Probabilidade

Valor Esperado, Momentos, Vetores Aleatórios e Desigualdades

Material adaptado dos slides 7 e 8

2025-07-06

Sumário

1 1. Valor Esperado (Esperança)

A Esperança é o conceito central que quantifica a "localização" de uma variável aleatória.

1.1 Definição Formal e Interpretação

No nível da Teoria da Medida, para um espaço de probabilidade $(\Omega, \mathcal{F}, P)$, a esperança de uma v.a. X é a integral de Lebesgue de X em relação à medida P:

$$E[X] = \int_\Omega X(\omega)\, dP(\omega)$$

Na prática, calculamos esta integral usando a **integral de Lebesgue-Stieltjes** em relação à Função de Distribuição Acumulada (FDA) $F(x)$ de X:

$$E[X] = \int_{-\infty}^{\infty} x\, dF(x)$$

Esta definição geral se especializa para os casos conhecidos:

- **Caso Discreto:** Se X tem função de massa de probabilidade (FMP) $p(x_i)$:

$$E[X] = \sum_i x_i \cdot p(x_i)$$

  A esperança existe se $\sum_i |x_i| p(x_i) < \infty$. Formalmente, a definição requer que pelo menos uma das somas para as partes positiva e negativa de X seja finita.

- **Caso Contínuo:** Se X tem função de densidade (FDP) $f(x)$:

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

  A esperança existe se $\int_{-\infty}^{\infty} |x| f(x)\, dx < \infty$.

Interpretação Geométrica: A esperança pode ser vista como a área líquida entre a função de sobrevivência $S(x) = 1 - F(x)$ e o eixo horizontal:

$$E[X] = \int_0^{\infty} (1 - F(x))\, dx - \int_{-\infty}^{0} F(x)\, dx$$

1.2 Exemplos de Cálculo de Esperança

1.2.1 Exemplo 1: Distribuição Binomial, $X \sim \textbf{Bin}(n, p)$

Usando o "truque do fatorial", $k\binom{n}{k} = n\binom{n-1}{k-1}$:

$$\begin{aligned}
E[X] &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\
&\text{Seja } j = k-1: \\
&= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{(n-1)-j} \\
&= np \cdot (p + (1-p))^{n-1} = np
\end{aligned}$$

1.2.2 Exemplo 2: Distribuição Geométrica, $X \sim \textbf{Geo}(\beta)$

Seja X o número de falhas *antes* do primeiro sucesso, com $P(X = k) = (1-\beta)^k \beta$ para $k = 0, 1, ...$ (Nota: Existem duas parametrizações para a Geométrica. A dos slides parece ser $P(X = k) = (1-\beta)\beta^k$ para $k \geq 1$. Vamos usar a parametrização dos slides para $E[X] = \beta/(1-\beta)$). Seja $p = (1-\beta)$ a prob. de sucesso, e $q = \beta$ a de falha. $P(X = k) = pq^{k-1}$ para $k = 1, 2, ...$

$$E[X] = \sum_{k=1}^{\infty} kpq^{k-1} = p \sum_{k=1}^{\infty} kq^{k-1}$$

Lembramos da soma da série geométrica e sua derivada: $\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$. Derivando em relação a q, obtemos $\sum_{k=1}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2}$.

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

2 2. Momentos, Variância e Propriedades

A esperança de potências de uma v.a., $g(X) = X^k$, define os **momentos**, que caracterizam a forma da distribuição. Para calcular $E[g(X)]$ usamos a **Lei do Estatístico Inconsciente (LOTUS)**:

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

2.1 Momentos Brutos e Centrais

- **k-ésimo Momento (Bruto):** $\mu'_k = E[X^k]$
- **k-ésimo Momento Central:** $\mu_k = E[(X - \mu'_1)^k]$

2.2 Variância

A variância é o segundo momento central, μ_2, e mede a dispersão em torno da média.

- **Definição:** $\mathrm{Var}(X) = \mu_2 = E[(X - E[X])^2]$
- **Fórmula Computacional:** $\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \mu'_2 - (\mu'_1)^2$

2.2.1 Propriedades da Variância

1. **Não-negatividade:** $\mathrm{Var}(X) \geq 0$.
2. **Transformação Afim:** $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$.

2.2.2 Exemplo: Variância da Poisson, $X \sim$ **Poi**(λ)

Já sabemos que $E[X] = \lambda$. Para a variância, precisamos de $E[X^2]$. Usando LOTUS e o truque $k^2 = k(k-1) + k$:

$$\begin{aligned}
E[X^2] &= \sum_{k=0}^{\infty} k^2 \frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty} k \frac{e^{-\lambda}\lambda^k}{(k-1)!} \\
&= \sum_{k=1}^{\infty} [k(k-1) + k] \frac{e^{-\lambda}\lambda^k}{k!} \\
&= \sum_{k=2}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k-2)!} + \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k-1)!} \\
&= \lambda^2 \sum_{k=2}^{\infty} \frac{e^{-\lambda}\lambda^{k-2}}{(k-2)!} + \lambda \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!} \\
&= \lambda^2(1) + \lambda(1) = \lambda^2 + \lambda
\end{aligned}$$

Portanto, a variância é:

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

Para a distribuição de Poisson, média e variância são iguais.

3 3. Medidas de Associação: Covariância e Correlação

3.1 Covariância

Mede a direção da relação *linear* entre duas v.a. X e Y.

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

Propriedade fundamental: $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$. Se X e Y são independentes, $\mathrm{Cov}(X, Y) = 0$, mas a recíproca **não** é verdadeira.

3.1.1 Exemplo: Covariância Zero com Dependência

Seja $X \sim \text{Uniforme}(-1, 1)$, então $E[X] = 0$. Seja $Y = X^2$. Y é claramente dependente de X. $E[Y] = E[X^2] = \int_{-1}^{1} x^2 \frac{1}{2} dx = \frac{1}{2}[\frac{x^3}{3}]_{-1}^{1} = \frac{1}{3}$. $E[XY] = E[X^3] = \int_{-1}^{1} x^3 \frac{1}{2} dx = \frac{1}{2}[\frac{x^4}{4}]_{-1}^{1} = 0$. $\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0 - (0)(\frac{1}{3}) = 0$.

3.2 Coeficiente de Correlação (ρ)

A medida adimensional da associação linear, sempre entre -1 e 1.

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

4 4. Vetores e Matrizes Aleatórias

A teoria se estende elegantemente para o caso multivariado.

4.1 Vetor de Médias

Para um vetor aleatório $\mathbf{X} = [X_1, \dots, X_p]^T$, o vetor de médias μ é:

$$E[\mathbf{X}] = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_p] \end{pmatrix} = \mu$$

4.2 Matriz de Variância-Covariância

A matriz de covariância Σ captura todas as variâncias e covariâncias do vetor $\mathbf{X}$.

$$\Sigma = \text{Var}(\mathbf{X}) = E[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T]$$

Sua estrutura é:

$$\Sigma_{ij} = \text{Cov}(X_i, X_j) \implies \Sigma = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

Propriedades de Σ: 1. **Simétrica:** $\Sigma = \Sigma^T$. 2. **Positiva Semi-Definida:** Para qualquer vetor de constantes $\mathbf{a}$, temos $\mathbf{a}^T \Sigma \mathbf{a} \geq 0$. 3. **Transformação Linear:** Para matriz constante $\mathbf{A}$, $\text{Var}(\mathbf{A}\mathbf{X} + \mathbf{b}) = \mathbf{A}\Sigma\mathbf{A}^T$.

4.2.1 Variância da Soma (Revisitada)

A variância da soma de n variáveis aleatórias pode ser escrita de forma compacta:

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{i<j} \text{Cov}(X_i, X_j)$$

Se as v.a. são não-correlacionadas (um caso especial é a independência), o termo de covariância se anula e $\text{Var}(\sum X_i) = \sum \text{Var}(X_i)$.

5 5. Ferramentas Teóricas e Desigualdades

5.1 Função Geradora de Momentos (FGM)

Se existe, $M_X(t) = E[e^{tX}]$ define unicamente a distribuição e gera momentos através de suas derivadas em $t = 0$:

$$E[X^n] = \left. \frac{d^n}{dt^n} M_X(t) \right|_{t=0}$$

5.2 Desigualdade de Cauchy-Schwarz

Uma das desigualdades mais importantes da matemática, com uma versão probabilística fundamental.

Teorema: Para quaisquer v.a. X e Y com segundos momentos finitos:

$$(E[XY])^2 \leq E[X^2]E[Y^2]$$

Esboço da Prova: A função $g(t) = E[(tX - Y)^2] = t^2E[X^2] - 2tE[XY] + E[Y^2]$ é uma parábola em t que nunca pode ser negativa. Logo, seu discriminante deve ser $\Delta \leq 0$, o que leva diretamente à desigualdade. Esta desigualdade, quando aplicada a variáveis centradas, prova que $|\rho(X, Y)| \leq 1$.

5.3 Desigualdade de Jensen

Conecta o valor esperado de uma função com a função do valor esperado.

Teorema: Se ϕ é uma função **convexa**, então:

$$\phi(E[X]) \leq E[\phi(X)]$$

Prova: A propriedade definidora de uma função convexa ϕ é que ela sempre se situa acima de suas retas tangentes. Seja $a = E[X]$. A reta tangente a ϕ em a é $y(x) = \phi(a) + \phi'(a)(x - a)$. Pela convexidade, temos que $\phi(x) \geq \phi(a) + \phi'(a)(x - a)$ para todo x. Substituindo x pela variável aleatória X e a por $E[X]$:

$$\phi(X) \geq \phi(E[X]) + \phi'(E[X])(X - E[X])$$

Tomando a esperança em ambos os lados e usando a linearidade da esperança:

$$\begin{aligned}
E[\phi(X)] &\geq E[\phi(E[X])] + E[\phi'(E[X])(X - E[X])] \\
&= \phi(E[X]) + \phi'(E[X]) \cdot E[X - E[X]] \\
&= \phi(E[X]) + \phi'(E[X]) \cdot 0 \\
E[\phi(X)] &\geq \phi(E[X])
\end{aligned}$$

Aplicação Imediata: Se $\phi(x) = x^2$ (convexa), temos $(E[X])^2 \leq E[X^2]$, o que é outra forma de mostrar que $\text{Var}(X) \geq 0$.

5.4 Desigualdade de Chebyshev (Generalizada)

Fornece um limite superior para a probabilidade de uma v.a. se afastar de sua média (medida de concentração), usando apenas a variância. Para uma função não-negativa $g(x)$ e um conjunto A:

$$P(X \in A) \leq \frac{E[g(X)]}{\inf_{x \in A} g(x)}$$

A versão clássica usa $g(x) = (x - \mu)^2$ e $A = \{x : |x - \mu| \geq \epsilon\}$, resultando em:

$$P(|X - E[X]| \geq \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}$$

É uma ferramenta teórica poderosa por ser "livre de distribuição".