Fundamentos de Probabilidad y Funciones de Distribución

Notas de Probabilidad

26 de septiembre de 2025

1. Teoría de Conjuntos en Probabilidad

La probabilidad se construye sobre la teoría de conjuntos. Algunos conceptos fundamentales:

- **Espacio muestral** (Ω): conjunto de todos los resultados posibles de un experimento aleatorio.
- **Evento**: subconjunto de Ω.
- **Complemento**: $A^c = \Omega \setminus A$.
- **Unión**: $A \cup B = \{x \in \Omega : x \in A \text{ o } x \in B\}$.
- **Intersección**: $A \cap B = \{x \in \Omega : x \in A \text{ y } x \in B\}$.

2. Conceptos Fundamentales de Probabilidad

2.1. Experimento Aleatorio

Un **experimento aleatorio** es un proceso o procedimiento que:

- Puede repetirse indefinidamente bajo las mismas condiciones
- Tiene un conjunto bien definido de resultados posibles (espacio muestral)
- No se puede predecir con certeza cuál será el resultado antes de realizarlo
- Está sujeto a variabilidad aleatoria inherente

Características fundamentales:

- **Reproducibilidad**: Se puede repetir en condiciones idénticas o similares
- **Incertidumbre**: El resultado no es determinístico
- **Regularidad estadística**: Aunque individual es impredecible, en muchas repeticiones emergen patrones

Ejemplos de experimentos aleatorios:

- Lanzar una moneda o un dado
- Medir el tiempo de vida de un componente electrónico
- Seleccionar una muestra aleatoria de una población
- Contar el número de llamadas a un centro de atención en una hora
- Medir el error en una transmisión de datos

Elementos de un experimento aleatorio:

1. **Espacio muestral** (Ω): Conjunto de todos los resultados posibles
2. **Eventos**: Subconjuntos del espacio muestral de interés
3. **Probabilidad**: Medida de la posibilidad de ocurrencia de cada evento

Experimento determinístico vs. aleatorio:

- **Determinístico**: El resultado está completamente determinado por las condiciones iniciales (ej: caída libre en el vacío)
- **Aleatorio**: Incluso conociendo todas las condiciones, existe incertidumbre en el resultado

La teoría de probabilidad proporciona el marco matemático para analizar y modelar experimentos aleatorios de manera rigurosa.

2.2. Variable Aleatoria

Una **variable aleatoria** es una función que asigna un valor numérico a cada resultado de un experimento aleatorio. Formalmente, es una función $X : \Omega \to \mathbb{R}$ que mapea el espacio muestral a los números reales.

Tipos de variables aleatorias:

- **Discretas**: Toman valores contables (finitos o infinitos numerables). Ejemplos: número de caras al lanzar monedas, número de clientes en una cola.
- **Continuas**: Pueden tomar cualquier valor en un intervalo de números reales. Ejemplos: tiempo de espera, altura, temperatura.

Interpretación: Una variable aleatoria no es un número fijo, sino una función cuyo valor depende del resultado del experimento aleatorio. Antes de realizar el experimento, su valor es incierto; después, toma un valor específico llamado realización.

2.3. Eventos Independientes

Dos eventos A y B son **independientes** si la ocurrencia de uno no afecta la probabilidad de ocurrencia del otro. Matemáticamente:

$$P(A \cap B) = P(A) \cdot P(B) \tag{1}$$

Equivalentemente, si $P(B) > 0$:

$$P(A|B) = P(A) \tag{2}$$

Independencia de variables aleatorias:
Dos variables aleatorias X e Y son independientes si:

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y) \quad \forall x, y \in \mathbb{R} \tag{3}$$

Para variables discretas: $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$
Para variables continuas: $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$
Propiedades importantes:

- Si X e Y son independientes: $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$
- Para n eventos independientes: $P(A_1 \cap A_2 \cap \cdots \cap A_n) = \prod_{i=1}^{n} P(A_i)$
- La independencia es simétrica: si A es independiente de B, entonces B es independiente de A

2.4. Proceso Estocástico

Un **proceso estocástico** (o proceso aleatorio) es una colección de variables aleatorias indexadas por un parámetro, típicamente el tiempo. Se denota como $\{X_t : t \in T\}$, donde T es el conjunto de índices.

Clasificación según el conjunto de índices:

- **Tiempo discreto**: $T = \{0, 1, 2, \ldots\}$ o $\mathbb{Z}$. Ejemplo: $\{X_n\}_{n=0}^{\infty}$
- **Tiempo continuo**: $T = [0, \infty)$ o $\mathbb{R}$. Ejemplo: $\{X(t)\}_{t \geq 0}$

Clasificación según el espacio de estados:

- **Espacio de estados discreto**: Las variables toman valores en un conjunto contable
- **Espacio de estados continuo**: Las variables toman valores en $\mathbb{R}$ o $\mathbb{R}^n$

Ejemplos de procesos estocásticos:

- **Caminata aleatoria**: Suma acumulada de variables aleatorias independientes
- **Proceso de Poisson**: Cuenta el número de eventos en intervalos de tiempo
- **Movimiento Browniano**: Modelo de partículas en suspensión, fundamental en finanzas
- **Cadenas de Markov**: Procesos donde el futuro depende solo del presente (propiedad de Markov)

Interpretación: Un proceso estocástico modela sistemas que evolucionan de manera aleatoria en el tiempo. Para cada instante t, X_t es una variable aleatoria, y para cada resultado $\omega \in \Omega$, $X_t(\omega)$ describe una trayectoria o realización del proceso.

Caracterización: Un proceso estocástico se caracteriza completamente por sus distribuciones finito-dimensionales:

$$F_{t_1,t_2,\ldots,t_n}(x_1, x_2, \ldots, x_n) = P(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \ldots, X_{t_n} \leq x_n) \tag{4}$$

para cualquier colección finita de tiempos $t_1 < t_2 < \cdots < t_n$.

3. Función de Probabilidad

Una función de probabilidad asigna a cada evento un número real entre 0 y 1 que mide la posibilidad de que dicho evento ocurra.

3.1. Funciones de Masa y Densidad

- **Función de Masa de Probabilidad (pmf)**: utilizada para variables aleatorias discretas.
$$P(X = x_i) = p(x_i), \quad \sum_i p(x_i) = 1 \tag{5}$$

- **Función de Densidad de Probabilidad (pdf)**: utilizada para variables aleatorias continuas.
$$P(a \leq X \leq b) = \int_a^b f(x)\,dx, \quad \int_{-\infty}^{\infty} f(x)\,dx = 1 \tag{6}$$

3.2. Función de Distribución Acumulada (CDF)

La función de distribución acumulada (CDF) es una función fundamental en probabilidad que proporciona la probabilidad de que una variable aleatoria X tome un valor menor o igual a x:
$$F(x) = P(X \leq x) \tag{7}$$

3.2.1. Propiedades de la CDF

- $\lim_{x\to-\infty} F(x) = 0$ y $\lim_{x\to\infty} F(x) = 1$
- $F(x)$ es no decreciente: si $x_1 < x_2$, entonces $F(x_1) \leq F(x_2)$
- $F(x)$ es continua por la derecha: $\lim_{h\to0^+} F(x + h) = F(x)$
- $P(a < X \leq b) = F(b) - F(a)$

3.2.2. Relación con la PMF (Variables Discretas)

Para una variable aleatoria discreta con función de masa de probabilidad $p(x_i)$:

$$F(x) = \sum_{x_i \leq x} p(x_i) = \sum_{x_i \leq x} P(X = x_i) \tag{8}$$

La CDF es una función escalonada que presenta saltos en los puntos donde la variable aleatoria tiene masa de probabilidad. El tamaño del salto en x_i es exactamente $p(x_i)$.

Inversamente, podemos recuperar la pmf desde la CDF:

$$p(x_i) = F(x_i) - \lim_{x \to x_i^-} F(x) = F(x_i) - F(x_i^-) \tag{9}$$

3.2.3. Relación con la PDF (Variables Continuas)

Para una variable aleatoria continua con función de densidad $f(x)$:

$$F(x) = \int_{-\infty}^{x} f(t)\, dt \tag{10}$$

La pdf se puede obtener derivando la CDF (cuando la derivada existe):

$$f(x) = \frac{dF(x)}{dx} = F'(x) \tag{11}$$

Esta relación es fundamental: la pdf es la derivada de la CDF, y la CDF es la integral de la pdf. Esto significa que:

- La pdf nos da la "densidad"de probabilidad en cada punto
- La CDF nos da la probabilidad acumulada hasta cada punto
- El área bajo la curva de la pdf entre dos puntos a y b es igual a $F(b) - F(a)$

3.2.4. Ejemplo Ilustrativo

Consideremos una variable aleatoria exponencial con parámetro λ:

- PDF: $f(x) = \lambda e^{-\lambda x}$ para $x \geq 0$
- CDF: $F(x) = 1 - e^{-\lambda x}$ para $x \geq 0$

Podemos verificar que $F'(x) = \lambda e^{-\lambda x} = f(x)$.

La CDF caracteriza completamente la distribución de una variable aleatoria y es especialmente útil para:

- Calcular probabilidades de intervalos
- Generar números aleatorios mediante el método de la transformada inversa
- Comparar distribuciones mediante pruebas estadísticas

4. Medidas de Tendencia y Dispersión

Las medidas de tendencia central y dispersión son fundamentales para caracterizar el comportamiento de una variable aleatoria.

4.1. Esperanza Matemática

La **esperanza** o **valor esperado** de una variable aleatoria X, denotada como $\mathbb{E}[X]$ o μ, representa el valor promedio que esperamos obtener si repetimos el experimento aleatorio un número infinito de veces. Es el çentro de masa"de la distribución de probabilidad.

Interpretación:

- Es el valor promedio ponderado por las probabilidades
- Representa el punto de equilibrio de la distribución
- En juegos de azar, indica la ganancia o pérdida promedio a largo plazo
- No necesariamente es un valor que la variable pueda tomar (ej: esperanza de un dado es 3.5)

Definición matemática:

$$\mathbb{E}[X] = \begin{cases} \sum_i x_i \, p(x_i), & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} x f(x) \, dx, & \text{si } X \text{ es continua} \end{cases} \tag{12}$$

Definición alternativa usando probabilidades de cola:

Para variables aleatorias no negativas ($X \geq 0$), la esperanza también se puede calcular como:

$$\mathbb{E}[X] = \int_0^{\infty} P(X > x) \, dx = \int_0^{\infty} [1 - F(x)] \, dx \tag{13}$$

Para el caso discreto con valores no negativos:

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} P(X > k) = \sum_{k=1}^{\infty} P(X \geq k) \tag{14}$$

Para variables aleatorias que pueden tomar valores positivos y negativos:

$$\mathbb{E}[X] = \int_0^{\infty} P(X > x) \, dx - \int_{-\infty}^{0} P(X < x) \, dx \tag{15}$$

Esta representación es particularmente útil cuando:

- Es más fácil calcular $P(X > x)$ que la densidad o masa de probabilidad
- Se trabaja con tiempos de vida o tiempos de espera
- Se analiza la función de supervivencia $S(x) = P(X > x)$

4.2. Varianza y Desviación Estándar

La **varianza** de una variable aleatoria X, denotada como $\text{Var}(X)$ o σ^2, mide la dispersión de los valores de la variable alrededor de su esperanza. Cuantifica qué tan alejados están típicamente los valores de la media.

Interpretación:

- Mide el promedio de las desviaciones cuadradas respecto a la media

- Una varianza pequeña indica que los valores están concentrados cerca de la media
- Una varianza grande indica que los valores están dispersos
- Siempre es no negativa: $\mathrm{Var}(X) \geq 0$
- $\mathrm{Var}(X) = 0$ si y solo si X es constante (no aleatoria)

Definición matemática:

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \tag{16}$$

La segunda igualdad es la fórmula computacional, a menudo más práctica para calcular.

Desviación estándar:

La desviación estándar $\sigma = \sqrt{\mathrm{Var}(X)}$ tiene las mismas unidades que la variable original, lo que la hace más interpretable que la varianza.

Propiedades importantes:

- $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$ (la varianza no cambia con traslaciones)
- $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$ (linealidad de la esperanza)
- Para variables independientes: $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

4.3. Covarianza y Correlación

4.3.1. Covarianza

La **covarianza** entre dos variables aleatorias X e Y mide el grado de variación conjunta respecto a sus medias:

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \tag{17}$$

Interpretación:

- $\mathrm{Cov}(X, Y) > 0$: Las variables tienden a variar en la misma dirección
- $\mathrm{Cov}(X, Y) < 0$: Las variables tienden a variar en direcciones opuestas
- $\mathrm{Cov}(X, Y) = 0$: No hay relación lineal entre las variables

Propiedades:

- $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$ (simetría)
- $\mathrm{Cov}(aX + b, cY + d) = ac \cdot \mathrm{Cov}(X, Y)$
- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$

4.3.2. Coeficiente de Correlación

El **coeficiente de correlación de Pearson** es una versión normalizada de la covarianza:

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \tag{18}$$

Propiedades:

- $-1 \leq \rho_{X,Y} \leq 1$ (acotado)
- $|\rho_{X,Y}| = 1$ si y solo si existe una relación lineal perfecta: $Y = aX + b$
- $\rho_{X,Y} = 1$: correlación lineal positiva perfecta
- $\rho_{X,Y} = -1$: correlación lineal negativa perfecta
- $\rho_{X,Y} = 0$: no hay correlación lineal (variables no correlacionadas)

4.3.3. Relación entre Independencia y Correlación

La relación entre independencia y correlación es asimétrica:
Teorema fundamental:

$$X \text{ e } Y \text{ independientes} \Rightarrow \mathrm{Cov}(X,Y) = 0 \text{ y } \rho_{X,Y} = 0 \tag{19}$$

Pero la implicación inversa NO es cierta en general:

$$\mathrm{Cov}(X,Y) = 0 \not\Rightarrow X \text{ e } Y \text{ independientes} \tag{20}$$

Explicación:

- La independencia es una condición más fuerte que implica que no hay **ningún tipo** de relación entre las variables
- La correlación cero solo indica ausencia de relación **lineal**
- Pueden existir relaciones no lineales fuertes con correlación cero

Ejemplo clásico de correlación cero sin independencia:
Sea $X \sim \mathcal{U}(-1, 1)$ (uniforme) y $Y = X^2$. Entonces:

- Y está completamente determinada por X (máxima dependencia)
- Sin embargo, $\mathrm{Cov}(X,Y) = \mathbb{E}[X^3] - \mathbb{E}[X]\mathbb{E}[X^2] = 0 - 0 = 0$
- Por tanto, $\rho_{X,Y} = 0$ a pesar de la dependencia perfecta

Caso especial - Variables normales:
Para variables aleatorias con distribución normal conjunta:

$$X, Y \text{ normales conjuntas: } \rho_{X,Y} = 0 \Leftrightarrow X \text{ e } Y \text{ independientes} \tag{21}$$

Este es uno de los pocos casos donde correlación cero implica independencia.
Resumen de implicaciones:

- Independencia $\Rightarrow$ Incorrelación (siempre)
- Incorrelación $\not\Rightarrow$ Independencia (en general)
- Incorrelación $\Rightarrow$ Independencia (solo para normales conjuntas)
- Correlación $\neq 0 \Rightarrow$ Dependencia (por contraposición)

5. Teoremas Fundamentales

5.1. Teorema de la Probabilidad Total

Si $\{B_1, B_2, \dots, B_n\}$ es una partición de Ω, entonces:

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i) \tag{22}$$

5.2. Teorema de Bayes

$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)} \tag{23}$$

6. Distribuciones Importantes

6.1. Experimento de Bernoulli

Variable aleatoria que toma el valor 1 con probabilidad p y 0 con probabilidad $1 - p$.

$$P(X = x) = p^x(1-p)^{1-x}, \quad x \in \{0, 1\} \tag{24}$$

Parámetros:

- Esperanza: $\mathbb{E}[X] = p$
- Varianza: $\mathrm{Var}(X) = p(1-p)$

6.2. Distribución Binomial

Número de éxitos en n ensayos de Bernoulli independientes.

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \dots, n \tag{25}$$

Coeficiente binomial (combinatoria):

El coeficiente binomial $\binom{n}{k}$ representa el número de formas de elegir k elementos de un conjunto de n elementos, sin importar el orden:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdot (n-2) \cdots (n-k+1)}{k \cdot (k-1) \cdot (k-2) \cdots 1} \tag{26}$$

Propiedades importantes:

- $\binom{n}{0} = \binom{n}{n} = 1$
- $\binom{n}{k} = \binom{n}{n-k}$ (simetría)
- $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ (identidad de Pascal)
- $\sum_{k=0}^{n} \binom{n}{k} = 2^n$

Parámetros:

- Esperanza: $\mathbb{E}[X] = np$
- Varianza: $\text{Var}(X) = np(1-p)$

6.3. Distribución de Poisson

Modela el número de eventos que ocurren en un intervalo de tiempo o espacio.
Forma estándar:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots \tag{27}$$

Forma con tasa y tiempo:
Si λ representa la tasa promedio de eventos por unidad de tiempo y t es el intervalo de tiempo considerado:

$$P(X(t) = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \quad k = 0, 1, 2, \ldots \tag{28}$$

donde $X(t)$ es el número de eventos en el intervalo $[0, t]$.
Parámetros:

- Tasa de eventos: λ (eventos por unidad de tiempo)
- Parámetro de la distribución: $\mu = \lambda t$ (número esperado de eventos en tiempo t)
- Esperanza: $\mathbb{E}[X] = \lambda t$
- Varianza: $\text{Var}(X) = \lambda t$

Nota: En la distribución de Poisson, la media y la varianza son siempre iguales. Cuando el contexto es claro, se usa λ directamente como el parámetro que representa el número esperado de eventos.
Notación:

- Para la forma estándar: $X \sim \text{Poisson}(\lambda)$ o $X \sim \mathcal{P}(\lambda)$
- Para la forma con tasa y tiempo: $X(t) \sim \text{Poisson}(\lambda t)$ o $X(t) \sim \mathcal{P}(\lambda t)$

6.4. Distribución Normal

Distribución continua simétrica alrededor de su media μ.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R} \tag{29}$$

Parámetros:

- Esperanza: $\mathbb{E}[X] = \mu$
- Varianza: $\mathrm{Var}(X) = \sigma^2$
- Desviación estándar: σ

La notación común es $X \sim N(\mu, \sigma^2)$.