\omega t}X$ para X aleatorio y ω fijo es débilmente estacionario (media constante, autocovarianza depende solo de h), pero su distribución conjunta varía con t , por lo que no es estrictamente estacionario [Durrett2019].

12.4 Ruido Blanco

Un proceso de **ruido blanco** e_t es una secuencia de variables aleatorias independientes e idénticamente distribuidas (iid) con media cero y varianza constante, típicamente asumidas gaussianas:

$$E[e_t] = 0, \quad Var(e_t) = \sigma^2, \quad Cov(e_t, e_s) = 0 \text{ para } t \neq s.$$

El ruido blanco es estrictamente estacionario (puesto que todas las distribuciones de dimensión finita son invariantes bajo desplazamiento temporal), y trivialmente débilmente estacionario [BoxJenkins1970]. Para secuencias iid no gaussianas, el ruido blanco sigue siendo estrictamente estacionario, pero algunos autores reservan “ruido blanco” para el caso gaussiano.

12.4.1 Demostración: El Ruido Blanco es Débilmente Estacionario

Dado e_t iid con $E[e_t] = 0$ y $Var(e_t) = \sigma^2$:

- $E[e_t] = 0$ para todo t ,
- $Var(e_t) = \sigma^2$ para todo t ,
- $Cov(e_t, e_{t+h}) = 0$ para todo $h \neq 0$. Por tanto, todas las condiciones para estacionaridad débil se satisfacen.

12.5 Métodos de Preprocesamiento

Transformar una serie de tiempo no estacionaria en una estacionaria es frecuentemente un prerequisite para modelado efectivo. Dos métodos centrales son:

- **Promedios Móviles:** Suavizan fluctuaciones a corto plazo vía

$$T_t = \frac{1}{2k+1} \sum_{j=-k}^k Y_{t+j}.$$

Esta operación ayuda a aislar tendencias a más largo plazo y periodicidad [Hyndman2021].

- **Diferenciación:** El operador de diferencia ∇ o Δ se usa para remover tendencias:

$$\Delta Y_t = Y_t - Y_{t-1}, \quad \Delta^p Y_t = Y_t - Y_{t-p}.$$

Diferencias de orden superior pueden remover tendencias polinomiales de grado creciente. El operador de retroceso general B se define por $BY_t = Y_{t-1}$, por lo que $\Delta = 1 - B$. Transformaciones estabilizadoras de varianza, como la transformada de Box-Cox, también se usan comúnmente [BoxCox1964].

Ejemplo: Diferenciando una Caminata Aleatoria. Sea $Y_t = Y_{t-1} + e_t$ con $Y_0 = 0$ y $e_t \sim \text{iid}(0, \sigma^2)$. La diferenciación produce $\Delta Y_t = Y_t - Y_{t-1} = e_t$, que es ruido blanco estacionario.

12.6 Caminata Aleatoria

Una **caminata aleatoria** se define por $Y_t = Y_{t-1} + e_t$ para $e_t \sim \text{iid}(0, \sigma^2)$. Su media es constante, pero su varianza crece linealmente con t :

$$Var(Y_t) = t\sigma^2.$$

Por tanto, el proceso es no estacionario, con memoria infinita—cada innovación e_t afecta permanentemente todos los valores futuros, haciendo la predicción a largo plazo altamente incierta [Hamilton1994].

12.7 Modelos ARIMA

Un modelo **ARIMA**(p, d, q) (Autoregresivo Integrado de Promedio Móvil) generaliza tanto procesos autoregresivos como de promedio móvil, acomodando no estacionaridad vía diferenciación [BoxJenkins1970]. El modelo general se escribe como:

$$\Phi(B)\nabla^d Y_t = \Theta(B)e_t, \quad e_t \sim \text{iid}(0, \sigma^2)$$

donde

- B es el operador de retroceso ($BY_t = Y_{t-1}$),
- $\nabla^d = (1 - B)^d$ es el operador de diferencia d -ésima,
- $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$,
- $\Theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$.

Aquí, p es el orden autoregresivo, d es el número de diferencias para estacionaridad, y q es el orden de promedio móvil. La identificación del modelo típicamente se basa en análisis de la función de autocorrelación (ACF), función de autocorrelación parcial (PACF), y criterios de información como AIC y BIC [Hyndman2021].

12.8 Proceso Autoregresivo (AR)

Un **proceso autoregresivo de orden p** (**AR**(p)) se define por:

$$Y_t = \mu + \sum_{i=1}^p \phi_i Y_{t-i} + e_t, \quad e_t \sim \text{iid}(0, \sigma^2).$$

El proceso es estacionario si y solo si las raíces del polinomio característico $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ yacen fuera del círculo unitario ($|z| > 1$).

12.8.1 Demostración: Condición de Estacionaridad

El proceso **AR**(p) puede escribirse como $\Phi(B)Y_t = e_t$. La estacionaridad requiere que las soluciones de la ecuación en diferencias homogénea decaigan a cero, lo que ocurre si y solo si todas las raíces de $\Phi(z) = 0$ tienen módulo mayor que uno [Hamilton1994].

Ejemplo: Modelo AR(2) Considerar $Y_t = 0.5Y_{t-1} - 0.3Y_{t-2} + e_t$, $e_t \sim \mathcal{N}(0, 1)$. Datos simulados de este modelo tendrán una ACF que se corta después del rezago 2 y una PACF que se desvanece, característico del comportamiento **AR**(2) [Hyndman2021].

12.9 Proceso de Promedio Móvil (MA)

Un proceso de promedio móvil de orden q ($\text{MA}(q)$) se define como:

$$Y_t = \mu + \sum_{i=1}^q \theta_i e_{t-i} + e_t, \quad e_t \sim \text{iid}(0, \sigma^2).$$

La invertibilidad se cumple si las raíces del polinomio $\Theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ yacen dentro del círculo unitario ($|z| < 1$). Los procesos $\text{MA}(q)$ son siempre estacionarios por construcción.

12.9.1 Demostración: Autocovarianza de $\text{MA}(q)$

La función de autocovarianza es:

$$\gamma(h) = \begin{cases} \sigma^2 \left(1 + \sum_{i=1}^q \theta_i^2\right), & h = 0 \\ \sigma^2 \sum_{i=1}^{q-h} \theta_i \theta_{i+h}, & 1 \leq h \leq q \\ 0, & h > q \end{cases}$$

12.9.2 Ejemplo: Ajuste de Modelo $\text{MA}(1)$

Suponer $Y_t = e_t + 0.6e_{t-1}$, $e_t \sim \mathcal{N}(0, 1)$. Ajustar este modelo a datos reales (ej., retornos diarios de un índice financiero) produce una ACF que cae a cero después del rezago 1, indicando la apropiabilidad de la especificación $\text{MA}(1)$.

12.10 Extensiones de Modelos ARIMA

Varias extensiones avanzadas de modelos ARIMA abordan fenómenos específicos:

- **SARIMA:** Incorpora términos autoregresivos y de promedio móvil estacionales para modelar periodicidades.
- **ARFIMA:** Permite diferenciación fraccionaria (d no entero) para capturar procesos de memoria larga [Beran1994].
- **ARIMAX:** Añade covariables exógenas para explicar variación no explicada por la estructura autoregresiva/promedio móvil [Hyndman2021].

12.11 Ejercicios

1. *Descomposición*: Dada una serie mensual con fuerte estacionalidad, describir la descomposición STL y calcular los componentes de tendencia, estacional e irregular.
2. *Estacionaridad*: Demostrar que la primera diferencia de una caminata aleatoria produce un proceso estacionario.
3. *Identificación AR*: Simular un proceso $AR(2)$, graficar su ACF y PACF, y verificar la condición de estacionaridad.
4. *Modelo MA*: Para un proceso $MA(1)$, derivar y graficar la función de autocorrelación.
5. *Selección de Modelo*: Dados gráficos empíricos de ACF/PACF, seleccionar y justificar un modelo $ARIMA(p, d, q)$ apropiado.

13 Modelos ARIMA: Fundamentos, Teoría y Aplicación

13.1 Estacionaridad

Un proceso estocástico $X_{t \in \mathbb{Z}}$ es **estrictamente estacionario** si para cada $n \in \mathbb{N}$ y cada conjunto de índices temporales t_1, \dots, t_n y $k \in \mathbb{Z}$, la distribución conjunta de $(X_{t_1}, \dots, X_{t_n})$ es igual a la de $(X_{t_1+k}, \dots, X_{t_n+k})$ [Billingsley1995].

Estacionaridad débil (de covarianza) requiere que:

- $E[X_t] = \mu$ (media constante para todo t),
- $\text{Var}[X_t] = \sigma^2 < \infty$ (varianza constante para todo t),
- $\text{Cov}(X_t, X_{t+h}) = \gamma(h)$ (autocovarianza depende solo del rezago h).

Ejemplo: Considerar el proceso $X_t = \epsilon_t$ con $\epsilon_t \sim \mathcal{N}(0, 1)$ i.i.d. Este proceso es estricta y débilmente estacionario. En contraste, $Y_t = t + \epsilon_t$ es no estacionario en ambos sentidos, ya que su media depende de t .

Ejercicio: Sea $Z_t = 0.5Z_{t-1} + \epsilon_t$ con $Z_0 = 0$ y $\epsilon_t \sim \mathcal{N}(0, 1)$ i.i.d. ¿Es Z_t débilmente estacionario? Justificar analíticamente.

13.2 Ruido Blanco

Un proceso de **ruido blanco** e_t es una secuencia de variables aleatorias no correlacionadas que satisface:

- $E[e_t] = 0$,
- $\text{Var}[e_t] = \sigma_e^2$,
- $\text{Cov}(e_t, e_s) = 0$ para todo $t \neq s$.

Si $e_t \sim \mathcal{N}(0, \sigma_e^2)$ i.i.d., se llama **ruido blanco gaussiano**.

Demostración: Un proceso de ruido blanco con innovaciones i.i.d. es estrictamente estacionario: Para cualquier colección $(e_{t_1}, \dots, e_{t_n})$, su distribución conjunta es invariante a desplazamientos temporales, puesto que todas las e_t son i.i.d. [Durrett2019].

Ejemplo: Simular $e_t \sim \mathcal{N}(0, 1)$ para $t = 1, \dots, 100$. La secuencia es ruido blanco; las autocorrelaciones empíricas son aproximadamente cero para $h \neq 0$.

Ejercicio: Dada una secuencia de residuos de un modelo ARIMA ajustado, proponer y justificar una prueba estadística para ruido blanco.

13.3 Caminatas Aleatorias y Raíces Unitarias

Una **caminata aleatoria** Y_t se define por $Y_t = Y_{t-1} + e_t$, con e_t como ruido blanco. Su varianza crece sin límite:

$$\text{Var}(Y_t) = t \cdot \sigma_e^2,$$

que es una característica distintiva de la no estacionaridad.

La presencia de una **raíz unitaria** (la raíz en $z = 1$ en el polinomio AR) significa que se requiere diferenciación para estacionaridad.

Ejemplo: Sea $Y_0 = 0$, $e_t \sim \mathcal{N}(0, 1)$. Simular Y_t para $t = 1, \dots, 100$; $\text{Var}(Y_{100}) = 100$.

Ejercicio: Dado $W_t = W_{t-1} + \epsilon_t$ con $W_0 = 0$, calcular $\text{Cov}(W_t, W_s)$ y discutir estacionaridad.

13.4 Preprocesamiento y Suavizado

Sea $y_{t=1}^N$ una serie de tiempo.

Promedio Móvil Simple (PMS):

$$S_t = \frac{1}{2k+1} \sum_{j=-k}^k y_{t+j}, \quad k \in \mathbb{N}, k < t \leq N - k$$

Mediana Móvil:

$$M_t = \text{mediana}(y_{t-k}, \dots, y_{t+k})$$

La mediana móvil es robusta a valores atípicos, mientras que el PMS no lo es.

Nota de Coherencia: Usar notación consistente y_t para la serie de tiempo observada.

Ejemplo: Supongamos que y_t es temperatura diaria. El suavizado con $k = 3$ produce S_t y M_t que reducen fluctuaciones a corto plazo. A continuación se presenta una demostración con $y_t = 20 + \sin(2\pi t/30) + \epsilon_t$ simulado, $\epsilon_t \sim \mathcal{N}(0, 1)$.

Ejercicio: Aplicar tanto PMS como mediana móvil a un conjunto de datos ruidoso. Comparar e interpretar los resultados.

13.5 Diferenciación para Estacionaridad

El operador de **primera diferencia** se define como

$$\Delta y_t = y_t - y_{t-1}$$

La **diferencia estacional** (para período p) es

$$\Delta_p y_t = y_t - y_{t-p}$$

Demostración: Para una caminata aleatoria $Y_t = Y_{t-1} + e_t$,

$$\Delta Y_t = Y_t - Y_{t-1} = e_t$$

Dado que e_t es ruido blanco (estacionario), diferenciar la caminata aleatoria produce un proceso estacionario.

13.6 Descomposición Aditiva y Multiplicativa

Una serie de tiempo puede descomponerse como:

Modelo Aditivo: $y_t = T_t + S_t + I_t$

Modelo Multiplicativo: $y_t = T_t \cdot S_t \cdot I_t$

Donde T_t es tendencia, S_t es estacional, I_t es irregular.

Criterios: Usar el modelo aditivo cuando las fluctuaciones estacionales son aproximadamente constantes; usar el modelo multiplicativo si la variación estacional crece con la tendencia [Hyndman2021]. Una transformación logarítmica puede convertir multiplicativo a aditivo.

Ejemplo: Dadas ventas mensuales con una clara tendencia ascendente y amplitud estacional creciente, el modelo multiplicativo es apropiado. Descomponer un $y_t = (10 + 0.5t) \times (1 + 0.2 \sin(2\pi t/12)) + \epsilon_t$ simulado.

Ejercicio: Descomponer una serie de tiempo proporcionada usando enfoques tanto aditivos como multiplicativos. Justificar el modelo apropiado.

13.7 El Modelo ARIMA: Definición y Propiedades

Sea B el **operador de retroceso**: $By_t = y_{t-1}$, $B^k y_t = y_{t-k}$.

Un modelo **ARIMA**(p, d, q) está dado por

$$\Phi(B)(1 - B)^d y_t = \mu + \Theta(B)e_t,$$

donde

- $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ es el polinomio autoregresivo (AR),
- $(1 - B)^d$ es el operador de diferencia de orden d ,
- $\Theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ es el polinomio de promedio móvil (MA),
- μ es una constante, y
- e_t es ruido blanco.

Identificación y Estimación del Modelo El procedimiento estándar sigue la metodología de Box-Jenkins:

1. Identificación del modelo vía ACF/PACF y pruebas de raíz unitaria (ej., Dickey-Fuller Aumentado, Phillips-Perron) [Hamilton1994].
2. Estimación de parámetros, típicamente vía máxima verosimilitud.
3. Diagnóstico del modelo (análisis de residuos).

Estacionaridad e Invertibilidad:

- **Estacionaridad:** Todas las raíces de $\Phi(B)$ deben yacer fuera del círculo unitario en el plano complejo [Durrett2019].
- **Invertibilidad:** Todas las raíces de $\Theta(B)$ también deben yacer fuera del círculo unitario.

13.8 Modelos Autoregresivos (AR)

Un proceso **AR**(p) satisface

$$y_t = \mu + \sum_{i=1}^p \phi_i y_{t-i} + e_t$$

Demostración de las Ecuaciones de Yule-Walker: Sea $\gamma(h)$ la autocovarianza en el rezago h .

$$\gamma(h) = \sum_{i=1}^p \phi_i \gamma(h-i) + \sigma_e^2 \delta_{h,0}$$

para $h \geq 0$, donde $\delta_{h,0}$ es la delta de Kronecker.

Ejemplo: Simular **AR**(2): $y_t = 0.5y_{t-1} - 0.3y_{t-2} + e_t$. Analizar la ACF y PACF muestrales.

Ejercicio: Estimar parámetros para un proceso AR(2) dado una muestra. Probar la condición de estacionaridad.

13.9 Modelos de Media Móvil (MA)

Un proceso MA(q) es

$$y_t = \mu + e_t + \sum_{j=1}^q \theta_j e_{t-j}$$

Invertibilidad: Las raíces de $\Theta(B)$ deben yacer fuera del círculo unitario para garantizar una representación única en términos de y_t 's pasados [Hamilton1994].

Demostración de la Estructura de Autocorrelación: La ACF de MA(q) es cero más allá del rezago q .

Ejercicio: Dada una muestra de un proceso MA(1), estimar θ_1 y probar la invertibilidad.

13.10 Procesos Integrados

Una serie es **integrada de orden d** ($I(d)$) si $(1 - B)^d y_t$ es estacionaria pero $(1 - B)^{d-1} y_t$ no lo es.

Pruebas ADF/PP: Las pruebas de Dickey-Fuller Aumentado y Phillips-Perron se usan para determinar el número mínimo de diferencias requeridas para estacionaridad [Hamilton1994].

Nota de Terminología: Usar “diferencias” o “diferenciación”—no “diferenciaciones”—para series de tiempo discretas.

13.11 Ejercicios

1. Demostrar que diferenciar una caminata aleatoria produce un proceso estacionario.
2. Dados datos de series de tiempo, probar la presencia de raíces unitarias y determinar el orden de integración usando la prueba ADF.
3. Simular y graficar un proceso AR(2) y MA(1). Analizar la ACF y PACF muestrales.
4. Para una serie de residuos dada, realizar una prueba de Ljung-Box para ruido blanco.