Métodos Computacionais

Thiago Rodrigo Ramos

12 de agosto de 2025

Sumário

Capítulo 1

Introdução

1.1 Um conselho: a importância de ser ruim antes de ser bom

É natural que, quando começamos a fazer algo, a gente faça essa coisa muito malfeita ou cheia de defeitos. Isso é comum em qualquer processo de aprendizagem, e sempre foi assim, desde o início dos tempos.

Quando comecei a programar em Python, muita coisa sobre a linguagem eu aprendi por conta própria, apesar de já ter feito alguns cursos básicos em C. Programei de forma amadora em Python por muitos anos, até que, no doutorado, precisei aprender a programar de forma mais organizada e profissional. Lembro que, nessa época, um amigo da pós-graduação me apresentou ao "submundo da programação". Foi aí que aprendi muito do que sei hoje sobre terminal do Linux, Git, e foi também quando comecei a usar o Vim.

Uma das coisas que esse amigo me mostrou foi o Pylint, que nada mais é do que um verificador de bugs e qualidade de código para Python. O Pylint é bem rigoroso na análise, e ainda te dá, ao final, uma nota que vai até 10. Nessa fase, apesar de já ter evoluído bastante, meus códigos ainda recebiam notas por volta de 6 ou 7. Resolvi então rodar o Pylint nos meus códigos antigos pra ter uma noção de quão ruins eles eram — e a nota final foi -900. Pois é, existe um limite superior para o quão bem você consegue fazer algo, mas aparentemente o fundo do poço é infinito.

O que eu queria mostrar com essa história é que faz parte do processo de aprendizado ser ruim no começo e melhorar com o tempo. Falo isso porque, hoje em dia, com o crescimento dos LLMs, a gente fica tentado a pular essa etapa de errar muito até acertar, e ir direto pra fase em que escrevemos códigos limpos, bem comentados, identados e organizados. Mas não se enganem: apesar da aparência profissional, depender de LLMs pra escrever tudo atrapalha justamente essa parte essencial de aprender errando.

Neste curso, vários exercícios envolvem escrever códigos em Python. Meu conselho é: não tenham vergonha de errar, de escrever soluções ruins ou confusas. Isso é absolutamente normal. Vocês estão aqui para evoluir — e errar faz parte do processo.

Capítulo 2

Elementos básicos de probabilidade

2.1 Axiomas da probabilidade

Um *espaço de probabilidade* é uma tupla composta por três elementos: o *espaço amostral*, o *conjunto de eventos* e uma *distribuição de probabilidade*:

- **Espaço amostral** Ω: Ω é o conjunto de todos os eventos elementares ou resultados possíveis de um experimento. Por exemplo, ao lançar um dado, $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- **Conjunto de eventos** $\mathcal{F}$: $\mathcal{F}$ é uma σ-álgebra, ou seja, um conjunto de subconjuntos de Ω que contém Ω e é fechado sob complementação e união enumerável (e, consequentemente, também sob interseção enumerável). Um exemplo de evento é: "o dado mostra um número ímpar".
- **Distribuição de probabilidade** $\mathbb{P}$: $\mathbb{P}$ é uma função que associa a cada evento de $\mathcal{F}$ um número em $[0, 1]$, tal que $\mathbb{P}[\Omega] = 1$, $\mathbb{P}[\emptyset] = 0$ e, para eventos mutuamente exclusivos $A_1, \ldots, A_n$, temos:

$$\mathbb{P}\left[A_1 \cup \cdots \cup A_n\right] = \sum_{i=1}^{n} \mathbb{P}[A_i].$$

A distribuição de probabilidade discreta associada ao lançamento de um dado justo pode ser definida como $\mathbb{P}[A_i] = 1/6$ para $i \in \{1, \ldots, 6\}$, onde A_i é o evento "o dado mostra o valor i".

2.1.1 Probabilidade condicional e independência

A probabilidade condicional do evento A dado o evento B é definida como a razão entre a probabilidade da interseção $A \cap B$ e a probabilidade de B, desde que $\mathbb{P}[B] \neq 0$:

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

Dois eventos A e B são ditos independentes quando a probabilidade conjunta $\mathbb{P}[A \cap B]$ pode ser fatorada como o produto $\mathbb{P}[A]\mathbb{P}[B]$:

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B].$$

De forma equivalente, a independência entre A e B pode ser expressa afirmando que $\mathbb{P}[A \mid B] = \mathbb{P}[A]$, sempre que $\mathbb{P}[B] \neq 0$.

Além disso, uma sequência de variáveis aleatórias é dita *i.i.d.* (independentes e identicamente distribuídas) quando todas as variáveis da sequência são mutuamente independentes e seguem a mesma distribuição de probabilidade.

Seguem algumas propriedades importantes:

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B] \qquad \text{(regra da soma)}$$

$$\mathbb{P}\left[\bigcup_{i=1}^{n} A_i\right] \leq \sum_{i=1}^{n} \mathbb{P}[A_i] \qquad \textbf{(desigualdade da união)}$$

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[B \mid A]\mathbb{P}[A]}{\mathbb{P}[B]} \qquad \text{(fórmula de Bayes)}$$

$$\mathbb{P}\left[\bigcap_{i=1}^{n} A_i\right] = \mathbb{P}[A_1]\mathbb{P}[A_2 \mid A_1] \cdots \mathbb{P}\left[A_n \mid \bigcap_{i=1}^{n-1} A_i\right] \qquad \text{(regra da cadeia)}$$

Exercício 1. *Prove os resultados acima.*

2.2 Variáveis aleatórias

Uma *variável aleatória* X é uma função mensurável $X : \Omega \to \mathbb{R}$, ou seja, tal que, para qualquer intervalo $I \subset \mathbb{R}$, o conjunto

$$\{\omega \in \Omega : X(\omega) \in I\}$$

pertence à σ-álgebra de eventos.

No caso discreto, a *função de massa de probabilidade* de X é dada por

$$x \mapsto \mathbb{P}[X = x].$$

Quando a distribuição de X é *absolutamente contínua*, existe uma *função densidade de probabilidade* f tal que, para todo $a, b \in \mathbb{R}$,

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

A função f é chamada *função densidade de probabilidade* da variável aleatória X. A relação entre a *função de distribuição acumulada* $F(\cdot)$ e a densidade $f(\cdot)$ é

$$F(a) = \mathbb{P}\{X \leq a\} = \int_{-\infty}^{a} f(x)\, dx.$$

Derivando ambos os lados, obtemos

$$\frac{d}{da}F(a) = f(a),$$

ou seja, a densidade é a derivada da função de distribuição acumulada.

Uma interpretação mais intuitiva de f pode ser obtida observando que, para $\varepsilon > 0$ pequeno,

$$\mathbb{P}\left(a - \frac{\varepsilon}{2} < X < a + \frac{\varepsilon}{2}\right) = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f(x)\, dx \approx \varepsilon f(a).$$

Assim, $f(a)$ quantifica a probabilidade de X assumir valores próximos de a.

Em muitos contextos, o interesse recai não apenas sobre variáveis aleatórias individuais, mas também sobre o relacionamento entre duas ou mais variáveis. Para descrever a dependência entre X e Y, definimos a *função de distribuição acumulada conjunta* como

$$F(x,y) = \mathbb{P}\{X \leq x,\, Y \leq y\},$$

que fornece a probabilidade de X ser menor ou igual a x e, simultaneamente, Y ser menor ou igual a y.

Se X e Y forem variáveis aleatórias discretas, a *função de massa de probabilidade conjunta* é

$$p(x,y) = \mathbb{P}\{X = x,\, Y = y\}.$$

Se forem *conjuntamente contínuas*, existe uma *função densidade de probabilidade conjunta* $f(x,y)$ tal que, para quaisquer conjuntos $C, D \subset \mathbb{R}$,

$$\mathbb{P}\{X \in C,\, Y \in D\} = \iint_{x \in C,\, y \in D} f(x,y)\, dx\, dy.$$

As variáveis X e Y são *independentes* se, para quaisquer $C, D \subset \mathbb{R}$,

$$\mathbb{P}\{X \in C,\, Y \in D\} = \mathbb{P}\{X \in C\}\, \mathbb{P}\{Y \in D\}.$$

De forma intuitiva, isso significa que conhecer o valor de uma das variáveis não altera a distribuição da outra.

No caso discreto, X e Y são independentes se, e somente se, para todo x, y,

$$\mathbb{P}\{X = x,\, Y = y\} = \mathbb{P}\{X = x\}\, \mathbb{P}\{Y = y\}.$$

Se forem conjuntamente contínuas, a independência é equivalente a

$$f(x,y) = f_X(x)\, f_Y(y), \quad \forall x, y,$$

onde f_X e f_Y são as densidades marginais de X e Y, respectivamente.

2.3 Valor esperado

A esperança ou valor esperado de uma variável aleatória X é denotada por $\mathbb{E}[X]$ e, no caso discreto, é definida como

$$\mathbb{E}[X] = \sum_x x\, \mathbb{P}[X = x].$$

Exemplo 1. *Se I é uma variável aleatória indicadora do evento A, isto é,*

$$I = \begin{cases} 1, & \text{se } A \text{ ocorre}, \\ 0, & \text{se } A \text{ não ocorre}, \end{cases}$$

então

$$\mathbb{E}[I] = 1 \cdot \mathbb{P}(A) + 0 \cdot \mathbb{P}(A^c) = \mathbb{P}(A).$$

Portanto, a esperança de uma variável indicadora de um evento A é exatamente a probabilidade de que A ocorra.

No caso contínuo, quando X possui uma função densidade de probabilidade $f(x)$, a esperança é dada por

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

Além disso, dado uma função qualquer g, temos que:

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

Uma propriedade fundamental da esperança é sua linearidade. Isto é, para quaisquer variáveis aleatórias X e Y e constantes $a, b \in \mathbb{R}$, temos:

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

2.4 Variância

A variância de uma variável aleatória X é denotada por $\mathrm{Var}[X]$ e definida como

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

O desvio padrão de X é denotado por σ_X e definido como

$$\sigma_X = \sqrt{\mathrm{Var}[X]}.$$

Para qualquer variável aleatória X e qualquer constante $a \in \mathbb{R}$, as seguintes propriedades básicas são válidas:

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2,$$

$$\mathrm{Var}[aX] = a^2\, \mathrm{Var}[X].$$

Além disso, se X e Y forem independentes, então

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y].$$

Exercício 2. *Prove os resultados anteriores.*

2.4.1 Covariância

A covariância entre duas variáveis aleatórias X e Y é denotada por $\mathrm{Cov}(X, Y)$ e definida por

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right].$$

Exercício 3. *Prove que*

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

Dizemos que X e Y são *não correlacionadas* quando $\mathrm{Cov}(X, Y) = 0$. Se X e Y forem independentes, então certamente são não correlacionadas, mas a recíproca nem sempre é verdadeira.

Exercício 4. *Seja X uniforme no intervalo $[-1, 1]$ e seja $Y = X^2$. Mostre que $\mathrm{Cov}\,(X, Y) = 0$ mas X, Y não são independentes.*

Observação 1. *Considere uma variável aleatória contínua X centrada em zero, ou seja, $\mathbb{E}[X] = 0$, com densidade de probabilidade par e definida em um intervalo do tipo $(-a, a)$, com $a > 0$. Seja $Y = g(X)$ para uma função g. A questão é: para quais funções $g(X)$ temos $\mathrm{Cov}(X, g(X)) = 0$?*

Sabemos que

$$\mathrm{Cov}(X, g(X)) = \mathbb{E}[Xg(X)] - \mathbb{E}[X]\mathbb{E}[g(X)].$$

Como $\mathbb{E}[X] = 0$, segue que $\mathrm{Cov}(X, g(X)) = \mathbb{E}[Xg(X)]$. Denotando a densidade de X por $f(x)$, temos

$$\mathrm{Cov}(X, g(X)) = \int_{-a}^{a} xg(x)f(x)dx.$$

Uma maneira de garantir que $\mathrm{Cov}(X, g(X)) = 0$ é exigir que $g(x)$ seja uma função par. Assim, $xg(x)f(x)$ será uma função ímpar e a integral em $(-a, a)$ se anulará, ou seja,

$$\int_{-a}^{a} xg(x)f(x)dx = 0.$$

Portanto, $\mathrm{Cov}(X, f(X)) = 0$ e como $Y = g(X)$, teremos que ambas são dependentes.

Dessa forma, podemos concluir que a distribuição precisa de X não afeta a condição, desde que $p(x)$ seja simétrica em torno da origem. Qualquer função par $f(\cdot)$ satisfará $\mathrm{Cov}(X, f(X)) = 0$.

A covariância é uma forma bilinear simétrica e semi-definida positiva, com as seguintes propriedades:

- **Simetria**: $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$ para quaisquer variáveis X e Y.
- **Bilinearidade**: $\mathrm{Cov}(X + X', Y) = \mathrm{Cov}(X, Y) + \mathrm{Cov}(X', Y)$ e $\mathrm{Cov}(aX, Y) = a\,\mathrm{Cov}(X, Y)$ para qualquer $a \in \mathbb{R}$.
- **Semi-definida positiva**: $\mathrm{Cov}(X, X) = \mathrm{Var}[X] \geq 0$ para qualquer variável X.

Além disso, vale a desigualdade de Cauchy-Schwarz, que afirma que para variáveis X e Y com variância finita,

$$|\mathrm{Cov}(X, Y)| \leq \sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}.$$

Perceba a semelhança do resultado acima com a desigualdade de Cauchy-Schwarz!

Exercício 5. *Prove os resultados acima.*

A matriz de covariância de um vetor de variáveis aleatórias $\mathbf{X} = (X_1, \ldots, X_p)$ é a matriz em $\mathbb{R}^{n\times n}$ denotada por $\mathbf{C}(\mathbf{X})$ e definida por

$$\mathbf{C}(\mathbf{X}) = \mathbb{E}\left[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^\top\right].$$

Portanto, $\mathbf{C}(\mathbf{X})$ é a matriz cujos elementos são $\mathrm{Cov}(X_i, X_j)$. Além disso, é imediato mostrar que

$$\mathbf{C}(\mathbf{X}) = \mathbb{E}[\mathbf{X}\mathbf{X}^\top] - \mathbb{E}[\mathbf{X}]\,\mathbb{E}[\mathbf{X}]^\top.$$

2.5 Desigualdades básicas de concentração

Nesta seção, apresentamos duas desigualdades fundamentais que estabelecem limites superiores para a probabilidade de uma variável aleatória assumir valores distantes de sua média. Tais resultados são amplamente utilizados em probabilidade, estatística e teoria da informação para analisar o comportamento de caudas de distribuições.

A primeira delas é a *Desigualdade de Markov*, que fornece um limite simples para variáveis aleatórias não-negativas em função apenas de sua esperança.

Teorema 1 (Desigualdade de Markov). *Seja X uma variável aleatória não-negativa ($X \geq 0$ quase certamente) com valor esperado $\mathbb{E}[X] < \infty$. Então, para todo $t > 0$, temos:*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

Exercício 6. *Prove a desigualdade de Markov.* ***Dica:*** *use o fato de que $\frac{x}{t} \geq \mathbb{I}\{x \geq t\}$.*

A próxima desigualdade é um refinamento da anterior. Conhecida como *Desigualdade de Chebyshev*, ela aplica a desigualdade de Markov à variável aleatória $(X - \mu)^2$ e relaciona o desvio da média com a variância da distribuição.

Teorema 2 (Desigualdade de Chebyshev). *Seja X uma variável aleatória com valor esperado $\mu = \mathbb{E}[X]$ e variância finita $\mathrm{Var}(X) = \sigma^2$. Então, para todo $\varepsilon > 0$, vale:*

$$\mathbb{P}(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Exercício 7. *Prove a desigualdade de Chebyshev a partir da desigualdade de Markov aplicada a $(X - \mu)^2$.*

2.6 Teoremas assintóticos

Em muitas aplicações de probabilidade e estatística, estamos interessados no comportamento de sequências de variáveis aleatórias quando o número de observações tende ao infinito. Os *teoremas assintóticos* fornecem resultados fundamentais que descrevem como certos estimadores ou somas de variáveis aleatórias se comportam no limite, ou seja, quando o tamanho da amostra n cresce indefinidamente.

Teorema 3 (Lei Fraca dos Grandes Números). *Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias independentes, todas com a mesma esperança μ e variância $\sigma^2 < \infty$. Definindo a média amostral por*

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

então, para qualquer $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\overline{X}_n - \mu\right| \geq \varepsilon\right) = 0.$$

Exercício 8. *Prove a Lei Fraca dos Grandes números utilizando a desigualdade de Chebyshev.*

Teorema 4 (Teorema Central do Limite)**.** *Seja $X_1, \dots, X_n$ uma sequência de variáveis aleatórias i.i.d. com esperança μ e desvio padrão σ. Definimos a média amostral como*

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

e a variância da média como $\sigma_n^2 = \sigma^2/n$. Então, a variável padronizada $(\overline{X}_n - \mu)/\sigma_n$ converge em distribuição para uma normal padrão $N(0,1)$. Mais precisamente, para todo $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\overline{X}_n - \mu}{\sigma_n} \le t\right) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Demonstração. □

2.6.1 Função geradora de momentos

A esperança $\mathbb{E}[X^p]$ é chamada de p-ésimo momento da variável aleatória X. A *função geradora de momentos* de uma variável aleatória X é uma ferramenta importante, pois permite obter seus diferentes momentos por meio de diferenciação em zero. Essa função é crucial tanto para descrever a distribuição de X quanto para analisar suas propriedades.

A função geradora de momentos de uma variável aleatória X é a função $M_X : t \mapsto \mathbb{E}[e^{tX}]$, definida para os valores de $t \in \mathbb{R}$ tais que a expectativa exista (seja finita).

Exercício 9. *Mostre que se M_X for diferenciável em zero, então o p-ésimo momento de X é dado por $\mathbb{E}[X^p] = M_X^{(p)}(0)$.*

Exercício 10. *Seja X uma variável aleatória com distribuição normal padrão, ou seja, $X \sim N(0,1)$. Mostre que a função geradora de momentos de X é dada por por:*

$$M_X(t) = e^{\frac{t^2}{2}}.$$

Capítulo 3

Geração de variáveis aleatórias

Capítulo 4

Redução de variância

Capítulo 5

Inferência

Capítulo 6

Bootstrap

Capítulo 7

Cadeias de Markov

Referências Bibliográficas