

Métodos Computacionais

Thiago Rodrigo Ramos

17 de agosto de 2025

Sumário

1	Introdução	5
1.1	Um conselho: a importância de ser ruim antes de ser bom	5
2	Elementos básicos de probabilidade	7
2.1	Axiomas da probabilidade	7
2.1.1	Probabilidade condicional e independência	7
2.2	Variáveis aleatórias	8
2.3	Valor esperado	9
2.4	Variância	10
2.4.1	Covariância	10
2.5	Desigualdades básicas de concentração	12
2.6	Teoremas assintóticos	12
3	Variáveis discretas e como simulá-las	15
3.1	Bernoulli	15
3.2	Distribuição Binomial	16
3.3	Distribuição Geométrica	16
3.4	Distribuição de Poisson	17
3.5	Distribuição Binomial Negativa	17
3.6	Distribuição Hipergeométrica	18
4	Simulação de variáveis discretas via inversão	19
4.1	Variáveis com suporte finito	21
4.2	Binomial	22
4.2.1	Simulando via Bernoullis	22
4.2.2	Simulando via identidade recursiva	22
4.2.3	Aspectos computacionais	24
5	Variáveis Contínuas e suas distribuições	27
5.0.1	Uniformes	27
5.0.2	Exponencial	27
5.0.3	Beta	27
5.0.4	Gamma	27
5.0.5	Normal	27

6	Geração de variáveis aleatórias	29
7	Redução de variância	31
8	Inferência	33
9	Bootstrap	35
10	Cadeias de Markov	37

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, \dots, A_n , temos:

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

A distribuição de probabilidade discreta associada ao lançamento de um dado justo pode ser definida como $\mathbb{P}[A_i] = 1/6$ para $i \in \{1, \dots, 6\}$, 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:

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

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

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

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

Exercício 1. Prove os resultados acima.

2.2 Variáveis aleatórias

Uma *variável aleatória* X é uma função mensurável $X : \Omega \rightarrow \mathbb{R}$, 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 $\text{Var}[X]$ e definida como

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

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

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

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

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

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

Além disso, se X e Y forem independentes, então

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

Exercício 2. Prove os resultados anteriores.

2.4.1 Covariância

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

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

Exercício 3. Prove que

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

Dizemos que X e Y são *não correlacionadas* quando $\text{Cov}(X, Y) = 0$. 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 $\text{Cov}(X, Y) = 0$ 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 $\text{Cov}(X, g(X)) = 0$?

Sabemos que

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

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

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

Uma maneira de garantir que $\text{Cov}(X, g(X)) = 0$ 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, $\text{Cov}(X, f(X)) = 0$ 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 $\text{Cov}(X, f(X)) = 0$.

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

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

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

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

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 $\mathbf{X} = (X_1, \dots, X_p)$ 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 $\text{Cov}(X_i, X_j)$. 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 $\text{Var}(X) = \sigma^2$. 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*

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

então, para qualquer $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) = 0.$$

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 variância σ^2 . Definimos a média amostral como*

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

Então,

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \rightarrow N(0, 1).$$

Demonstração. Suponha, sem perda de generalidade, que $\mu = 0$ e $\sigma = 1$. Para provar que uma sequência de variáveis aleatórias Z_n converge em distribuição para uma variável aleatória Z , basta mostrar que, para toda função de teste suficientemente regular f , vale

$$\mathbb{E}[f(Z_n)] \rightarrow \mathbb{E}[f(Z)].$$

Defina

$$A_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \quad \text{e} \quad B_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n N_i,$$

onde $N_i \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ independentes. Note que $B_n \sim N(0, 1)$ para todo n . Assim, o que precisamos mostrar é que, para qualquer f regular,

$$\mathbb{E}[f(A_n)] - \mathbb{E}[f(B_n)] \rightarrow 0.$$

O truque é considerar as seguintes variáveis:

$$\begin{aligned} C_n^{(0)} &= \frac{1}{\sqrt{n}} (X_1 + X_2 + X_3 + \dots + X_{n+1} + X_n), \\ C_n^{(1)} &= \frac{1}{\sqrt{n}} (N_1 + X_2 + X_3 + \dots + X_{n+1} + X_n), \\ C_n^{(2)} &= \frac{1}{\sqrt{n}} (N_1 + N_2 + X_3 + \dots + X_{n+1} + X_n), \\ &\vdots \\ C_n^{(n-1)} &= \frac{1}{\sqrt{n}} (N_1 + N_2 + N_3 + \dots + N_{n-1} + X_n), \\ C_n^{(n)} &= \frac{1}{\sqrt{n}} (N_1 + N_2 + N_3 + \dots + N_{n-1} + N_n), \end{aligned}$$

e escrever a diferença de esperanças como soma telescópica:

$$\begin{aligned} \mathbb{E}[f(A_n)] - \mathbb{E}[f(B_n)] &= \mathbb{E}[f(C_n^{(n)})] - \mathbb{E}[f(C_n^{(0)})] \\ &= \sum_{k=1}^n \Delta_k, \end{aligned}$$

onde,

$$\Delta_k := \mathbb{E} \left[f(C_n^{(k)}) - f(C_n^{(k-1)}) \right].$$

Note que, entre $C_n^{(j)}$ e $C_n^{(j-1)}$, o único termo que difere é o j -ésimo.

Definimos então

$$D_n^{(j)} = \frac{1}{\sqrt{n}} (N_1 + \cdots + N_{j-1} + 0 + X_{j+1} + \cdots + X_n),$$

isto é, a variável que coincide com $C_n^{(j)}$ exceto que o j -ésimo termo foi substituído por zero. Dessa forma, pela expansão de Taylor:

$$\begin{aligned} \mathbb{E} \left[f \left(C_n^{(k)} \right) - f \left(C_n^{(k-1)} \right) \right] &= \mathbb{E} \left[f \left(D_n^{(k)} + \frac{N_k}{\sqrt{n}} \right) - f \left(D_n^{(k)} + \frac{X_k}{\sqrt{n}} \right) \right] \\ &= \frac{1}{\sqrt{n}} \mathbb{E} \left[f' \left(D_n^{(k)} \right) (N_k - X_k) \right] \\ &\quad + \frac{1}{2n} \mathbb{E} \left[f'' \left(D_n^{(k)} \right) (N_k^2 - X_k^2) \right] \\ &\quad + \mathbb{E}[R_k], \end{aligned}$$

Agora note que

$$\begin{aligned} \mathbb{E} \left[f' \left(D_n^{(k)} \right) (N_k - X_k) \right] &= \mathbb{E} \left[\mathbb{E} \left[f' \left(D_n^{(k)} \right) (N_k - X_k) \mid D_n^{(k)} \right] \right] \\ &= \mathbb{E} \left[f' \left(D_n^{(k)} \right) \mathbb{E} [N_k - X_k \mid D_n^{(k)}] \right] \\ &= \mathbb{E} \left[f' \left(D_n^{(k)} \right) (\mathbb{E}[N_k] - \mathbb{E}[X_k]) \right] \\ &= 0, \end{aligned}$$

E o mesmo vale para o termo de segunda ordem, portanto só nos resta analisar o que acontece com o resto R_k . □

Capítulo 3

Variáveis discretas e como simulá-las

Nesta seção, apresentaremos as principais propriedades de algumas distribuições de probabilidade amplamente utilizadas, que servirão de base para o estudo e a implementação de métodos de simulação nas seções posteriores.

Iniciaremos com o estudo de algumas distribuições discretas, isto é, distribuições de probabilidade cuja variável aleatória associada assume apenas valores em um conjunto enumerável (finito ou infinito).

Nesse caso, a distribuição é completamente determinada pela função de probabilidade

$$p_X(x_k) = \mathbb{P}(X = x_k), \quad x_k \in S,$$

onde S é um conjunto de valores possíveis da variável aleatória X .

Essas probabilidades satisfazem

$$p_X(k) \geq 0 \quad \text{para todo } k \in S, \quad \sum_{k \in S} p_X(k) = 1.$$

Exemplo 2. Seja $S = \{x_1, x_2, \dots, x_K\}$ um conjunto de K valores distintos. Dizemos que X tem distribuição uniforme discreta em S quando

$$p_X(x_i) = \frac{1}{K}, \quad i = 1, 2, \dots, K.$$

Nesse caso, cada valor é igualmente provável e temos

$$\sum_{i=1}^K p_X(x_i) = \sum_{i=1}^K \frac{1}{K} = 1.$$

3.1 Bernoulli

A distribuição de Bernoulli modela experimentos com dois resultados possíveis, tipicamente denominados “sucesso” (valor 1) e “fracasso” (valor 0). Dizemos que $X \sim \text{Bernoulli}(p)$ se

$$\mathbb{P}(X = 1) = p \quad \text{e} \quad \mathbb{P}(X = 0) = 1 - p,$$

onde $0 \leq p \leq 1$ representa a probabilidade de sucesso.

A função de probabilidade (pmf) pode ser escrita de forma compacta como

$$p_X(k) = p^k(1-p)^{1-k}, \quad k \in \{0, 1\}.$$

As principais características dessa distribuição são:

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

Exercício 9. *Prove as propriedades acima.*

3.2 Distribuição Binomial

A distribuição binomial modela o número de sucessos em n repetições independentes de um experimento de Bernoulli com probabilidade de sucesso p .

Sejam X_1, X_2, \dots, X_n variáveis aleatórias independentes, todas com distribuição Bernoulli(p). Definimos

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

Nesse caso, dizemos que $X \sim \text{Binomial}(n, p)$, cuja função de probabilidade é

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

As principais propriedades são:

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

Exercício 10. *Prove as propriedades acima.*

3.3 Distribuição Geométrica

A distribuição geométrica modela o número de ensaios de Bernoulli até a ocorrência do primeiro sucesso. Seja p a probabilidade de sucesso em cada tentativa, com $0 < p \leq 1$. Definimos X como o número de ensaios necessários até o primeiro sucesso. Dizemos que $X \sim \text{Geom}(p)$ se

$$\mathbb{P}(X = k) = (1-p)^{k-1}p, \quad k = 1, 2, 3, \dots$$

Nesse caso:

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

Exercício 11. *Prove que a função de probabilidade acima satisfaz $\sum_{k=1}^{\infty} \mathbb{P}(X = k) = 1$.*

Exercício 12. *Prove as propriedades acima.*

3.4 Distribuição de Poisson

A distribuição de Poisson modela o número de ocorrências de um evento em um intervalo fixo de tempo ou espaço, assumindo que tais ocorrências sejam raras e independentes.

Dizemos que $X \sim \text{Poisson}(\lambda)$ se sua função de probabilidade for

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

onde $\lambda > 0$ representa a taxa média de ocorrências no intervalo considerado.

A média e a variância são dadas por

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

Exercício 13. Prove que $\sum_{k=0}^{\infty} \mathbb{P}(X = k) = 1$.

Exercício 14. Prove as propriedades acima. Dica: para a variância, calcule primeiro $\mathbb{E}[X(X-1)]$ e use o fato de que

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

3.5 Distribuição Binomial Negativa

Seja X o número de ensaios necessários para obter um total de r sucessos, considerando que cada ensaio é independente e resulta em sucesso com probabilidade p . Nesse caso, dizemos que X segue uma distribuição *binomial negativa* (também chamada *Pascal*) com parâmetros p e r .

Sua função de probabilidade é dada por:

$$\mathbb{P}(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n = r, r+1, r+2, \dots$$

Essa fórmula é justificada pelo fato de que, para que sejam necessários exatamente n ensaios para obter r sucessos, os primeiros $n-1$ ensaios devem conter exatamente $r-1$ sucessos — o que ocorre com probabilidade

$$\binom{n-1}{r-1} p^{r-1} (1-p)^{n-r}$$

— e, em seguida, o n -ésimo ensaio deve ser um sucesso, com probabilidade p .

Seja X_i , $i = 1, \dots, r$, o número de ensaios necessários após o $(i-1)$ -ésimo sucesso para obter o i -ésimo sucesso. É fácil ver que X_1, X_2, \dots, X_r são variáveis aleatórias independentes com distribuição $\text{Geom}(p)$. Assim, como

$$X = X_1 + X_2 + \dots + X_r,$$

temos, usando os resultados da distribuição geométrica:

$$\mathbb{E}[X] = \sum_{i=1}^r \mathbb{E}[X_i] = \frac{r}{p}, \quad \text{Var}(X) = \sum_{i=1}^r \text{Var}(X_i) = \frac{r(1-p)}{p^2}.$$

Exercício 15. Prove que a função de probabilidade acima é válida, isto é, que $\sum_{n=r}^{\infty} \mathbb{P}(X = n) = 1$.

Exercício 16. Prove as fórmulas da média e variância usando o fato de que X é a soma de r variáveis independentes com distribuição $\text{Geom}(p)$.

3.6 Distribuição Hipergeométrica

A distribuição hipergeométrica modela experimentos de seleção *sem reposição* a partir de uma população finita contendo dois tipos de elementos. Por exemplo, suponha uma urna com $N + M$ bolas, das quais N são claras e M são escuras. Retiramos, de forma aleatória e sem reposição, uma amostra de tamanho n . Seja X o número de bolas claras na amostra.

Nesse caso, cada subconjunto de tamanho n é igualmente provável, e a probabilidade de observar exatamente k bolas claras é

$$\mathbb{P}(X = k) = \frac{\binom{N}{k} \binom{M}{n-k}}{\binom{N+M}{n}}, \quad \max(0, n-M) \leq k \leq \min(n, N).$$

Dizemos então que $X \sim \text{Hipergeom}(N, M, n)$.

As principais propriedades dessa distribuição são:

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

Exercício 17. Prove que a função de probabilidade acima é válida, isto é, que

$$\sum_{k=\max(0, n-M)}^{\min(n, N)} \mathbb{P}(X = k) = 1.$$

Exercício 18. Prove as fórmulas da média e variância acima. Dica: considere o sorteio sequencial das n bolas e defina X_i como a variável indicadora do evento “a i -ésima bola é clara”. Para a variância, use a decomposição

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

Capítulo 4

Simulação de variáveis discretas via inversão

O truque fundamental para simular variáveis aleatórias a partir de uma variável uniforme $U \sim \text{Uniforme}(0, 1)$ é a seguinte propriedade:

$$\mathbb{P}(a < U < b) = b - a, \quad 0 \leq a < b \leq 1.$$

Isto é, a probabilidade de U cair em um subintervalo do intervalo $(0, 1)$ é igual ao comprimento desse subintervalo.

Para variáveis discretas, essa ideia pode ser usada da seguinte forma: suponha que X assumam valores x_1, x_2, \dots, x_m com probabilidades p_1, p_2, \dots, p_m , onde

$$p_k = \mathbb{P}(X = x_k), \quad p_k \geq 0, \quad \sum_{k=1}^m p_k = 1.$$

Definimos as probabilidades acumuladas

$$F_k = \sum_{i=1}^k p_i, \quad k = 1, \dots, m.$$

Então, o algoritmo de simulação é:

1. Gerar $U \sim \text{Uniforme}(0, 1)$;
2. Encontrar o menor índice k tal que $U \leq F_k$;
3. Retornar $X = x_k$.

A propriedade $\mathbb{P}(a < U < b) = b - a$ garante que

$$\mathbb{P}(X = x_k) = p_k.$$

De forma intuitiva, dividimos o intervalo $(0, 1)$ em subintervalos consecutivos de comprimentos p_k . Ao sortearmos $U \sim \text{Uniforme}(0, 1)$, o valor de X será aquele correspondente ao subintervalo no qual U cair. Esse procedimento é conhecido como *método da inversão* para variáveis discretas.

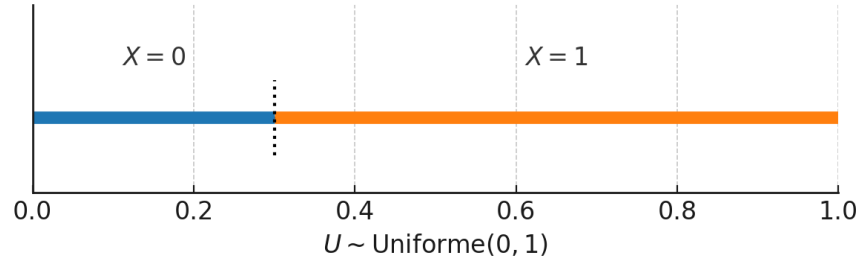


Figura 4.1: Particionamento do intervalo $(0,1)$ para simular uma variável Bernoulli com $p = 0,3$. Sorteia-se $U \sim \text{Uniforme}(0,1)$; se U cair na região azul, definimos $X = 0$, e caso contrário, $X = 1$.

A Figura 4.1 ilustra esse processo para uma variável Bernoulli.

A mesma ideia se aplica quando o conjunto de valores possíveis de X é infinito (ou muito grande). Nesse caso, o intervalo $(0,1)$ é particionado em uma sequência de subintervalos, cada um correspondente a um valor de X , como ilustrado na Figura 4.2.

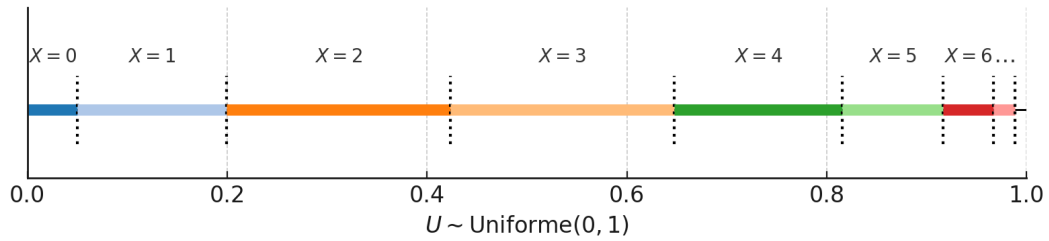


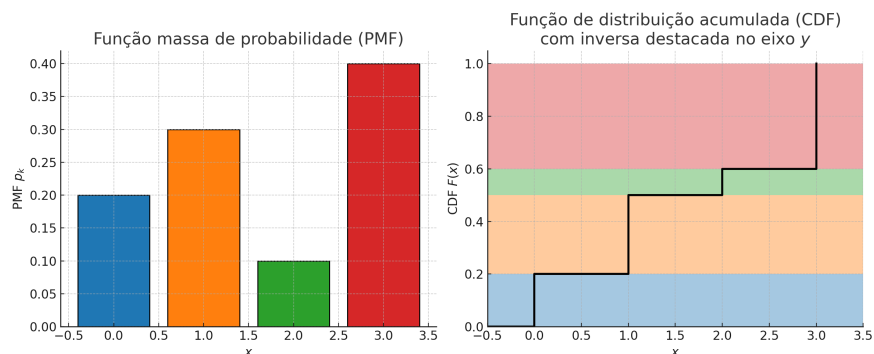
Figura 4.2: Particionamento do intervalo $(0,1)$ para simular uma variável discreta com suporte infinito.

O nome *método da inversão* vem do fato de que a simulação utiliza a *função de distribuição acumulada* (CDF) e sua *inversa generalizada*. Seja X uma variável aleatória com CDF $F(x)$. Então, se $U \sim \text{Uniforme}(0,1)$, vale que

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

onde a inversa generalizada é definida por

$$F^{-1}(u) = \min\{x : F(x) \geq u\}, \quad 0 < u < 1.$$



No caso discreto, isto corresponde exatamente ao passo do algoritmo em que escolhemos o menor k tal que $U \leq F_k$. Ou seja, sorteamos U , e depois “invertemos” a CDF para recuperar uma realização de X na sua escala original.

Esse procedimento pode parecer um pouco abstrato neste momento, já que a noção de inversa de uma função acumulada fica mais clara quando lidamos com variáveis contínuas. Por isso, retornaremos a esse método mais adiante, ao estudarmos a simulação de variáveis contínuas via inversão.

4.1 Variáveis com suporte finito

Considere o caso em que X assume um número finito de valores x_1, x_2, \dots, x_m , cada um com probabilidade $p_j = \mathbb{P}(X = x_j)$.

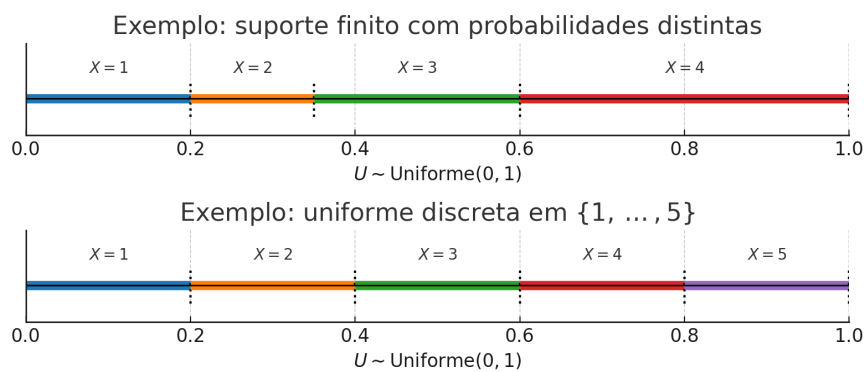
Por exemplo, suponha que

$$p_1 = 0.20, \quad p_2 = 0.15, \quad p_3 = 0.25, \quad p_4 = 0.40.$$

Uma maneira direta de simular X é gerar $U \sim \text{Uniforme}(0, 1)$ e aplicar:

- Se $U < 0.20$, definir $X = 1$ e pare;
- Se $U < 0.35$, definir $X = 2$ e pare;
- Se $U < 0.60$, definir $X = 3$ e pare;
- Caso contrário, definir $X = 4$.

Embora possamos reordenar os testes para tornar a verificação mais eficiente, a ideia central permanece a mesma: dividir o intervalo $(0, 1)$ em partes de comprimentos p_j e identificar onde U caiu.



Um caso especial é o da *uniforme discreta* nos valores $1, 2, \dots, n$, em que

$$\mathbb{P}(X = j) = \frac{1}{n}, \quad j = 1, \dots, n.$$

Neste cenário, o método se torna extremamente simples: basta gerar $U \sim \text{Uniforme}(0, 1)$ e definir

$$X = \lfloor nU \rfloor + 1,$$

onde $\lfloor x \rfloor$ indica a parte inteira de x (maior inteiro menor ou igual a x).

De fato, $X = j$ se e somente se $j - 1 \leq nU < j$, o que ocorre com probabilidade $\frac{1}{n}$. Variáveis uniformes discretas são particularmente importantes em simulação, pois permitem gerar inteiros equiprováveis de forma extremamente eficiente.

Um caso ainda mais simples é a simulação de uma variável **Bernoulli** com parâmetro p , que é equivalente a uma uniforme discreta em $\{0, 1\}$ com probabilidades $1 - p$ e p , respectivamente.

4.2 Binomial

4.2.1 Simulando via Bernoullis

Uma forma simples e direta de simular uma variável aleatória binomial é a partir de variáveis de Bernoulli independentes.

Recorde que se $X \sim \text{Binomial}(n, p)$, então X pode ser escrito como

$$X = \sum_{i=1}^n B_i,$$

onde B_1, B_2, \dots, B_n são variáveis independentes e identicamente distribuídas, cada uma com

$$B_i \sim \text{Bernoulli}(p).$$

Assim, o algoritmo de simulação da binomial segue naturalmente:

1. Para $i = 1, \dots, n$, gerar $B_i \sim \text{Bernoulli}(p)$;
2. Retornar $X = \sum_{i=1}^n B_i$.

Em outras palavras, uma variável binomial conta o número de sucessos em n tentativas independentes, cada uma com probabilidade de sucesso p . Portanto, simular uma binomial se reduz a repetir n vezes o procedimento de simulação da Bernoulli e somar os resultados.

4.2.2 Simulando via identidade recursiva

Uma alternativa mais eficiente utiliza o *método da inversão*, aproveitando a identidade recursiva da função massa de probabilidade da Binomial.

Se $X \sim \text{Binomial}(n, p)$, então

$$\mathbb{P}(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}, \quad i = 0, 1, \dots, n.$$

Essas probabilidades satisfazem uma relação de recorrência simples. De fato, começando em

$$\mathbb{P}(X = i + 1) = \binom{n}{i+1} p^{i+1} (1 - p)^{n-i-1},$$

observamos que ¹

¹Por exemplo, se $n = 10$, $i = 6$ e $i + 1 = 7$, então

$$\frac{10!}{7!3!} = \frac{10! \cdot 4}{7 \cdot 6! \cdot 4 \cdot 3!} = \frac{10!}{6! \cdot 4!} \frac{4}{7}$$

$$\binom{n}{i+1} = \frac{n!}{(i+1)!(n-i-1)!} = \frac{n-i}{i+1} \frac{n!}{i!(n-i)!} = \frac{n-i}{i+1} \binom{n}{i}.$$

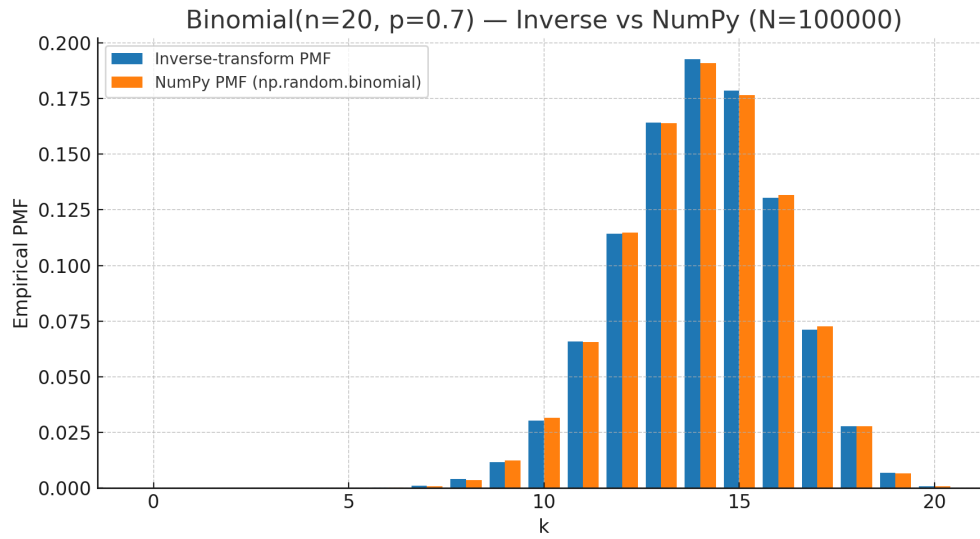
Substituindo essa relação,

$$\mathbb{P}(X = i+1) = \frac{n-i}{i+1} \binom{n}{i} p^{i+1} (1-p)^{n-i-1}.$$

Reorganizando,

$$\mathbb{P}(X = i+1) = \frac{n-i}{i+1} \cdot \frac{p}{1-p} \mathbb{P}(X = i).$$

Assim, conhecendo $\mathbb{P}(X = 0) = (1-p)^n$, podemos calcular $\mathbb{P}(X = 1), \mathbb{P}(X = 2), \dots$ de forma recursiva, sem reavaliar coeficientes binomiais nem potências.



Isso leva ao seguinte algoritmo de simulação via inversão:

1. Gerar $U \sim \text{Uniforme}(0,1)$;
2. Inicializar o índice $i = 0$, a probabilidade atual $p_i = (1-p)^n$ e a soma acumulada $F = p_i$;
3. Enquanto $U > F$, atualizar

$$p_{i+1} = \frac{n-i}{i+1} \cdot \frac{p}{1-p} p_i, \quad i \leftarrow i+1, \quad F \leftarrow F + p_i;$$

4. Retornar $X = i$.

Esse procedimento verifica primeiro se $X = 0$, depois se $X = 1$, e assim por diante, até encontrar o valor de X sorteado. Em média, o número de passos necessários é aproximadamente $1 + np$, o que pode ser bem mais eficiente do que gerar n variáveis de Bernoulli quando n é grande.

Exemplo 3. Considere $n = 5$ e $p = 0.3$. Temos $\mathbb{P}(X = 0) = (1-0.3)^5 = 0.16807$. Suponha que geramos $U = 0.4$. Como $U > 0.16807$, passamos ao próximo valor:

$$p_1 = \frac{5-0}{1} \cdot \frac{0.3}{0.7} \cdot 0.16807 \approx 0.36015, \quad F = 0.16807 + 0.36015 = 0.52822.$$

Agora $U = 0.4 < F$, logo o algoritmo retorna $X = 1$.

Portanto, neste caso específico, o sorteio resultou em exatamente um sucesso entre as cinco tentativas.

4.2.3 Aspectos computacionais

A escolha do método para simular variáveis binomiais tem implicações diretas em termos de eficiência. Dois fatores fundamentais influenciam o desempenho: o número de tentativas n e a probabilidade de sucesso p .

No método da soma de Bernoullis, o custo de cada amostra é proporcional a n , já que é necessário realizar n sorteios independentes. Esse custo não depende do valor de p : tanto para valores pequenos quanto grandes de p , o algoritmo precisa sempre gerar todas as n Bernoullis.

Já no método da inversão recursiva, o número médio de passos é da ordem de $1 + np$, pois o procedimento acumula probabilidades até ultrapassar o valor sorteado U . Quando p é pequeno, o valor típico da variável X também é pequeno, e o algoritmo tende a parar cedo, podendo ser competitivo em relação à soma de Bernoullis. Por outro lado, quando p é moderado ou grande, o valor esperado np cresce e, com ele, o número de passos, tornando a inversão significativamente mais lenta.

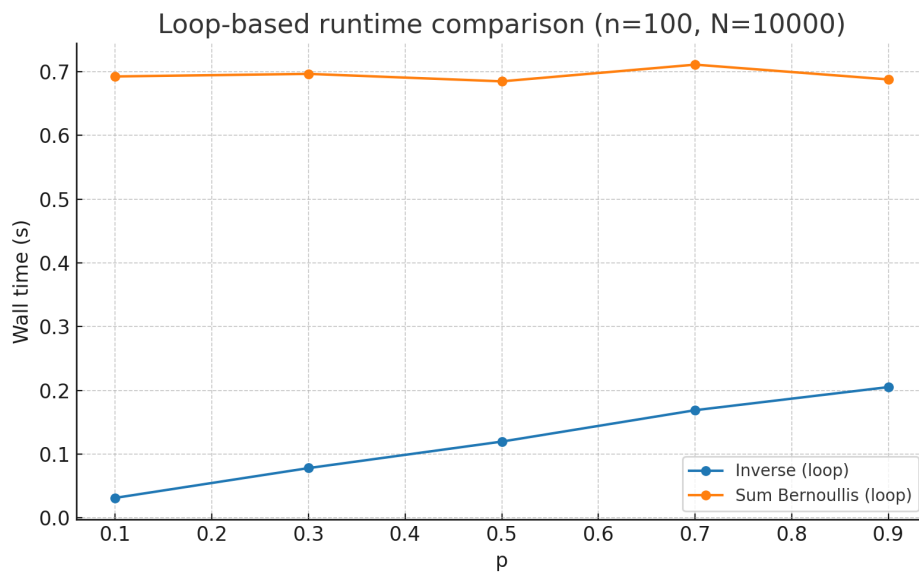


Figura 4.3: Comparação de tempo de execução (em segundos) entre o método da inversão recursiva e a soma de Bernoullis para $n = 100$ e $N = 10,000$ amostras, variando p .

A Figura 4.3 ilustra essa comparação em implementações com *loops* explícitos, para $n = 100$ e diferentes valores de p . Enquanto o tempo da soma de Bernoullis cresce linearmente apenas com n e não é afetado por p , o tempo do método da inversão cresce proporcionalmente a np , aumentando de forma acentuada à medida que p se aproxima de 1. Na prática, bibliotecas como NumPy utilizam algoritmos especializados para a binomial, ainda mais rápidos do que ambos os métodos discutidos aqui, de modo que a utilidade principal desses algoritmos é didática e comparativa, permitindo compreender os diferentes custos computacionais associados a cada abordagem.

Tabela de referência

Distribuição	Técnica utilizada	Dica/Obs
Suporte finito	Inversão simples	Separar em intervalos
Bernoulli	Inversão simples	Caso particular do suporte finito com $m = 2$
Binomial	Soma de Bernoullis ou inversão recursiva	Relação de recorrência evita coeficientes binomiais
Poisson		
Geométrica		
Binomial negativa		
Hipergeométrica		
Uniforme discreta		

Capítulo 5

Variáveis Contínuas e suas distribuições

5.0.1 Uniformes

5.0.2 Exponencial

5.0.3 Beta

5.0.4 Gamma

5.0.5 Normal

Capítulo 6

Geração de variáveis aleatórias

Capítulo 7

Redução de variância

Capítulo 8

Inferência

Capítulo 9

Bootstrap

Capítulo 10

Cadeias de Markov

Referências Bibliográficas