UnB | Departamento de Matemática

# Momentos e Funções Geradoras de Momentos

**L. Cioletti**

Outubro de 2026

**Resumo**

Nestas notas, estudamos os momentos de uma variável aleatória e as funções geradoras de momentos. Começamos retomando a expressão da esperança como integral da cauda, que permite tratar distribuições gerais por meio de integrais impróprias de Riemann. Nos cálculos, usamos principalmente densidades e funções de probabilidade. Definimos os momentos $\mathbb{E}[X^k]$, calculamos os momentos da distribuição exponencial, observamos que a distribuição de Cauchy não possui momentos e discutimos em que sentido a sequência de momentos determina a distribuição, com a distribuição log-normal como contraexemplo. Em seguida, estudamos a variância, a covariância e o coeficiente de correlação $\rho(X, Y)$, mostramos que $|\rho(X, Y)| \leqslant 1$ a partir da desigualdade de Cauchy–Schwarz e caracterizamos os casos $\rho(X, Y) = \pm 1$. Por fim, introduzimos a função geradora de momentos $M_X(t) = \mathbb{E}[e^{tX}]$, mostramos que os momentos podem ser recuperados a partir das derivadas de $M_X$ na origem, que $M_{aX+b}(t) = e^{tb} M_X(at)$ e que a geradora da soma de variáveis aleatórias independentes é o produto das geradoras. Com o teorema de unicidade, mostramos que a soma de duas variáveis normais independentes é normal. Encerramos com uma aplicação da geradora da uniforme discreta: uma única expansão de Taylor produz a fórmula de Faulhaber para as somas de potências, com os números de Bernoulli como coeficientes. A exposição segue de perto as Seções 7.1 a 7.4 do Capítulo 7 da referência [1], com exceção desta última seção, que é uma aplicação adicional.

## 1. Introdução

Nas notas anteriores, estudamos as variáveis aleatórias discretas e, em notas separadas, as absolutamente contínuas. No caso discreto, descrevemos a distribuição por uma função de probabilidade $p_X$ e calculamos esperanças por meio de somas; no caso absolutamente contínuo, descrevemos a distribuição por uma densidade $f_X$ e calculamos as mesmas grandezas por meio de integrais. Essa separação mantém cada dedução próxima das ferramentas disponíveis no momento, mas as duas classes não esgotam as possibilidades. A distribuição de Cantor, construída nas notas sobre funções de distribuição, por exemplo, não é discreta nem admite densidade: sua função de distribuição é contínua, mas a distribuição é singular. Também há distribuições que combinam uma parte discreta e uma parte absolutamente contínua. Para estudar momentos sem restringir o tipo da variável, precisamos de uma noção de

CC BY-NC-SA 4.0- Attribution-NonCommercial-ShareAlike 4.0 International

esperança que inclua esses casos. A Figura 1 situa as duas classes já estudadas nesse quadro mais geral.

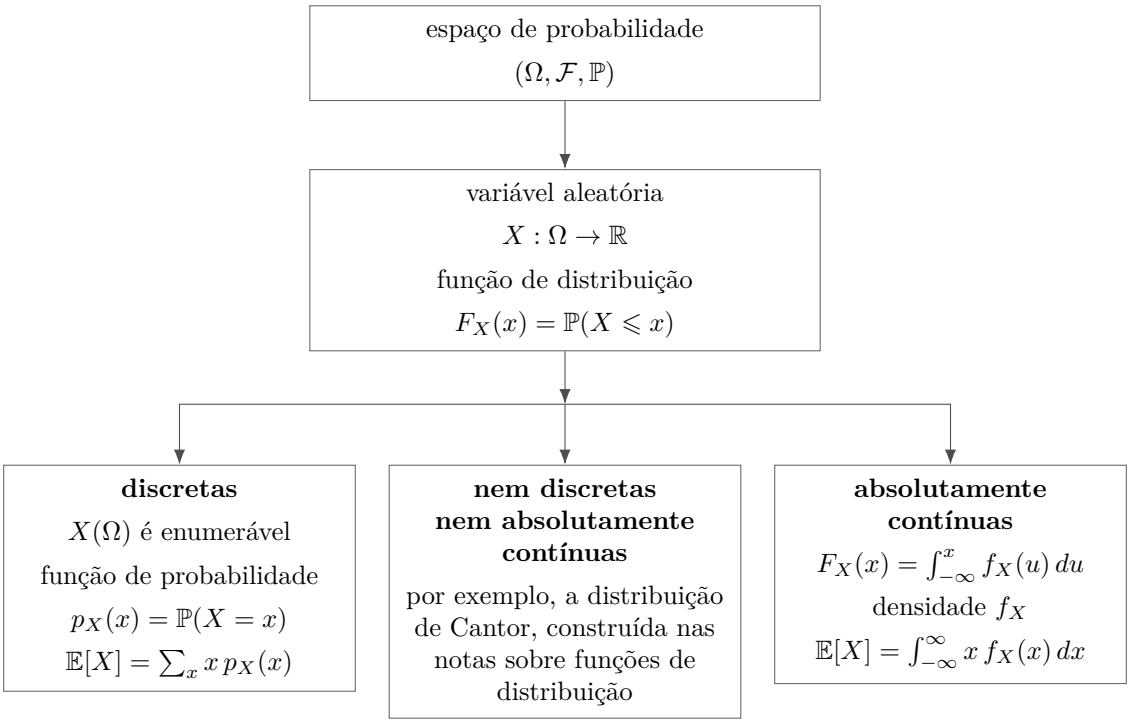



Figura 1: As classes de distribuições estudadas até aqui. Toda variável aleatória é um mapa $X : \Omega \to \mathbb{R}$ e induz uma função de distribuição $F_X$. As distribuições discretas e as absolutamente contínuas foram estudadas em notas anteriores; a distribuição de Cantor, também estudada antes, mostra que essas duas classes não esgotam todas as possibilidades.

A expressão (1) fornece uma maneira de fazer isso com as ferramentas do cálculo. Nas notas sobre funções de variáveis aleatórias contínuas e esperança, demonstramos essa identidade para variáveis discretas ou absolutamente contínuas e comentamos sua extensão ao caso geral. Podemos agora tomá-la como definição: para qualquer variável aleatória $Y \geqslant 0$, definimos

$$\mathbb{E}[Y] = \int_0^\infty \mathbb{P}(Y > t)\, dt = \lim_{R\to\infty} \int_0^R [1 - F_Y(t)]\, dt \in [0, +\infty]. \tag{1}$$

A função $t \mapsto \mathbb{P}(Y > t)$ é não crescente e limitada entre 0 e 1. Seu conjunto de descontinuidades é no máximo enumerável, e ela é integrável no sentido de Riemann em todo intervalo compacto. Essa integrabilidade também se verifica diretamente: em uma partição de $[0, R]$ em $n$ subintervalos iguais, a diferença entre as somas superior e inferior é no máximo $R/n$. Como o integrando é não negativo, o limite em (1) sempre existe, podendo ser infinito. Assim, essa definição usa apenas integrais de Riemann e um limite monótono, sem exigir a teoria abstrata da integral de Lebesgue, e coincide com as definições anteriores nos casos discreto e absolutamente contínuo.

Podemos definir a esperança para variáveis aleatórias gerais, que assumem tanto valores positivos quanto negativos, separando $X$ em suas partes positiva e negativa:

$$X^+ = \max\{X, 0\} \qquad \text{e} \qquad X^- = \max\{-X, 0\}.$$

Dizemos que $X$ é *integrável*, ou que possui esperança finita, quando

$$\mathbb{E}[|X|] = \mathbb{E}[X^+] + \mathbb{E}[X^-] < \infty.$$

Neste caso, definimos

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-] = \int_0^\infty \mathbb{P}(X > t)\,dt - \int_0^\infty \mathbb{P}(X < -t)\,dt. \tag{2}$$

As duas integrais devem ser finitas separadamente. A igualdade usada no critério de integrabilidade decorre de $\mathbb{P}(|X| > t) = \mathbb{P}(X > t) + \mathbb{P}(X < -t)$ para $t \geqslant 0$. Aplicando essas mesmas definições a $g(X)$, sempre que $g(X)$ for uma variável aleatória, também damos sentido a $\mathbb{E}[g(X)]$ sem supor que $X$ tenha densidade ou função de probabilidade.

Em particular, para $g(x) = x^n$, com $n \geqslant 1$ inteiro, a mudança de variável $t = u^n$ na expressão (1), aplicada a $X^n$, dá, quando $X \geqslant 0$,

$$\mathbb{E}[X^n] = \int_0^\infty \mathbb{P}(X^n > t)\,dt = n\int_0^\infty u^{n-1}\mathbb{P}(X > u)\,du. \tag{3}$$

Essa igualdade vale também com valor $+\infty$. Para uma variável $X$ de sinal qualquer, supondo $\mathbb{E}[|X|^n] < \infty$, obtemos

$$\mathbb{E}[X^n] = n\int_0^\infty u^{n-1}\Big[\mathbb{P}(X > u) + (-1)^n\mathbb{P}(X < -u)\Big]\,du. \tag{4}$$

De fato, para $n$ par, $X^n = |X|^n$ e as duas caudas se somam; para $n$ ímpar, aplicamos (2) a $X^n$ e fazemos a mesma mudança de variável em cada integral. A hipótese de integrabilidade garante que as integrais das duas caudas, com o peso $nu^{n-1}$, sejam finitas separadamente.

Esse ponto de vista permite definir a esperança sem restringir o tipo de variável aleatória e sem pressupor a construção da integral de Lebesgue. Para manter o foco nas ideias desta nota, não desenvolveremos toda a teoria da esperança a partir de (2). Nos cálculos concretos, continuaremos usando sobretudo densidades e funções de probabilidade. Sempre que um argumento se limitar a um caso particular, indicaremos essa restrição e, quando pertinente, o que seria necessário para tratar o caso geral. Também distinguiremos as demonstrações completas das ideias de prova e dos resultados usados sem demonstração.

Em particular, usaremos a linearidade da esperança para variáveis integráveis e o fato de que, para $U$ e $V$ integráveis e independentes, o produto $UV$ é integrável e satisfaz $\mathbb{E}[UV] = \mathbb{E}[U]\mathbb{E}[V]$. Essas propriedades foram estudadas nos casos discreto e absolutamente contínuo; aqui admitiremos, sem demonstração, sua validade para distribuições gerais. Uma dedução a partir de (2) pode ser feita por aproximação e passagem ao limite, mas não a desenvolveremos aqui. Já a monotonicidade da esperança segue da comparação das probabilidades nas integrais que a definem. Assim, os argumentos algébricos baseados nessas propriedades serão válidos para quaisquer distribuições, sob as hipóteses de integrabilidade indicadas, mesmo quando não escrevermos a esperança como soma ou integral contra uma densidade.

Uma regra prática para converter um argumento sobre variáveis absolutamente contínuas em um argumento sobre variáveis discretas é substituir "$f_X(x)\,dx$" por "$p_X(x)$" e a integral pela soma:

$$f_X(x)\,dx \;\longmapsto\; p_X(x), \qquad \int \;\longmapsto\; \sum. \tag{5}$$

A correspondência em (5) não é um teorema, e sim um guia de leitura: ela diz que, nos argumentos que faremos, o caso discreto se obtém do caso absolutamente contínuo trocando

a densidade pela função de probabilidade e o símbolo de integração pelo de soma, sem que a estrutura da dedução se altere.

Passamos ao estudo dos momentos, usando a expressão (2) para a esperança e sua aplicação às potências em (4) como base para a definição geral dos momentos. As expressões por somas e densidades continuarão sendo nossas principais ferramentas de cálculo.

# 2. Momentos

## 2.1. Definição e Primeiros Exemplos

A média de uma variável aleatória $X$ é apenas o primeiro de uma família de números que descrevem sua distribuição. Nestas notas, vamos estudar os *momentos* de $X$, isto é, as esperanças de suas potências não negativas, e ver como essa sequência de números permite extrair informações sobre a distribuição. Começamos pelos cálculos diretos, que mostram o que a definição exige e o que ela eventualmente não nos fornece.

**Definição 1** (Momento de ordem $k$)**.** Seja $X$ uma variável aleatória e seja $k \geqslant 1$ um inteiro. O *momento de ordem* $k$ de $X$ é o número $\mathbb{E}[X^k]$, sempre que essa esperança existir e for finita. Nesse caso dizemos que $X$ possui momento de ordem $k$; caso contrário, dizemos que o momento de ordem $k$ não existe. Por conveniência, também falamos do momento de ordem 0, que é sempre $\mathbb{E}[X^0] = 1$.

Como a esperança é calculada a partir da distribuição de $X$, os momentos dependem apenas dela, e não do espaço de probabilidade particular em que $X$ está definida. Por isso também faz sentido falar dos momentos de uma distribuição ou de uma densidade.

Para qualquer distribuição, a Definição 1 exige $\mathbb{E}[|X|^k] < \infty$. Pela expressão (1), aplicada a $Y = |X|^k$, e pela mudança de variável $t = u^k$, temos

$$\mathbb{E}[|X|^k] = \int_0^\infty \mathbb{P}(|X|^k > t)\, dt = k \int_0^\infty u^{k-1} \mathbb{P}(|X| > u)\, du. \tag{6}$$

Essa identidade vale também com valor $+\infty$ e dá um critério de existência dos momentos mesmo quando $X$ não é discreta nem absolutamente contínua. Quando esse valor é finito, calculamos $\mathbb{E}[X^k]$ por (4), que leva em conta a paridade de $k$.

Nos dois casos já estudados, recuperamos as expressões usuais. Se $X$ é discreta, então

$$\mathbb{E}[X^k] = \sum_{x \in \operatorname{Im} X} x^k\, \mathbb{P}(X = x),$$

e a soma deve convergir absolutamente. Se $X$ tem densidade $f$, então

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

e a integral de $|x|^k f(x)$ precisa ser finita. Assim, a convergência absoluta nas fórmulas por somas ou densidades expressa o mesmo critério de integrabilidade. Mesmo no caso de potências ímpares, os eventuais cancelamentos entre valores positivos e negativos não são suficientes para substituir a condição (6).

**Exemplo 2** (Momentos da distribuição exponencial)**.** Fixe $\lambda > 0$ e suponha que $X$ tenha densidade $f(x) = \lambda e^{-\lambda x}$ para $x > 0$ e $f(x) = 0$ para $x \leqslant 0$. Como $X \geqslant 0$, a condição de integrabilidade de $X^k$ coincide com a convergência da integral que define $\mathbb{E}[X^k]$, e podemos escrever, para $k \geqslant 1$,

$$\mathbb{E}[X^k] = \int_0^\infty x^k \lambda e^{-\lambda x}\, dx.$$

Vamos calcular essa integral por partes, escolhendo derivar o fator polinomial $x^k$ e integrar o fator exponencial. Tomando $u = x^k$ e $dv = \lambda e^{-\lambda x}\, dx$, temos $du = kx^{k-1}\, dx$ e $v = -e^{-\lambda x}$, logo

$$\mathbb{E}[X^k] = \Big[-x^k e^{-\lambda x}\Big]_0^\infty + \int_0^\infty kx^{k-1} e^{-\lambda x}\, dx.$$

O termo de fronteira se anula nas duas extremidades: em $x = 0$, porque $x^k$ se anula e $e^{-\lambda x}$ vale 1; e quando $x \to \infty$, porque a exponencial decai mais rápido do que qualquer polinômio cresce, isto é, $\lim_{x\to\infty} x^k e^{-\lambda x} = 0$. Quanto à integral que sobra,

$$\int_0^\infty kx^{k-1} e^{-\lambda x}\, dx = \frac{k}{\lambda} \int_0^\infty x^{k-1} \lambda e^{-\lambda x}\, dx = \frac{k}{\lambda}\, \mathbb{E}[X^{k-1}],$$

onde a última igualdade usa novamente a definição do momento de ordem $k-1$. Como o integrando é não negativo, a integração por partes pode ser feita em $[0, R]$ e seguida da passagem ao limite $R \to \infty$, de modo que a recorrência abaixo vale em $[0, +\infty]$, isto é, ela é válida mesmo que um dos dois lados seja infinito. Reunindo os dois passos, obtemos a recorrência

$$\mathbb{E}[X^k] = \frac{k}{\lambda}\, \mathbb{E}[X^{k-1}], \qquad k \geqslant 1.$$

Aplicando a identidade acima recursivamente, ficamos com

$$\begin{aligned} \mathbb{E}[X^k] &= \frac{k}{\lambda}\, \mathbb{E}[X^{k-1}] \\ &= \frac{k(k-1)}{\lambda^2}\, \mathbb{E}[X^{k-2}] \\ &= \cdots \\ &= \frac{k!}{\lambda^k}\, \mathbb{E}[X^0] \\ &= \frac{k!}{\lambda^k}, \end{aligned}$$

pois $X^0 = 1$ e a esperança da constante 1 é 1. Para $k = 1$ recuperamos $\mathbb{E}[X] = 1/\lambda$, o valor já conhecido para a média da exponencial. Mais do que isso, como $\mathbb{E}[X^0] = 1$ é finita, a recorrência fornece, por indução, $\mathbb{E}[X^k] < \infty$ para todo $k$: se $\mathbb{E}[X^{k-1}] < \infty$, então $\mathbb{E}[X^k] = (k/\lambda)\mathbb{E}[X^{k-1}] < \infty$. Logo a distribuição exponencial possui momentos de todas as ordens, dados por

$$\mathbb{E}[X^k] = \frac{k!}{\lambda^k}, \qquad \forall k \in \{0, 1, 2, \ldots\}.$$

**Exemplo 3** (A distribuição de Cauchy não possui momentos)**.** Dizemos que $X$ tem distribuição de Cauchy quando sua densidade é

$$f(x) = \frac{1}{\pi(1 + x^2)}, \qquad x \in \mathbb{R}.$$

Se $X$ tivesse momento de ordem $k$, ele seria dado por

$$\mathbb{E}[X^k] = \int_{-\infty}^{\infty} \frac{x^k}{\pi(1+x^2)}\,dx.$$

Vamos mostrar que essa integral não converge absolutamente para nenhum $k \geqslant 1$, de modo que a distribuição de Cauchy não possui momento algum.

Para $|x| \geqslant 1$ temos $1 + x^2 \leqslant 2x^2$, logo

$$\left|\frac{x^k}{\pi(1+x^2)}\right| = \frac{|x|^k}{\pi(1+x^2)} \geqslant \frac{|x|^k}{2\pi x^2} = \frac{1}{2\pi}\,|x|^{k-2}.$$

Como $k - 2 \geqslant -1$, a integral de $|x|^{k-2}$ em $\{|x| \geqslant 1\}$ diverge, e a comparação garante que

$$\int_{-\infty}^{\infty} \left|\frac{x^k}{\pi(1+x^2)}\right| dx = \infty.$$

Para $k = 1$ vale a pena notar por que isso não é um detalhe técnico. O integrando $x/(\pi(1+x^2))$ é uma função ímpar, e o valor principal simétrico da integral se anula; mas a esperança exige convergência absoluta, e podemos refazer o argumento acima separando $\{|x| \geqslant 1\}$ do intervalo limitado $[-1, 1]$ para ver que a integral de $|x|/(\pi(1+x^2))$ diverge. Concluímos que $\mathbb{E}[X]$ não existe, e o mesmo vale para todo $k \geqslant 1$.

O exemplo anterior sugere que o comportamento da densidade no infinito controla quais momentos existem. Vamos tornar isso preciso com uma família de densidades de cauda polinomial, para as quais os momentos existem apenas até uma certa ordem.

**Exemplo 4** (Densidades com cauda polinomial)**.** Fixe um inteiro $m \geqslant 2$ e considere

$$f(x) = \frac{c_m}{1+|x|^m}, \qquad x \in \mathbb{R},$$

onde a constante

$$c_m = \left(\int_{-\infty}^{\infty} \frac{dx}{1+|x|^m}\right)^{-1}$$

é escolhida de modo que a integral de $f$ seja igual a 1. A constante é positiva e finita: o integrando é positivo e contínuo, e para $|x| \geqslant 1$ ele é dominado por $|x|^{-m}$, cuja integral converge em $\{|x| \geqslant 1\}$ porque $m \geqslant 2$. Portanto, $f$ é de fato uma densidade de probabilidade.

Vamos determinar para quais $k$ a integral

$$\int_{-\infty}^{\infty} \frac{|x|^k}{1+|x|^m}\,dx$$

converge. O comportamento no infinito é o único ponto delicado, pois o integrando é contínuo e limitado em qualquer intervalo limitado. Para $|x| \geqslant 1$ temos $|x|^m \leqslant 1 + |x|^m \leqslant 2|x|^m$, e portanto

$$\frac{1}{2}\,|x|^{k-m} \leqslant \frac{|x|^k}{1+|x|^m} \leqslant |x|^{k-m}.$$

As duas desigualdades mostram que a integral em $\{|x| \geqslant 1\}$ converge se, e somente se, $\int_1^\infty x^{k-m}\,dx$ converge, o que ocorre exatamente quando $k - m < -1$, isto é, $k < m - 1$. Como $k$ e $m$ são inteiros, essa condição é equivalente a $k \leqslant m - 2$.

Concluímos que a densidade $f$ possui momentos apenas para as ordens $1 \leqslant k \leqslant m-2$: ela tem momento de ordem $k$ para todo inteiro $k$ nesse intervalo e nenhum momento para $k \geqslant m-1$. Conforme $m$ aumenta, a cauda fica mais leve e mais momentos passam a existir; para $m = 4$, por exemplo, existem $\mathbb{E}[X]$ e $\mathbb{E}[X^2]$, mas não $\mathbb{E}[X^3]$. No extremo oposto, $m = 2$ dá $f(x) = c_2/(1+x^2)$, que é a densidade de Cauchy do Exemplo 3, e a condição $k \leqslant 0$ não deixa nenhuma ordem admissível, como já sabíamos.

Conhecida a função de distribuição de $X$, podemos calcular seus momentos, sempre que eles existirem, pela expressão (4), qualquer que seja o tipo da distribuição. Cabe então perguntar se vale a recíproca: dada a sequência dos momentos finitos

$$\mathbb{E}[X],\ \mathbb{E}[X^2],\ \mathbb{E}[X^3],\ \ldots,$$

é possível reconstruir a distribuição de $X$? A resposta geral é não: existem variáveis aleatórias distintas que compartilham toda a sequência de momentos. A resposta passa a ser sim quando acrescentamos alguma informação extra sobre a sequência, e a próxima subseção trata exatamente disso.

## 2.2. Os Momentos e a Determinação da Distribuição

Os momentos de uma variável aleatória $X$, isto é, as esperanças $\mathbb{E}[X]$, $\mathbb{E}[X^2]$, $\mathbb{E}[X^3]$, $\ldots$, quando existem, codificam informações sobre a distribuição de $X$. Uma pergunta natural é se a sequência de momentos determina unicamente essa distribuição. Vamos ver que a resposta é afirmativa sob uma hipótese de convergência, que explicita a relação entre a sequência de momentos e uma função geradora.

**Teorema 5** (Unicidade dos momentos)**.** Suponha que todos os momentos $\mathbb{E}[X]$, $\mathbb{E}[X^2]$, $\ldots$ da variável aleatória $X$ existam e que a série

$$\sum_{k=0}^{\infty} \frac{1}{k!}\, t^k\, \mathbb{E}[X^k] \tag{7}$$

seja absolutamente convergente para algum $t > 0$. Então a sequência de momentos determina unicamente a distribuição de $X$.

A demonstração usual deste fato utiliza resultados de análise real e complexa que não desenvolveremos nestas notas. Por isso, omitimos a prova, que pode ser encontrada em textos de Análise.

Observamos que o enunciado vale para distribuições gerais, sem hipótese de existência de densidade. O ponto de contato com a probabilidade está na forma de (7): essa série é a função geradora exponencial da sequência $(\mathbb{E}[X^k])_{k\geqslant 0}$, e o teorema é o análogo, para momentos, do teorema de unicidade da função geradora de probabilidade, que estudamos nas notas sobre funções geradoras de probabilidade. Assim como lá a função geradora determinava a distribuição, aqui a convergência absoluta da série faz a sequência de momentos determinar a distribuição.

O papel lógico da hipótese é o de uma condição suficiente para a conclusão: a convergência absoluta de (7) para algum $t > 0$ basta para que os momentos determinem a distribuição. O teorema não afirma que essa convergência seja necessária, e o exemplo da subseção seguinte mostra que a hipótese não é supérflua: sem hipótese alguma sobre a convergência da série, os momentos podem deixar de determinar a distribuição.

## 2.3. Um Contraexemplo: A Distribuição Log-Normal

Vamos construir uma distribuição que não é determinada pelos seus momentos. Para isso, nosso ponto de partida será considerar uma variável aleatória com distribuição log-normal. Isto é, dizemos que $Y$ tem a distribuição log-normal (padrão) se $Y$ é da forma $Y = e^X$, com $X \sim \mathrm{N}(0,1)$. Vamos verificar que se $Y$ é uma variável aleatória log-normal, então $Y$ possui uma função de densidade. De fato, para $x > 0$, a função de distribuição de $Y$ é dada por $\mathbb{P}(Y \leqslant x) = \mathbb{P}(X \leqslant \log x)$, e derivando em relação a $x$ (ou pela fórmula de mudança de variável para transformações crescentes e diferenciáveis) obtemos a densidade

$$f(x) = \begin{cases} \dfrac{1}{x\sqrt{2\pi}} \exp\Big[-\dfrac{1}{2}(\log x)^2\Big], & \text{se } x > 0, \\ 0, & \text{se } x \leqslant 0. \end{cases} \tag{8}$$

**Exemplo 6** (Momentos que não determinam a distribuição)**.** Fixe $a$ com $-1 \leqslant a \leqslant 1$ e defina

$$f_a(x) = \Big[1 + a \operatorname{sen}(2\pi \log x)\Big] f(x) \qquad \text{para } x > 0,$$

com $f_a(x) = 0$ para $x \leqslant 0$. Vamos verificar que

(i) $f_a$ é uma densidade,

(ii) $f$ tem momentos finitos de todas as ordens,

(iii) $f_a$ e $f$ têm momentos iguais de todas as ordens.

O fator $1 + a \operatorname{sen}(2\pi \log x)$ modifica a densidade $f$. Vamos mostrar que essa modificação preserva tanto a integral da densidade quanto todos os seus momentos. A presença de $\log x$ sugere a substituição $u = \log x$: ela transforma a expressão $e^{-(\log x)^2/2}$ em $e^{-u^2/2}$, que aparece na densidade da normal padrão. Faremos essa substituição explicitamente em cada uma das contas.

**(i) A função $f_a$ é uma densidade.** Precisamos verificar que $f_a$ é mensurável, não negativa e tem integral igual a 1. A mensurabilidade decorre de ela ser contínua em $(0, \infty)$ e nula em $(-\infty, 0]$. Para verificar o sinal, usamos $|\operatorname{sen} t| \leqslant 1$ e $|a| \leqslant 1$:

$$1 + a \operatorname{sen}(2\pi \log x) \geqslant 1 - |a \operatorname{sen}(2\pi \log x)| \geqslant 1 - |a| \geqslant 0.$$

Como $f(x) \geqslant 0$, segue que $f_a(x) \geqslant 0$ para $x > 0$. Para $x \leqslant 0$, isso também vale, pois $f_a(x) = 0$.

Agora vamos calcular a integral de $f_a$ ao longo de toda a reta real. Primeiro, vamos conferir diretamente que a integral de $f$ é 1. Fazemos

$$u = \log x, \qquad x = e^u, \qquad dx = e^u\, du.$$

Quando $x \to 0^+$, temos $u \to -\infty$; quando $x \to +\infty$, temos $u \to +\infty$. Substituindo também o fator $1/x$ por $1/e^u$, obtemos

$$\begin{aligned} \int_0^\infty f(x)\, dx &= \int_0^\infty \frac{1}{x\sqrt{2\pi}} e^{-(\log x)^2/2}\, dx \\ &= \int_{-\infty}^\infty \frac{1}{e^u\sqrt{2\pi}} e^{-u^2/2}\, e^u\, du \\ &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du = 1. \end{aligned}$$

Na segunda linha, os fatores $1/e^u$ e $e^u$ se cancelam. A última integral vale 1 porque é a integral da densidade da normal padrão.

Para calcular a integral de $f_a$, falta examinar o termo que contém o seno. Essa integral é absolutamente convergente, pois

$$\int_0^\infty \Big|\operatorname{sen}(2\pi \log x) f(x)\Big|\, dx \leqslant \int_0^\infty f(x)\, dx = 1.$$

Usando novamente $x = e^u$, $\log x = u$ e $dx = e^u\, du$, temos

$$\begin{aligned}\int_0^\infty \operatorname{sen}(2\pi \log x) f(x)\, dx &= \int_{-\infty}^\infty \operatorname{sen}(2\pi u) \frac{1}{e^u\sqrt{2\pi}} e^{-u^2/2}\, e^u\, du \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \operatorname{sen}(2\pi u) e^{-u^2/2}\, du \\ &= 0.\end{aligned}$$

Essa última integral é zero: o integrando é ímpar, por ser o produto de uma função par por uma função ímpar, e é absolutamente integrável sobre a reta, que é simétrica em torno da origem.

Usando os dois resultados obtidos acima, concluímos que

$$\begin{aligned}\int_{\mathbb{R}} f_a(x)\, dx &= \int_0^\infty \Big[1 + a \operatorname{sen}(2\pi \log x)\Big] f(x)\, dx \\ &= \int_0^\infty f(x)\, dx + a \int_0^\infty \operatorname{sen}(2\pi \log x) f(x)\, dx \\ &= 1 + a \cdot 0 = 1.\end{aligned}$$

Isso conclui a verificação de que $f_a$ é uma densidade.

**(ii) Cálculo dos momentos de $f$.** Fixe um inteiro $k \geqslant 1$. Como a densidade é nula para $x \leqslant 0$, o momento de ordem $k$ é $\int_0^\infty x^k f(x)\, dx$. Para calcular estes momentos vamos usar mais uma vez a substituição

$$u = \log x, \qquad x = e^u, \qquad x^k = e^{ku}, \qquad dx = e^u\, du.$$

Os limites 0 e $+\infty$ passam a $-\infty$ e $+\infty$, respectivamente. Assim,

$$\begin{aligned}\int_0^\infty x^k f(x)\, dx &= \int_0^\infty x^k \frac{1}{x\sqrt{2\pi}} e^{-(\log x)^2/2}\, dx \\ &= \int_{-\infty}^\infty e^{ku} \frac{1}{e^u\sqrt{2\pi}} e^{-u^2/2}\, e^u\, du \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{ku - u^2/2}\, du.\end{aligned}$$

Queremos reconhecer novamente uma integral da densidade normal. Para isso, vamos escrever o expoente como o negativo de um quadrado mais uma constante. Como $(u-k)^2 = u^2 - 2ku + k^2$, temos

$$ku - \frac{u^2}{2} = -\frac{1}{2}(u^2 - 2ku) = -\frac{1}{2}\Big[(u-k)^2 - k^2\Big] = -\frac{(u-k)^2}{2} + \frac{k^2}{2}.$$

Consequentemente,

$$e^{ku-u^2/2} = e^{k^2/2} e^{-(u-k)^2/2}.$$

O fator $e^{k^2/2}$ não depende de $u$ e pode sair da integral. Logo,

$$\int_0^\infty x^k f(x)\,dx = e^{k^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-(u-k)^2/2}\,du.$$

Para calcular a integral restante, fazemos

$$v = u - k, \qquad u = v + k, \qquad du = dv.$$

Subtrair a constante $k$ não altera os limites infinitos. Portanto,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-(u-k)^2/2}\,du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-v^2/2}\,dv = 1.$$

Concluímos que

$$\int_0^\infty x^k f(x)\,dx = e^{k^2/2} < \infty \qquad \text{para } k = 1, 2, \ldots.$$

Assim, $f$ tem momentos finitos de todas as ordens.

**(iii) Os momentos de $f_a$ são iguais aos de $f$.** Fixe um inteiro $k \geqslant 1$. Antes de separar os termos da integral, observamos que $0 \leqslant f_a(x) \leqslant 2f(x)$ e, portanto,

$$0 \leqslant \int_0^\infty x^k f_a(x)\,dx \leqslant 2\int_0^\infty x^k f(x)\,dx = 2e^{k^2/2} < \infty.$$

Além disso, $|x^k \operatorname{sen}(2\pi \log x) f(x)| \leqslant x^k f(x)$, de modo que o termo com seno também é absolutamente integrável. Pela definição de $f_a$,

$$\int_0^\infty x^k f_a(x)\,dx = \int_0^\infty x^k f(x)\,dx + a\int_0^\infty x^k \operatorname{sen}(2\pi \log x) f(x)\,dx.$$

O primeiro termo já foi calculado em (ii). Para mostrar que os momentos são iguais, precisamos mostrar que o segundo termo é zero. Vamos chamar a integral desse termo de

$$J_k = \int_0^\infty x^k \operatorname{sen}(2\pi \log x) f(x)\,dx.$$

Fazemos novamente $u = \log x$, de modo que $x = e^u$, $x^k = e^{ku}$ e $dx = e^u\,du$. Os limites passam de $(0, \infty)$ para $(-\infty, \infty)$, e obtemos

$$\begin{aligned} J_k &= \int_0^\infty x^k \operatorname{sen}(2\pi \log x) \frac{1}{x\sqrt{2\pi}} e^{-(\log x)^2/2}\,dx \\ &= \int_{-\infty}^\infty e^{ku} \operatorname{sen}(2\pi u) \frac{1}{e^u\sqrt{2\pi}} e^{-u^2/2}\, e^u\,du \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \operatorname{sen}(2\pi u) e^{ku-u^2/2}\,du. \end{aligned}$$

Usamos a mesma identidade de (ii),

$$ku - \frac{u^2}{2} = -\frac{(u-k)^2}{2} + \frac{k^2}{2},$$

para obter

$$J_k = \frac{e^{k^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \operatorname{sen}(2\pi u) e^{-(u-k)^2/2}\, du.$$

Para aproveitar o cancelamento por simetria usado em (i), queremos que o fator exponencial tenha a forma $e^{-v^2/2}$. Por isso, fazemos

$$v = u - k, \qquad u = v + k, \qquad du = dv.$$

Nessa substituição, precisamos trocar $u$ nos dois fatores do integrando:

$$e^{-(u-k)^2/2} = e^{-v^2/2}, \qquad \operatorname{sen}(2\pi u) = \operatorname{sen}\Big(2\pi(v+k)\Big).$$

Os limites continuam sendo $-\infty$ e $+\infty$. Assim,

$$J_k = \frac{e^{k^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \operatorname{sen}\Big(2\pi(v+k)\Big) e^{-v^2/2}\, dv.$$

É neste ponto que usamos o fato de $k$ ser inteiro: somar $2\pi k$ ao argumento do seno corresponde a dar um número inteiro de voltas no círculo trigonométrico e não muda seu valor. Portanto,

$$\operatorname{sen}\Big(2\pi(v+k)\Big) = \operatorname{sen}(2\pi v + 2\pi k) = \operatorname{sen}(2\pi v).$$

Com isso,

$$J_k = \frac{e^{k^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \operatorname{sen}(2\pi v) e^{-v^2/2}\, dv = 0.$$

A integral é zero pelo mesmo motivo explicado em (i): o seno é ímpar, o fator exponencial é par, e as integrais sobre as duas metades da reta se cancelam. Voltando ao momento de $f_a$, concluímos que

$$\int_0^{\infty} x^k f_a(x)\, dx = e^{k^2/2} + aJ_k = e^{k^2/2} = \int_0^{\infty} x^k f(x)\, dx.$$

Isso vale para todo inteiro $k \geqslant 1$. Para $k = 0$, a igualdade também vale, pois ambas as densidades têm integral 1.

**As distribuições são diferentes.** Falta verificar que variar $a$ realmente produz distribuições diferentes. Suponha que $a > a'$. Pela definição,

$$f_a(x) - f_{a'}(x) = (a - a') \operatorname{sen}(2\pi \log x) f(x), \qquad x > 0.$$

Se $1 < x < e^{1/2}$, então $0 < \log x < 1/2$ e, portanto, $0 < 2\pi \log x < \pi$. Nesse intervalo, $\operatorname{sen}(2\pi \log x) > 0$. Como também $a - a' > 0$ e $f(x) > 0$, temos $f_a(x) > f_{a'}(x)$ em todo esse intervalo. Integrando, obtemos

$$\int_1^{e^{1/2}} f_a(x)\, dx > \int_1^{e^{1/2}} f_{a'}(x)\, dx.$$

Essas integrais são as probabilidades que as duas distribuições atribuem ao intervalo $(1, e^{1/2})$. Como as probabilidades são diferentes, as distribuições também são diferentes. Se $a < a'$, basta trocar os papéis dos parâmetros no argumento.

A figura abaixo ilustra essa comparação para $a = 1$ e $a' = -1$. A região sombreada fica entre as duas densidades no intervalo $(1, e^{1/2})$: sua área é a diferença entre as probabilidades que elas atribuem a esse intervalo.

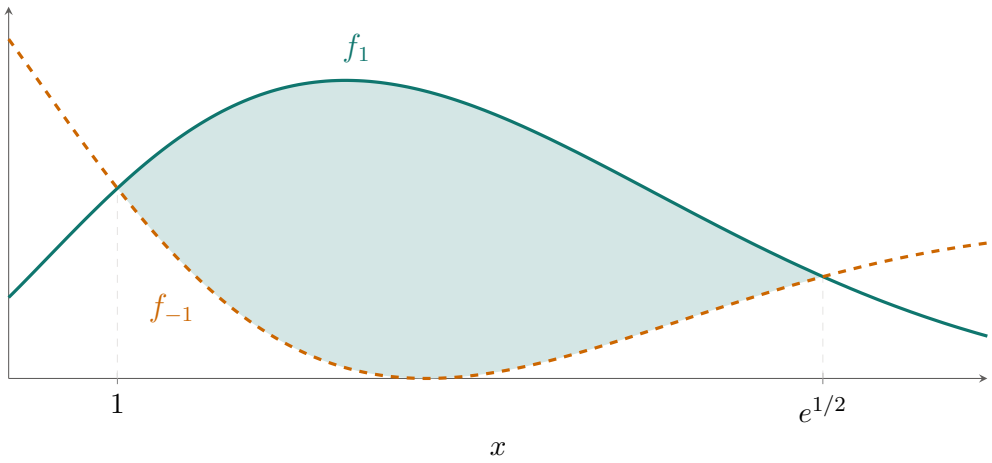



Figura 2: Um trecho de duas densidades com os mesmos momentos de todas as ordens. Entre 1 e $e^{1/2}$, a curva de $f_1$ fica acima da curva de $f_{-1}$. A área sombreada é positiva, mostrando que as duas distribuições atribuem probabilidades diferentes a esse intervalo.

Construímos, assim, uma família de distribuições distintas, cujas densidades correspondentes são dadas pela família de funções $\{f_a : -1 \leqslant a \leqslant 1\}$, que possuem os mesmos momentos de todas as ordens. Logo, conhecer todos os momentos não basta, em geral, para determinar uma distribuição.

Concluímos que a hipótese de Teorema 5 não é supérflua: as densidades $f_a$ acima são distintas, mas compartilham a sequência de momentos

$$\big(\mathbb{E}[Y^k]\big)_{k\geqslant 1} = \big(e^{k^2/2}\big)_{k\geqslant 1}.$$

Para esta família a série de potências (7) diverge para todo $t > 0$: a razão entre termos consecutivos é $\frac{t\,e^{k+1/2}}{k+1}$, que tende a $+\infty$, de modo que os termos gerais nem sequer tendem a zero. Temos assim o contraponto do teorema de unicidade: sem a convergência absoluta em (7), a sequência de momentos pode deixar de determinar a distribuição.

## 2.4. A Variância e suas Propriedades

A variância de uma variável aleatória já apareceu duas vezes no curso: nas notas sobre esperança de variáveis discretas, nas quais foi definida por meio da função de probabilidade, e nas notas sobre funções de variáveis contínuas e esperança, em que foi definida por meio da densidade. Naquele momento, era necessário escolher uma forma de somar ou integrar segundo o tipo da variável. Agora que dispomos de uma noção de esperança que não depende do tipo, podemos reapresentar a definição de uma só vez e estudar as propriedades da variância de maneira unificada.

**Definição 7** (Variância)**.** Seja $X$ uma variável aleatória com $\mathbb{E}[X^2] < \infty$ e escreva $\mu = \mathbb{E}[X]$. A *variância* de $X$ é

$$\operatorname{Var}(X) = \mathbb{E}\Big[(X-\mu)^2\Big]. \tag{9}$$

A exigência $\mathbb{E}[X^2] < \infty$ garante primeiro que $\mu$ é finita, pois $|X| \leqslant 1 + X^2$. Em seguida, $(X-\mu)^2 \leqslant 2X^2 + 2\mu^2$ garante que a esperança em (9) também é finita. Usamos aqui a monotonicidade e a linearidade da esperança, com o alcance fixado na Seção 1. Essas estimativas valem para qualquer distribuição; não precisamos supor que $X$ seja discreta ou tenha densidade.

A variância mede a dispersão de $X$ em torno de sua esperança $\mu$. Se $X$ assume com frequência valores consideravelmente distantes de $\mu$, então $|X-\mu|$ é frequentemente grande e $\mathbb{E}[(X-\mu)^2]$ também será grande; se $X$ costuma ficar perto de $\mu$, então $|X-\mu|$ é tipicamente pequeno e $\mathbb{E}[(X-\mu)^2]$ é pequena. O caso extremo é o de uma variável *concentrada* em um único ponto, e é isso que caracterizamos a seguir.

Começamos por um critério de anulamento que vale para qualquer variável aleatória quadrado-integrável.

**Lema 8.** Para uma variável aleatória $Y$ com $\mathbb{E}[Y^2] < \infty$,

$$\mathbb{E}[Y^2] = 0 \quad \text{se, e somente se,} \quad \mathbb{P}(Y = 0) = 1. \tag{10}$$

**Prova**. Uma das implicações é imediata. Se $\mathbb{P}(Y=0)=1$, então $Y^2 = 0$ quase certamente e, pela definição de esperança de uma variável que se anula em um evento de probabilidade 1, temos $\mathbb{E}[Y^2] = 0$.

Para a recíproca, suponha que $\mathbb{E}[Y^2] = 0$ e fixe $\varepsilon > 0$. Para $0 \leqslant t < \varepsilon^2$, temos $\mathbb{P}(Y^2 > t) \geqslant \mathbb{P}(Y^2 \geqslant \varepsilon^2)$. Integrando essa comparação e usando (1), obtemos

$$\varepsilon^2 \, \mathbb{P}\Big(Y^2 \geqslant \varepsilon^2\Big) \leqslant \int_0^{\varepsilon^2} \mathbb{P}(Y^2 > t)\, dt \leqslant \mathbb{E}[Y^2] = 0,$$

e portanto $\mathbb{P}(|Y| \geqslant \varepsilon) = 0$ para todo $\varepsilon > 0$. Escrevemos agora o evento $\{Y \neq 0\}$ como união enumerável dos eventos aninhados $\{|Y| \geqslant 1/n\}$, com $n \in \mathbb{N}$:

$$\{Y \neq 0\} = \bigcup_{n \geqslant 1} \{|Y| \geqslant 1/n\}.$$

Pela continuidade por baixo da probabilidade aplicada a essa sequência crescente de eventos,

$$\mathbb{P}(Y \neq 0) = \lim_{n \to \infty} \mathbb{P}\Big(|Y| \geqslant 1/n\Big) = \lim_{n \to \infty} 0 = 0.$$

Logo $\mathbb{P}(Y \neq 0) = 0$ e, tomando o complementar, $\mathbb{P}(Y = 0) = 1$. ∎

No caso de uma variável aleatória discreta, a recíproca admite uma leitura puramente somatória, que registramos porque esclarece de onde vem o argumento acima. Escrevendo a esperança como soma sobre a imagem de $Y$,

$$\mathbb{E}[Y^2] = \sum_{y \in \operatorname{Im} Y} y^2 \, \mathbb{P}(Y = y),$$

vemos que todos os termos dessa soma são não negativos. A soma, portanto, só pode se anular se cada termo for nulo; como $y^2 = 0$ apenas para $y = 0$, concluímos que $\mathbb{P}(Y = y) = 0$ para todo $y \neq 0$ e, somando sobre a parte não nula da imagem, $\mathbb{P}(Y \neq 0) = 0$. O argumento do Lema 8 apenas substitui essa soma por uma estimativa integral, o que permite tratar ao mesmo tempo os casos discreto, contínuo e singular.

Aplicando (10) a $Y = X - \mu$, cuja esperança do quadrado é $\operatorname{Var}(X)$, obtemos a caracterização do caso sem dispersão alguma:

$$\operatorname{Var}(X) = 0 \quad \text{se, e somente se,} \quad \mathbb{P}(X = \mathbb{E}[X]) = 1. \tag{11}$$

Isto é, a variância se anula exatamente quando $X$ é constante com probabilidade 1, de modo que "variância nula" significa "nenhuma dispersão".

Existem outras medidas possíveis de dispersão, como $\mathbb{E}[|X - \mu|]$ ou $\mathbb{E}[|X - \mu|^3]$, mas é com a variância que se trabalha com mais facilidade. O motivo aparece na fórmula a seguir.

Quando calculamos a variância de uma variável concreta, é em geral mais simples usar os momentos de $X$ do que retornar à definição (9). Expandindo o quadrado e usando a linearidade da esperança, obtemos uma identidade que expressa a variância em termos do segundo momento e da média. Como todos os termos envolvidos são integráveis, a expansão é legítima e resulta em

$$\operatorname{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \tag{12}$$

De fato, escrevendo $\mu = \mathbb{E}[X]$,

$$\begin{aligned} \mathbb{E}\Big[(X-\mu)^2\Big] &= \mathbb{E}\Big[X^2 - 2\mu X + \mu^2\Big] \\ &= \mathbb{E}[X^2] - 2\mu\,\mathbb{E}[X] + \mu^2 \\ &= \mathbb{E}[X^2] - 2\mu^2 + \mu^2 = \mathbb{E}[X^2] - \mu^2. \end{aligned}$$

A vantagem de (12) é que ela separa o cálculo em dois momentos, $\mathbb{E}[X]$ e $\mathbb{E}[X^2]$, que costumam ser obtidos por somas ou integrais diretas, sem que seja preciso lidar com o quadrado de $X - \mathbb{E}[X]$.

A mesma linearidade controla o efeito de uma transformação afim.

**Proposição 9** (Variância de uma transformação afim)**.** Sejam $X$ uma variável aleatória com $\mathbb{E}[X^2] < \infty$ e $a, b \in \mathbb{R}$. Então

$$\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X). \tag{13}$$

**Prova**. Pela linearidade da esperança, $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$, de modo que o desvio de $aX + b$ em relação à sua média é $aX + b - \mathbb{E}[aX + b] = a\Big(X - \mathbb{E}[X]\Big)$. Logo

$$\begin{aligned} \operatorname{Var}(aX+b) &= \mathbb{E}\Big[\Big(aX + b - \mathbb{E}[aX+b]\Big)^2\Big] \\ &= \mathbb{E}\Big[\Big(aX + b - a\mathbb{E}[X] - b\Big)^2\Big] \\ &= \mathbb{E}\Big[a^2\Big(X - \mathbb{E}[X]\Big)^2\Big] = a^2\,\mathbb{E}\Big[\Big(X - \mathbb{E}[X]\Big)^2\Big] \\ &= a^2 \operatorname{Var}(X). \end{aligned}$$

A penúltima igualdade usa a homogeneidade da esperança para retirar a constante real finita $a^2$. ∎

Como medida de dispersão, a variância tem um inconveniente: ela é quadrática em $X$ e, portanto, não linear. Pela Proposição 9, se $a > 1$, a variância de $aX$ é $a^2$ vezes a variância de $X$, e não $a$ vezes; além disso, a variância é medida em unidades ao quadrado das unidades de $X$. Se, por exemplo, $X$ é um comprimento em metros, a variância $\mathrm{Var}(X)$ é expressa em metros quadrados. Por esse motivo, quem precisa de uma medida de dispersão na mesma escala de $X$ usa o *desvio padrão*

$$\sigma_X = \sqrt{\mathrm{Var}(X)}.$$

O desvio padrão herda a não linearidade apenas no fator de escala: pela Proposição 9, $\sigma_{aX+b} = |a|\,\sigma_X$, o que é compatível com a mudança de unidades descrita por $a$.

Vamos voltar nossa atenção novamente para a variância de uma soma. A pergunta natural é se $\mathrm{Var}(X+Y)$ se decompõe em $\mathrm{Var}(X) + \mathrm{Var}(Y)$. Veremos que isso não é verdade em geral e que o termo que pode alterar esta igualdade tem significado próprio.

**Proposição 10** (Variância de uma soma)**.** Sejam $X$ e $Y$ variáveis aleatórias com $\mathbb{E}[X^2] < \infty$ e $\mathbb{E}[Y^2] < \infty$. Então

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathbb{E}\Big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\Big]. \tag{14}$$

**Prova**. Escreva $\mu_X = \mathbb{E}[X]$ e $\mu_Y = \mathbb{E}[Y]$. Pela linearidade da esperança, $\mathbb{E}[X+Y] = \mu_X + \mu_Y$, de modo que $(X+Y) - \mathbb{E}[X+Y] = (X-\mu_X) + (Y-\mu_Y)$. Expandindo o quadrado dessa soma,

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathbb{E}\Big[\big\{(X-\mu_X) + (Y-\mu_Y)\big\}^2\Big] \\
&= \mathbb{E}\Big[(X-\mu_X)^2\Big] + 2\,\mathbb{E}\Big[(X-\mu_X)(Y-\mu_Y)\Big] + \mathbb{E}\Big[(Y-\mu_Y)^2\Big] \\
&= \mathrm{Var}(X) + 2\,\mathbb{E}\Big[(X-\mu_X)(Y-\mu_Y)\Big] + \mathrm{Var}(Y).
\end{aligned}$$

A validade da expansão do quadrado segue da linearidade da esperança: as três parcelas são integráveis, pois $\mathbb{E}[(X-\mu_X)^2] < \infty$, $\mathbb{E}[(Y-\mu_Y)^2] < \infty$ e

$$|(X-\mu_X)(Y-\mu_Y)| \leqslant \tfrac{1}{2}\Big[(X-\mu_X)^2 + (Y-\mu_Y)^2\Big],$$

o que garante que o termo misto é absolutamente integrável. ■

A Proposição 10 mostra que a variância da soma é a soma das variâncias acrescida de duas vezes o número $\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]$. Esse número mede o quanto as duas variáveis flutuam conjuntamente em torno de suas médias: ele é positivo quando $X$ e $Y$ tendem a ficar simultaneamente acima ou simultaneamente abaixo de suas médias, e negativo quando tendem a se deslocar em direções opostas. É essa quantidade que se chama *covariância* de $X$ e $Y$, e é o objeto que estudamos na próxima subseção.

## 2.5. Covariância

**Definição 11** (Covariância)**.** A covariância das variáveis aleatórias $X$ e $Y$ é a quantidade denotada $\mathrm{Cov}(X,Y)$ e dada por

$$\mathrm{Cov}(X,Y) = \mathbb{E}\Big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\Big], \tag{15}$$

sempre que essas esperanças existem.

Note que $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$: a covariância generaliza a variância, que mede a dispersão de uma única variável em torno de sua média, para o caso de duas variáveis.

A expressão em (15) pode ser reescrita de forma mais simples. Suponha que $X$ e $Y$ tenham segundo momento finito. Como $|XY| \leqslant (X^2 + Y^2)/2$, o produto $XY$ é integrável. Podemos então expandir o produto e usar a linearidade da esperança, para distribuições gerais obtendo

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \mathbb{E}\Big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\Big] \\ &= \mathbb{E}\Big[XY - X\mathbb{E}[Y] - \mathbb{E}[X]\,Y + \mathbb{E}[X]\mathbb{E}[Y]\Big] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y],\end{aligned} \tag{16}$$

onde usamos que $\mathbb{E}[X]$ e $\mathbb{E}[Y]$ são constantes. Vale a pena guardar a forma final: ela dispensa o cálculo prévio das médias dentro do produto e é, em geral, a mais rápida de calcular.

Voltando à expressão da variância de uma soma, obtida na subseção anterior, o termo do meio é $2\,\mathbb{E}\Big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\Big]$, e portanto

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + 2\,\mathrm{Cov}(X, Y) + \mathrm{Var}(Y), \tag{17}$$

é uma identidade válida para quaisquer $X$ e $Y$ com segundo momento finito, sem hipótese de independência.

Suponhamos agora, ainda com segundos momentos finitos, que $X$ e $Y$ sejam independentes. Usamos a fatoração $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, cuja validade para distribuições gerais foi admitida na Seção 1. No caso discreto, ela se demonstra fatorando a soma conjunta; no caso de um par com densidade conjunta, a independência permite fatorar a densidade e separar as integrais. Essas verificações não cobrem todas as distribuições, mas, admitida a propriedade geral, o argumento a seguir não exige outra adaptação.

Das observações acima e da identidade (16) concluímos que no caso de variáveis aleatórias $X$ e $Y$ independentes e com segundo momento finito temos

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0. \tag{18}$$

Ou seja, variáveis aleatórias independentes, com segundo momento finito, não são correlacionadas. Usando esse resultado na igualdade (17), recuperamos a fórmula já conhecida

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y), \tag{19}$$

válida sempre que as variâncias existem e que $X$ e $Y$ forem independentes.

A recíproca é falsa: covariância nula não implica independência. Como esse ponto costuma causar confusão, vamos examiná-lo num exemplo com as contas feitas.

**Exemplo 12** (Covariância nula não implica independência)**.** Seja $X$ uniforme em $\{-1, 0, 1\}$, isto é, $\mathbb{P}(X = x) = \frac{1}{3}$ para cada $x \in \{-1, 0, 1\}$, e defina $Y = X^2$. Como $Y$ é função de $X$, as variáveis são dependentes. De fato, observe que $\mathbb{P}(X = 1, Y = 1) = \frac{1}{3}$, enquanto $\mathbb{P}(X = 1)\mathbb{P}(Y = 1) = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}$. O que mostra que $X$ e $Y$ são dependentes. Por outro lado, elas possuem covariância nula, como mostram os cálculos abaixo:

$$\begin{aligned}\mathbb{E}[X] &= \tfrac{1}{3}(-1) + \tfrac{1}{3} \cdot 0 + \tfrac{1}{3} \cdot 1 = 0; \\ \mathbb{E}[Y] = \mathbb{E}[X^2] &= \tfrac{1}{3} \cdot 1 + \tfrac{1}{3} \cdot 0 + \tfrac{1}{3} \cdot 1 = \tfrac{2}{3}; \\ \mathbb{E}[XY] = \mathbb{E}[X^3] &= \mathbb{E}[X] = 0,\end{aligned}$$

pois $X$ assume apenas os valores $-1$, $0$ e $1$, e nesse conjunto $X^3 = X$ ponto a ponto e finalmente por (16),

$$\operatorname{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0 - 0 \cdot \tfrac{2}{3} = 0.$$

Portanto este é um exemplo em que $X$ e $Y$ são dependentes e, ainda assim, não correlacionadas. A razão do cancelamento está na simetria: quando $X$ se afasta da média para a direita, $Y$ também se afasta para a direita, mas o mesmo ocorre quando $X$ se afasta para a esquerda, e as duas contribuições se anulam.

Apesar de a recíproca falhar, a covariância é muito usada como medida de dependência, e há uma razão para isso: ela é um único número, e não um objeto complicado como uma densidade conjunta, e ainda assim carrega informação sobre o comportamento conjunto de $X$ e $Y$. Se $\operatorname{Cov}(X, Y) > 0$, os desvios tendem a ter o mesmo sinal; se $\operatorname{Cov}(X, Y) < 0$, a tendência se inverte.

A covariância tem, porém, uma desvantagem: ela não é invariante por mudança de escala. Se $X$ e $Y$ registram duas grandezas em metros e $U$ e $V$ registram as mesmas grandezas em centímetros, então $U = 100X$, $V = 100Y$ e, pela linearidade da esperança,

$$\operatorname{Cov}(U, V) = 100^2 \operatorname{Cov}(X, Y) = 10\,000 \operatorname{Cov}(X, Y).$$

A covariância muda por um fator $10\,000$, embora os dois pares descrevam as mesmas quantidades; só a unidade de medida mudou. Para contornar isso, reescalamos a covariância de modo a eliminar as unidades.

## 2.6. Correlação

**Definição 13** (Coeficiente de correlação)**.** O coeficiente de correlação das variáveis aleatórias $X$ e $Y$ é a quantidade $\rho(X, Y)$ dada por

$$\rho(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}, \tag{20}$$

sempre que essas quantidades existem e $\operatorname{Var}(X)\operatorname{Var}(Y) \neq 0$.

A condição $\operatorname{Var}(X)\operatorname{Var}(Y) \neq 0$ garante que o denominador não se anule. Além do mais, se uma das variáveis fosse constante, ela não teria dispersão, e não haveria o que medir conjuntamente.

Vejamos que a correlação é invariante por translações e mudanças de escala. Sejam $a, b, c, d \in \mathbb{R}$ com $ac \neq 0$ e considere $aX + b$ e $cY + d$. Desenvolvendo a covariância com a forma (16),

$$\begin{aligned}
\operatorname{Cov}(aX + b,\, cY + d) &= \mathbb{E}[(aX + b)(cY + d)] - \mathbb{E}[aX + b]\,\mathbb{E}[cY + d] \\
&= ac\,\mathbb{E}[XY] + ad\,\mathbb{E}[X] + bc\,\mathbb{E}[Y] + bd - (a\mathbb{E}[X] + b)(c\mathbb{E}[Y] + d) \\
&= ac\operatorname{Cov}(X, Y),
\end{aligned}$$

pois os termos constantes e os termos lineares se cancelam. Para as variâncias, o mesmo desenvolvimento mostra que uma translação não altera a dispersão e que um fator multiplicativo sai ao quadrado:

$$\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X), \qquad \operatorname{Var}(cY + d) = c^2 \operatorname{Var}(Y).$$

Substituindo em (20),

$$\rho(aX+b,\,cY+d) = \frac{ac\,\operatorname{Cov}(X,Y)}{\sqrt{a^2c^2\,\operatorname{Var}(X)\,\operatorname{Var}(Y)}} = \operatorname{sgn}(ac)\,\rho(X,Y). \tag{21}$$

Em particular, se $ac > 0$, como em qualquer mudança de unidades aplicada às duas coordenadas (metros e centímetros, quilos e gramas), temos $\rho(aX+b, cY+d) = \rho(X,Y)$. Portanto, a correlação é invariante por mudanças de escala com fatores positivos. Se $ac < 0$, uma das coordenadas é invertida e o sinal de $\rho$ se inverte, o que é coerente com a interpretação de $\rho$ como medida da tendência de os desvios terem o mesmo sinal.

A correlação tem ainda outra propriedade atraente como medida de dependência: seus valores estão confinados ao intervalo $[-1,1]$, e os casos extremos admitem interpretação precisa em termos do comportamento conjunto de $X$ e $Y$.

**Teorema 14** (Faixa de variação do coeficiente de correlação)**.** Se $X$ e $Y$ são variáveis aleatórias e $\rho(X,Y)$ existe, então

$$-1 \leqslant \rho(X,Y) \leqslant 1. \tag{22}$$

A prova é uma aplicação direta da desigualdade de Cauchy–Schwarz, que apresentaremos na próxima subseção; por isso a demonstração fica adiada para lá.

## 2.7. A Desigualdade de Cauchy–Schwarz

Nas notas sobre funções de variáveis aleatórias contínuas, provamos uma desigualdade de Cauchy–Schwarz integrando $(g-\lambda h)^2$ em relação a uma densidade. O argumento não depende de representação alguma da distribuição e vale para variáveis aleatórias arbitrárias; é o que vamos registrar agora. A desigualdade controla a esperança do produto $UV$ pelos segundos momentos de $U$ e $V$, e não exige independência. A prova abaixo usa apenas a positividade e a linearidade da esperança, propriedades cuja validade foi discutida no caso geral na Seção 1.

**Teorema 15** (Desigualdade de Cauchy–Schwarz)**.** Sejam $U$ e $V$ variáveis aleatórias com segundos momentos finitos. Então $UV$ é integrável e

$$[\mathbb{E}[UV]]^2 \leqslant \mathbb{E}[U^2]\,\mathbb{E}[V^2]. \tag{23}$$

**Prova**. A integrabilidade do produto segue de $|UV| \leqslant (U^2+V^2)/2$. A ideia é produzir uma função quadrática em uma variável real $s$ que não assume valores negativos e extrair uma restrição sobre seus coeficientes. Para isso, fixe $s \in \mathbb{R}$ e defina a variável aleatória

$$W = sU + V.$$

Como $W^2 \geqslant 0$ em cada ponto de $\Omega$, vale $\mathbb{E}[W^2] \geqslant 0$ e, pela linearidade da esperança,

$$0 \leqslant \mathbb{E}[W^2] = \mathbb{E}[s^2U^2 + 2sUV + V^2] = as^2 + bs + c, \tag{24}$$

com coeficientes

$$a = \mathbb{E}[U^2], \qquad b = 2\mathbb{E}[UV], \qquad c = \mathbb{E}[V^2].$$

Claramente $a \geqslant 0$, pois $U^2 \geqslant 0$. Podemos supor $a > 0$: se $a = \mathbb{E}[U^2] = 0$, então $U = 0$ quase certamente (resultado da subseção anterior), de modo que $\mathbb{E}[UV] = 0$ e (23) vale como igualdade.

Considere então a função quadrática $g(s) = as^2 + bs + c$ com $a > 0$. Por (24), temos $g(s) \geqslant 0$ para todo $s \in \mathbb{R}$. Afirmamos que $g$ tem no máximo uma raiz real. De fato, se $g(s_1) = g(s_2) = 0$ com $s_1 \neq s_2$, então, como $a > 0$, vale $g(s) = a(s - s_1)(s - s_2)$, e para $s$ estritamente entre $s_1$ e $s_2$ os dois fatores têm sinais opostos, donde $g(s) < 0$, em contradição com (24). Pela fórmula quadrática, as raízes reais de $g$, quando existem, são

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a},$$

de modo que a existência de duas raízes reais distintas equivale a $b^2 - 4ac > 0$. Portanto $b^2 - 4ac \leqslant 0$. Além disso, se o discriminante $b^2 - 4ac = 0$, então o gráfico de $g$ toca o eixo horizontal em uma raiz dupla. Substituindo os coeficientes, concluímos que

$$0 \geqslant b^2 - 4ac = 4[\mathbb{E}[UV]]^2 - 4\,\mathbb{E}[U^2]\,\mathbb{E}[V^2],$$

e, dividindo por 4, obtemos (23). ■

Vamos agora completar o resultado anunciado na subseção anterior. Sejam $X$ e $Y$ variáveis aleatórias com $\operatorname{Var}(X) > 0$ e $\operatorname{Var}(Y) > 0$, de modo que $\rho(X, Y)$ está definido. Tomando $U = X - \mathbb{E}[X]$ e $V = Y - \mathbb{E}[Y]$ em Teorema 15, notamos que

$$\mathbb{E}[U^2] = \operatorname{Var}(X), \qquad \mathbb{E}[V^2] = \operatorname{Var}(Y), \qquad \mathbb{E}[UV] = \operatorname{Cov}(X, Y),$$

e, portanto,

$$\operatorname{Cov}(X, Y)^2 \leqslant \operatorname{Var}(X)\,\operatorname{Var}(Y).$$

Dividindo por $\operatorname{Var}(X)\operatorname{Var}(Y) > 0$ e usando a definição $\rho(X, Y) = \operatorname{Cov}(X, Y)/(\sigma_X \sigma_Y)$, concluímos que $\rho(X, Y)^2 \leqslant 1$, isto é,

$$-1 \leqslant \rho(X, Y) \leqslant 1.$$

## 2.8. A Caracterização dos Casos Extremos

Os valores $\rho(X, Y) = \pm 1$ só ocorrem em circunstâncias especiais. Para identificá-las, vamos reexaminar a prova de Teorema 15 com as escolhas $U = X - \mathbb{E}[X]$ e $V = Y - \mathbb{E}[Y]$ feitas acima; lembramos que $\operatorname{Var}(X) > 0$ e $\operatorname{Var}(Y) > 0$, pois fora disso $\rho(X, Y)$ não está definido. O polinômio quadrático (24) torna-se $g(s) = as^2 + bs + c$, com

$$a = \operatorname{Var}(X), \qquad b = 2\operatorname{Cov}(X, Y), \qquad c = \operatorname{Var}(Y),$$

e, para cada $s \in \mathbb{R}$,

$$g(s) = \mathbb{E}\Big[(s(X - \mathbb{E}[X]) + (Y - \mathbb{E}[Y]))^2\Big] \geqslant 0.$$

Suponha $\rho(X, Y) = \pm 1$. Então $ac = \operatorname{Var}(X)\operatorname{Var}(Y) > 0$ e o discriminante se anula:

$$b^2 - 4ac = 4\operatorname{Cov}(X, Y)^2 - 4\operatorname{Var}(X)\operatorname{Var}(Y) = 4\operatorname{Var}(X)\operatorname{Var}(Y)[\rho(X, Y)^2 - 1] = 0.$$

Como $a > 0$ e $b^2 - 4ac = 0$, a equação $g(s) = 0$ possui uma raiz dupla $s = s_0$, com $s_0 = -b/(2a)$. A variável aleatória

$$W = s_0(X - \mathbb{E}[X]) + (Y - \mathbb{E}[Y])$$

satisfaz $\mathbb{E}[W^2] = g(s_0) = 0$. Além disso, $\mathbb{E}[W] = 0$, de modo que $\mathrm{Var}(W) = \mathbb{E}[W^2] = 0$ e, por (10), concluímos que $\mathbb{P}(W = 0) = 1$. Isso significa que

$$\mathbb{P}(Y = \alpha X + \beta) = 1, \qquad \text{com} \qquad \alpha = -s_0 = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} \quad \text{e} \quad \beta = \mathbb{E}[Y] - \alpha\,\mathbb{E}[X].$$

O sinal de $\alpha$ é o sinal de $\rho(X,Y)$: de fato, como $\mathrm{Var}(Y) = \alpha^2\,\mathrm{Var}(X)$, vale $\sigma_Y = |\alpha|\sigma_X$ e, portanto,

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y} = \frac{\alpha\,\mathrm{Var}(X)}{\sigma_X\,|\alpha|\sigma_X} = \frac{\alpha}{|\alpha|}.$$

Assim, $\alpha > 0$ quando $\rho(X,Y) = 1$ e $\alpha < 0$ quando $\rho(X,Y) = -1$. Reciprocamente, se $\mathbb{P}(Y = \alpha X + \beta) = 1$ com $\alpha \neq 0$, então $\mathrm{Cov}(X,Y) = \alpha\,\mathrm{Var}(X)$ e $\mathrm{Var}(Y) = \alpha^2\,\mathrm{Var}(X)$, e o cálculo acima dá $\rho(X,Y) = \alpha/|\alpha|$. Ficam assim estabelecidas as duas equivalências

$$\rho(X,Y) = 1 \quad \text{se, e somente se,} \quad \mathbb{P}(Y = \alpha X + \beta) = 1 \text{ para algum } \alpha > 0 \text{ e } \beta \in \mathbb{R}, \tag{25}$$

$$\rho(X,Y) = -1 \quad \text{se, e somente se,} \quad \mathbb{P}(Y = \alpha X + \beta) = 1 \text{ para algum } \alpha < 0 \text{ e } \beta \in \mathbb{R}. \tag{26}$$

Recapitulando, quando $\mathrm{Var}(X) > 0$ e $\mathrm{Var}(Y) > 0$, o coeficiente $\rho(X,Y)$ pertence ao intervalo $[-1,1]$ e pode ser lido como uma medida da dependência entre $X$ e $Y$ à luz dos valores $-1$, 0 e 1:

- se $X$ e $Y$ são independentes, então $\rho(X,Y) = 0$;
- $\rho(X,Y) = 1$ se, e somente se, $Y$ é função linear crescente de $X$;
- $\rho(X,Y) = -1$ se, e somente se, $Y$ é função linear decrescente de $X$.

Nos dois últimos casos, a relação linear vale quase certamente, como estabelecem (25) e (26).

**Observação 16** (Variáveis não correlacionadas)**.** Dizemos que $X$ e $Y$ são "não correlacionadas" quando $\rho(X,Y) = 0$, o que equivale a $\mathrm{Cov}(X,Y) = 0$. A independência implica não correlação, pois $\mathrm{Cov}(X,Y) = 0$ quando $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, mas a recíproca é falsa: como mostra Exemplo 12, é possível que $\mathrm{Cov}(X,Y) = 0$ sem que $X$ e $Y$ sejam independentes. A não correlação registra apenas a ausência de relação linear entre as variáveis, enquanto a independência controla toda forma de dependência.

# 3. Funções Geradoras de Momentos

## 3.1. Definição e Exemplos

Nas notas sobre funções geradoras de probabilidade estudamos, em detalhe, a ferramenta que associa a uma variável aleatória $X$ com valores em $\{0,1,2,\ldots\}$ a função

$$G_X(s) = \mathbb{E}\left[s^X\right] = \sum_{k=0}^{\infty} s^k\,\mathbb{P}(X = k), \tag{27}$$

definida para os $s$ reais para os quais a série converge. Lá vimos que $G_X$ determina a distribuição de $X$, que suas derivadas em $s = 1$ produzem os momentos fatoriais de $X$ e que ela se comporta bem sob somas de variáveis independentes. Recordamos aqui apenas a definição, pois vamos precisar dela para comparar com o objeto novo desta seção.

A limitação de (27) é imediata: a expressão só faz sentido quando $X$ toma valores inteiros não negativos. Se $X$ é, por exemplo, uma variável com densidade ou uma variável discreta que assume valores negativos, não há uma maneira canônica de definir $s^X$ e portanto $G_X(s)$ poderia não estar bem-definida. Este problema ocorre, por exemplo, no caso $s = -1$ e $X = 1/2$.

Para trabalhar com uma variável aleatória $X$ que possa assumir quaisquer valores reais, vamos considerar apenas bases positivas: assim, $s^X$ está definida para todos os valores de $X$. É conveniente escrever $s = e^t$, com $t \in \mathbb{R}$, pois então $s^X = (e^t)^X = e^{tX}$. Isso nos leva a estudar a esperança $\mathbb{E}[e^{tX}]$ como função do parâmetro real $t$. Para cada $t$ fixado, $e^{tX}$ é uma variável aleatória positiva e portanto, sua esperança está bem-definida por (1), para qualquer que seja a distribuição de $X$. Essa esperança, porém, pode ser infinita. Na definição a seguir, consideramos apenas os valores de $t$ para os quais ela é finita.

**Definição 17** (Função geradora de momentos)**.** A *função geradora de momentos* da variável aleatória $X$ é a função $M_X$ definida por

$$M_X(t) = \mathbb{E}\Big[e^{tX}\Big], \tag{28}$$

para todo $t \in \mathbb{R}$ para o qual essa esperança existe e é finita.

Precisamos ser explícitos quanto ao domínio. Em (28) o símbolo $M_X(t)$ só está definido quando $\mathbb{E}\Big[e^{tX}\Big]$ converge; quando não converge, dizemos que a função geradora de momentos não existe naquele $t$. Além disso, como $e^{tX} > 0$, a esperança em (28) ou é finita ou é $+\infty$, e não há ambiguidade de sinal. Assim, o domínio de $M_X$ é um subconjunto de $\mathbb{R}$ que contém $t = 0$, pois $M_X(0) = \mathbb{E}[1] = 1$.

Para variáveis que tomam valores em $\{0, 1, 2, \ldots\}$, a definição nova é uma modificação da anterior. De fato, tomando $s = e^t$ em (27) obtemos

$$M_X(t) = \mathbb{E}\Big[e^{tX}\Big] = \mathbb{E}\Big[(e^t)^X\Big] = G_X(e^t), \tag{29}$$

desde que a série em (27) convirja em $s = e^t$, o que ocorre em particular para todo $t \leqslant 0$. Quando o raio de convergência da série é estritamente maior do que 1, a identidade vale também para uma faixa de valores positivos de $t$: no caso da distribuição de Poisson, por exemplo, a série converge para todo $s \in \mathbb{R}$ e a identidade vale em toda a reta.

Nos casos discreto e absolutamente contínuo, a definição se lê como

$$M_X(t) = \mathbb{E}\Big[e^{tX}\Big] = \begin{cases} \displaystyle\sum_x e^{tx}\, \mathbb{P}(X = x), & \text{se } X \text{ é discreta,} \\[2ex] \displaystyle\int_{-\infty}^{\infty} e^{tx}\, f_X(x)\, dx, & \text{se } X \text{ é absolutamente contínua,} \end{cases} \tag{30}$$

quando a soma ou a integral é finita. Como os termos e o integrando são não negativos, essa finitude equivale à convergência absoluta. As duas expressões são ferramentas de cálculo para essas classes, e não uma restrição da Definição 17; para outras distribuições, usamos (1) com $Y = e^{tX}$. Como veremos nos exemplos seguintes, a existência de $M_X(t)$ para $t \neq 0$ pode falhar, e falha com frequência suficiente para merecer atenção.

Começamos pelo caso em que a função existe em toda a reta.

**Exemplo 18** (A geradora da normal padrão)**.** Seja $X \sim \mathrm{N}(0,1)$. Vamos calcular $\mathbb{E}\left[e^{tX}\right]$ para $t \in \mathbb{R}$ fixado. Como o integrando é não negativo, a integral imprópria está bem definida, ainda que possa valer $+\infty$; vamos calcular e verificar que o valor é finito.

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx}\, \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(x^2 - 2tx\right)}\, dx. \end{aligned}$$

A ideia é completar o quadrado no expoente, escrevendo $x^2 - 2tx = (x-t)^2 - t^2$. Isso separa do integrando um fator que não depende de $x$ e deixa no lugar da exponencial a densidade de uma normal de média $t$ e variância 1. Obtemos

$$\begin{aligned} M_X(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2 + \frac{1}{2}t^2}\, dx \\ &= e^{\frac{1}{2}t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(x-t)^2}\, dx. \end{aligned} \tag{31}$$

A integral que restou é a integral, sobre toda a reta, da densidade de uma variável normal com média $t$ e variância 1. Por isso ela vale 1, e concluímos que

$$M_X(t) = e^{\frac{1}{2}t^2}, \qquad t \in \mathbb{R}.$$

Em particular, a função geradora de momentos da normal padrão existe para todo $t$ real, o que é o melhor comportamento possível.

**Exemplo 19** (A geradora da exponencial)**.** Seja $X \sim \mathrm{Exp}(\lambda)$, com densidade $f_X(x) = \lambda e^{-\lambda x}$ para $x > 0$ e $f_X(x) = 0$ para $x \leqslant 0$. Como $X > 0$ quase certamente, $e^{tX}$ é decrescente em $X$ quando $t < 0$, e a esperança deve ser finita nesse caso; para $t \geqslant 0$ a exponencial cresce e precisamos verificar diretamente. Escrevemos, usando $f_X(x) = 0$ para $x \leqslant 0$,

$$M_X(t) = \int_0^{\infty} e^{tx}\, \lambda e^{-\lambda x}\, dx = \lambda \int_0^{\infty} e^{-(\lambda - t)x}\, dx.$$

O integrando é positivo, logo a integral imprópria tem valor bem definido em $[0, +\infty]$, e o cálculo é uma integral exponencial elementar.

Se $t < \lambda$, o expoente $-(\lambda - t)x$ é negativo para $x > 0$ e a integral converge, valendo $1/(\lambda - t)$. Se $t \geqslant \lambda$, o expoente é não negativo; quando $t = \lambda$ o integrando é a constante $\lambda$, e quando $t > \lambda$ ele cresce exponencialmente, de modo que em ambos os casos a integral diverge para $+\infty$. Resumindo,

$$M_X(t) = \begin{cases} \dfrac{\lambda}{\lambda - t}, & \text{se } t < \lambda, \\[2ex] +\infty, & \text{se } t \geqslant \lambda. \end{cases}$$

Portanto, $M_X(t)$ existe apenas para $t < \lambda$, ou seja, na semirreta aberta $(-\infty, \lambda)$ que contém a origem. O domínio é assimétrico: não há simetria entre valores positivos e negativos de $t$, o que reflete a assimetria da própria distribuição exponencial.

**Exemplo 20** (Uma variável sem geradora de momentos)**.** Seja $X$ com densidade $f_X(x) = 1/[\pi(1+x^2)]$, isto é, $X$ com distribuição de Cauchy. Para $t = 0$ o integrando é a própria densidade e temos $M_X(0) = 1$. Vamos mostrar que, para todo $t \neq 0$, a esperança é infinita.

Suponha primeiro $t > 0$. Então $e^{tx} \to +\infty$ quando $x \to +\infty$ e $e^{tx} \to 0$ quando $x \to -\infty$. Para decidir a convergência, examinamos o comportamento do integrando

$$g_t(x) = \frac{e^{tx}}{\pi(1+x^2)}$$

quando $x \to +\infty$. Para obter uma cota inferior, observamos que $1 + x^2 \leqslant 2x^2$ para $x \geqslant 1$, de modo que

$$g_t(x) = \frac{e^{tx}}{\pi(1+x^2)} \geqslant \frac{e^{tx}}{2\pi x^2}, \qquad x \geqslant 1.$$

A função $x \mapsto e^{tx}/x^2$ tende a $+\infty$, pois a exponencial cresce mais rápido do que qualquer polinômio; logo existe $x_0 \geqslant 1$ tal que $e^{tx}/x^2 \geqslant 1$ para todo $x \geqslant x_0$, e portanto $g_t(x) \geqslant 1/(2\pi)$ nesse intervalo. A integral de $g_t$ sobre $[x_0, \infty)$ diverge para $+\infty$, e o mesmo ocorre com $\int_{-\infty}^{\infty} g_t$.

Suponha agora $t < 0$. Desta vez $e^{tx} \to +\infty$ quando $x \to -\infty$. No extremo $x \to -\infty$ escrevemos $u = -x \to +\infty$, e a mesma comparação se aplica: a densidade se comporta como $1/(\pi u^2)$ e $e^{tx} = e^{|t|u}$, de modo que o integrando é limitado inferiormente por uma constante positiva a partir de algum ponto e a integral diverge para $+\infty$. Reunindo os dois casos,

$$M_X(t) = \begin{cases} 1, & \text{se } t = 0, \\ +\infty, & \text{se } t \neq 0. \end{cases}$$

Assim, $M_X$ existe apenas em $t = 0$. Note ainda que a esperança de $X$ não existe, o que já sugere que a distribuição de Cauchy não pode ter uma geradora de momentos útil.

O último exemplo mostra que a dificuldade com a existência de $\mathbb{E}\left[e^{tX}\right]$ é intrínseca e não se resolve com uma escolha melhor de parametrização. Há uma forma de contornar este problema. Isto envolve estudar a função característica $\phi_X(t) = \mathbb{E}[e^{itX}]$, onde $i$ é a unidade imaginária no plano complexo. Neste caso, segue da fórmula de Euler que $\phi_X(t)$ existe para todo $t \in \mathbb{R}$, independentemente de qual é a distribuição de $X$, porque $|e^{itX}| = 1$. Esse caminho, porém, depende da teoria de funções de uma variável complexa. Vamos voltar a este contexto futuramente.

Vamos voltar nossas atenções para os casos em que $M_X(t)$ existe para todo $t$ em uma vizinhança $(-\delta, \delta)$ da origem, para algum $\delta > 0$. Essa condição é a que permite derivar $M_X$ sucessivamente em $t = 0$ e recuperar os momentos de $X$ e também consiste na hipótese do teorema de unicidade que garante que a função geradora determina a distribuição. No que segue sempre que usarmos uma função geradora de momentos, em valores de $t$ diferentes de zero em vizinhanças da origem, estamos admitindo que todos os momentos existem e que a série que define a função geradora de momentos é absolutamente convergente nesta vizinhança da origem.

Resta explicar por que $M_X$ é chamada de função geradora de momentos. A justificativa se deve ao seguinte argumento heurístico, baseado na expansão em série de potências da

função exponencial:

$$M_X(t) = \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left(1 + tX + \frac{1}{2!}(tX)^2 + \cdots\right)$$

$$= 1 + t\,\mathbb{E}[X] + \frac{1}{2!}\,t^2\,\mathbb{E}\left[X^2\right] + \cdots . \tag{32}$$

Os coeficientes da série de potências que aparece no lado direito da igualdade acima são, a menos dos fatoriais, os momentos de $X$. Mais precisamente, o termo de ordem $n$ em $t$ é dado por

$$\frac{1}{n!}\mathbb{E}[X^n]\,t^n.$$

Devemos ser cuidadosos quanto ao status de (32) neste ponto. Sem hipóteses adicionais o cálculo acima é apenas formal. Trocamos a esperança pela soma infinita e permutamos a soma sobre $n$ com a soma ou a integral que define a esperança, e nenhuma dessas passagens foi justificada. Elas exigem, entre outras coisas, que $\mathbb{E}\left[e^{tX}\right]$ exista em uma vizinhança da origem, o que já sabemos não ser verdadeiro em geral. A justificativa rigorosa será dada pelo teorema de unicidade das funções geradoras de momentos, mais adiante, do qual segue também que derivar (32) termo a termo produz os momentos: $\mathbb{E}[X^n] = M_X^{(n)}(0)$. Por ora, guardamos (32) como a motivação do nome e como um guia para o que esperamos que seja verdadeiro.

## 3.2. Momentos, Transformações Lineares e Somas

Na subseção anterior, vimos como usar a função geradora de momentos como uma forma de codificar a distribuição de uma variável aleatória. Mostraremos agora como tais funções podem ser usadas para: (i) obter os momentos de uma variável aleatória por meio de suas derivadas em $t = 0$; (ii) descrever o comportamento sob transformações afins; e (iii) calcular a distribuição da soma de variáveis aleatórias independentes.

Antes de enunciar o primeiro resultado, precisamos deixar claro o que significa existir. Como $e^{tX} > 0$, a esperança $\mathbb{E}[e^{tX}]$ está sempre bem definida, eventualmente valendo $+\infty$. Vamos dizer então que $M_X(t)$ *existe* em $t$ quando essa esperança for finita, isto é, quando $M_X(t) < \infty$. Analogamente, vamos dizer que $M_X(t)$ existe numa vizinhança de 0, quando existe algum $\delta > 0$ tal que $M_X(t) < \infty$ para todo $t \in (-\delta, \delta)$. Essa é a hipótese que garante a existência de todos os momentos de $X$.

**Teorema 21.** Suponha que $M_X(t)$ exista numa vizinhança de 0. Então $\mathbb{E}[|X|^k]$ é finita para todo $k \geqslant 1$ e

$$\mathbb{E}\left[X^k\right] = M_X^{(k)}(0), \tag{33}$$

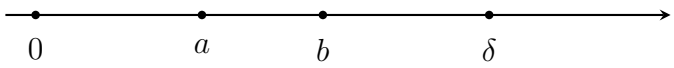
onde $M_X^{(k)}(0)$ denota a $k$-ésima derivada de $M_X$ calculada em $t = 0$.

*Ideia da prova.* O argumento vale para qualquer distribuição e geralmente é provado usando um resultado de convergência que é uma consequência do Teorema da Convergência Dominada cujo enunciado é o seguinte. Se $Z_h \to Z$, quando $h \to 0$, quase certamente e $|Z_h| \leqslant H$, para alguma variável aleatória integrável $H$, então $\mathbb{E}[Z_h] \to \mathbb{E}[Z]$. No caso discreto, este resultado justifica a troca do limite com o somatório. Já no caso absolutamente contínuo,

ele permite permutar a ordem do limite e de integração. Sua versão para distribuições gerais é o passo adicional que vamos assumir sem prova.

Vamos construir os majorantes que permitirão aplicar esse resultado. Ao derivar $e^{sX}$ repetidamente em relação a $s$, aparecem expressões da forma $X^j e^{sX}$. Precisamos limitar seus módulos por uma variável aleatória com esperança finita, que seja a mesma para todos os valores de $s$ em um pequeno intervalo em torno de 0.

Por hipótese, existe $\delta > 0$ tal que $M_X(t) < \infty$ sempre que $|t| < \delta$. Escolhemos dois números $a$ e $b$ com $0 < a < b < \delta$.




Vamos considerar $s$ no intervalo $[-a, a]$




e usar as esperanças finitas $M_X(b)$ e $M_X(-b)$ para construir o majorante. A escolha $b > a$ deixará uma margem no expoente para controlar também o fator $|X|^j$.

Primeiro, se $|s| \leqslant a$, então, para qualquer valor real $x$,

$$sx \leqslant |s|\,|x| \leqslant a|x|.$$

Como a função exponencial é crescente, segue que

$$|x|^j e^{sx} \leqslant |x|^j e^{a|x|}.$$

O lado direito já não depende de $s$. Falta mostrar que, ao substituir $x$ por $X$, obtemos uma variável aleatória com esperança finita.

Para controlar o fator $|x|^j$, vamos usar a diferença positiva $b - a$. Fixe um inteiro $j \geqslant 0$. Para $r \geqslant 0$, todos os termos da série da exponencial são não negativos; em particular, a soma é maior ou igual ao termo de ordem $j$:

$$e^{(b-a)r} = \sum_{n=0}^{\infty} \frac{((b-a)r)^n}{n!} \geqslant \frac{(b-a)^j r^j}{j!}.$$

Isolando $r^j$ e depois multiplicando por $e^{ar}$, obtemos

$$r^j \leqslant \frac{j!}{(b-a)^j} e^{(b-a)r}, \qquad r^j e^{ar} \leqslant \frac{j!}{(b-a)^j} e^{br}.$$

Assim, podemos tomar explicitamente $C_{j,a,b} = j!/(b-a)^j$. Essa constante é finita porque $b - a > 0$. Aplicando a desigualdade com $r = |x|$, chegamos à estimativa

$$|x|^j e^{a|x|} \leqslant C_{j,a,b} e^{b|x|}.$$

Agora relacionamos $e^{b|x|}$ às duas exponenciais cujas esperanças conhecemos. Se $x \geqslant 0$, então $e^{b|x|} = e^{bx}$; se $x < 0$, então $e^{b|x|} = e^{-bx}$. Em ambos os casos,

$$e^{b|x|} \leqslant e^{bx} + e^{-bx}.$$

Reunindo as desigualdades e substituindo $x$ por $X$, obtemos

$$\left|X^j e^{sX}\right| = |X|^j e^{sX} \leqslant |X|^j e^{a|X|} \leqslant C_{j,a,b}\Big(e^{bX} + e^{-bX}\Big), \qquad |s| \leqslant a.$$

Portanto, o majorante procurado é

$$H_j = C_{j,a,b}\Big(e^{bX} + e^{-bX}\Big). \tag{34}$$

Note que esta cota superior não depende de $s$ e tem esperança finita, pois

$$\mathbb{E}[H_j] = C_{j,a,b}\Big(\mathbb{E}[e^{bX}] + \mathbb{E}[e^{-bX}]\Big) = C_{j,a,b}\Big(M_X(b) + M_X(-b)\Big) < \infty.$$

Na última passagem usamos $b < \delta$, de modo que tanto $b$ quanto $-b$ pertencem ao intervalo onde $M_X$ é finita.

Essa estimativa também já prova a existência dos momentos absolutos. De fato, tomando $s = 0$, temos $e^{sX} = 1$ e, portanto,

$$0 \leqslant \mathbb{E}[|X|^j] \leqslant \mathbb{E}[H_j] < \infty.$$

Como podemos fazer isso para cada inteiro $j \geqslant 1$, todos os momentos absolutos de $X$ são finitos.

Vamos agora calcular a primeira derivada de $M_X$. Fixe $t$ com $|t| < a$. Pela definição de derivada, precisamos estudar o quociente

$$\frac{M_X(t+h) - M_X(t)}{h}, \qquad h \neq 0,$$

quando $h \to 0$. Escolhemos $h$ de modo que $0 < |h| < a - |t|$. Essa escolha é possível porque $|t| < a$ e garante que

$$|t + h| \leqslant |t| + |h| < a.$$

Assim, tanto $t$ quanto $t+h$ pertencem a $(-a, a)$, onde $M_X$ é finita. Podemos, portanto, usar a definição de $M_X$ e a linearidade da esperança:

$$\begin{aligned}\frac{M_X(t+h) - M_X(t)}{h} &= \frac{\mathbb{E}[e^{(t+h)X}] - \mathbb{E}[e^{tX}]}{h} \\ &= \mathbb{E}\left[\frac{e^{(t+h)X} - e^{tX}}{h}\right].\end{aligned}$$

Para aplicar o resultado de convergência enunciado no início da prova, vamos usar a mesma notação e definir

$$Z_h = \frac{e^{(t+h)X} - e^{tX}}{h}.$$

Para cada valor real $x$ de $X$, o quociente $(e^{(t+h)x} - e^{tx})/h$ converge para $xe^{tx}$, pois essa é a derivada da função $u \mapsto e^{ux}$ em $u = t$. Logo,

$$Z_h \longrightarrow Xe^{tX} \qquad \text{quando } h \to 0.$$

Queremos concluir que as esperanças também convergem. Para aplicar a convergência dominada, falta encontrar um majorante integrável para $|Z_h|$ que não dependa de $h$.

É aqui que usamos o teorema do valor médio. Fixado $x \in \mathbb{R}$, aplicamos esse teorema à função $u \mapsto e^{ux}$ entre $t$ e $t+h$. Existe um número $\xi$ entre esses dois pontos tal que

$$\frac{e^{(t+h)x} - e^{tx}}{h} = xe^{\xi x}.$$

O número $\xi$ pode depender de $x$ e de $h$, mas o que precisamos saber é apenas que $|\xi| < a$, pois os dois extremos estão em $(-a, a)$. Consequentemente, $\xi x \leqslant |\xi|\,|x| \leqslant a|x|$ e

$$\left|\frac{e^{(t+h)x} - e^{tx}}{h}\right| = |x|e^{\xi x} \leqslant |x|e^{a|x|}.$$

Como essa desigualdade vale para todo $x$, podemos substituir $x$ por $X$:

$$|Z_h| \leqslant |X|e^{a|X|} \leqslant H_1. \tag{35}$$

O majorante $H_1$ é o que construímos acima para $j = 1$: não depende de $h$ e satisfaz $\mathbb{E}[H_1] < \infty$. Temos, assim, as duas condições necessárias: $Z_h \to Xe^{tX}$ e $|Z_h| \leqslant H_1$. Na notação do resultado inicial, o limite é $Z = Xe^{tX}$ e o majorante é $H = H_1$. Pela consequência do Teorema da Convergência Dominada, podemos afirmar que

$$\lim_{h\to 0} \mathbb{E}[Z_h] = \mathbb{E}[Xe^{tX}].$$

Voltando ao quociente que define a derivada, obtemos

$$M_X'(t) = \lim_{h\to 0} \frac{M_X(t+h) - M_X(t)}{h} = \lim_{h\to 0} \mathbb{E}[Z_h] = \mathbb{E}[Xe^{tX}], \qquad |t| < a.$$

Para obter as derivadas seguintes, vamos explicitar como esse argumento se repete. Para cada inteiro $j \geqslant 0$, defina

$$F_j(t) = \mathbb{E}[X^j e^{tX}], \qquad |t| < a.$$

Essas esperanças são finitas, pois já mostramos que $|X^j e^{tX}| \leqslant H_j$ e $\mathbb{E}[H_j] < \infty$. Em particular, $F_0 = M_X$, e o cálculo anterior mostrou que $F_0' = F_1$.

Fixe agora $j \geqslant 0$ e considere novamente $0 < |h| < a - |t|$. Pela linearidade,

$$\frac{F_j(t+h) - F_j(t)}{h} = \mathbb{E}\left[X^j \frac{e^{(t+h)X} - e^{tX}}{h}\right]$$

$$= \mathbb{E}[X^j Z_h].$$

Dentro da esperança, temos a convergência

$$X^j Z_h \longrightarrow X^{j+1} e^{tX}.$$

Além disso, multiplicando a primeira desigualdade de (35) por $|X|^j$ e usando o majorante $H_{j+1}$ definido em (34), com $j$ substituído por $j+1$, obtemos

$$|X^j Z_h| \leqslant |X|^{j+1} e^{a|X|} \leqslant H_{j+1}.$$

Como $H_{j+1}$ também é integrável e não depende de $h$, podemos aplicar novamente a convergência dominada. Isso fornece

$$F_j'(t) = \mathbb{E}[X^{j+1} e^{tX}] = F_{j+1}(t).$$

Portanto, derivar $F_j$ produz $F_{j+1}$. Começando com $F_0 = M_X$, concluímos sucessivamente que $M_X' = F_1$, $M_X'' = F_2$ e, por indução,

$$M_X^{(k)}(t) = F_k(t) = \mathbb{E}[X^k e^{tX}], \qquad |t| < a.$$

Fazendo $t = 0$ na identidade acima, concluímos que

$$M_X^{(k)}(0) = \mathbb{E}[X^k e^{0X}] = \mathbb{E}[X^k].$$

■

Com o teorema acima, podemos pensar que os momentos deixam de ser quantidades que só podem ser obtidas por meio de somas ou integrais e passam a ser quantidades que ficam codificadas nas derivadas da função $M_X$ em 0.

Vamos ver a seguir como a geradora se comporta sob duas operações que aparecem o tempo todo em probabilidade.

**Transformações lineares.** Seja $X$ uma variável aleatória e considere $aX+b$, com $a, b \in \mathbb{R}$. Fixado $t$, o expoente na definição da função geradora se decompõe em um fator que depende de $X$ e um fator constante, e podemos usar a linearidade da esperança no fator constante para obter a seguinte igualdade

$$\begin{aligned} M_{aX+b}(t) &= \mathbb{E}\Big[e^{t(aX+b)}\Big] = \mathbb{E}\Big[e^{atX}e^{tb}\Big] \\ &= e^{tb}\,\mathbb{E}\Big[e^{(at)X}\Big] = e^{tb}M_X(at), \end{aligned}$$

sempre que $M_{aX+b}(t)$ seja finita, o que ocorre precisamente quando $M_X(at)$ é finita. Desta forma concluímos que

$$M_{aX+b}(t) = e^{tb}M_X(at). \tag{36}$$

Assim, uma mudança de escala em $X$ dilata o argumento de $M_X$, e uma translação multiplica a geradora por uma exponencial. Note que, se $a = 0$, a variável $aX + b$ é a constante $b$ e (36) fornece $e^{tb}M_X(0) = e^{tb}$, como esperado.

**Somas de variáveis independentes.** Boa parte da teoria da probabilidade gira em torno de somas de variáveis aleatórias independentes. Obter a distribuição de uma soma a partir das distribuições das parcelas costuma ser trabalhoso. Nas notas sobre somas de variáveis aleatórias contínuas, fizemos isso por meio de convoluções, identificando a densidade de $X+Y$ com a convolução das densidades de $X$ e de $Y$, cálculo que frequentemente envolve integrais bastante elaboradas. As funções geradoras de momentos são úteis aqui porque transformam somas em produtos, e produtos são fáceis de manipular.

**Teorema 22.** Sejam $X$ e $Y$ variáveis aleatórias independentes e suponha que $M_X(t)$ e $M_Y(t)$ sejam finitas. Então $M_{X+Y}(t)$ é finita e

$$M_{X+Y}(t) = M_X(t)\,M_Y(t). \tag{37}$$

**Prova**. Como $X$ e $Y$ são independentes, $e^{tX}$ e $e^{tY}$ também são, pois são funções mensuráveis de $X$ e de $Y$, respectivamente. Além disso, suas esperanças são finitas por hipótese. Podemos, portanto, usar a propriedade de que a esperança do produto de variáveis independentes

e integráveis é o produto das esperanças. Como $e^{t(X+Y)} = e^{tX}e^{tY}$, obtemos

$$\begin{aligned}\mathbb{E}\left[e^{t(X+Y)}\right] &= \mathbb{E}\left[e^{tX}e^{tY}\right] \\ &= \mathbb{E}\left[e^{tX}\right]\mathbb{E}\left[e^{tY}\right] \\ &= M_X(t)\,M_Y(t) < \infty.\end{aligned}$$

Assim, $M_{X+Y}(t)$ é finita e vale $M_X(t)M_Y(t)$, como queríamos provar. ■

O mesmo argumento se estende a qualquer número finito de parcelas independentes. Sejam $X_1, \ldots, X_n$ independentes e $S = X_1 + \cdots + X_n$. Provamos por indução sobre $n$ que

$$M_S(t) = M_{X_1}(t) \cdots M_{X_n}(t), \tag{38}$$

válida em todo $t$ em que todas as geradoras do lado direito sejam finitas. Para $n = 1$ a identidade é a definição de $M_{X_1}$. Suponha o resultado verdadeiro para $n-1$ e escreva $S = X_1 + \cdots + X_{n-1} + X_n$. Se $X_1, \ldots, X_n$ são independentes, então o vetor $(X_1, \ldots, X_{n-1})$ é independente de $X_n$ (é o que as notas sobre vetores aleatórios chamam de independência entre um vetor e uma variável). Em particular, a soma $S_{n-1} = X_1 + \cdots + X_{n-1}$ é independente de $X_n$, pois é função do vetor $(X_1, \ldots, X_{n-1})$. Podemos, portanto, aplicar o caso de duas variáveis, Teorema 22, com $X = S_{n-1}$ e $Y = X_n$, e depois a hipótese de indução:

$$M_S(t) = M_{S_{n-1}}(t)\,M_{X_n}(t) = M_{X_1}(t) \cdots M_{X_{n-1}}(t)\,M_{X_n}(t),$$

o que conclui a indução. A etapa de indução usa explicitamente o teorema para duas variáveis; é ele que faz o trabalho de transformar a soma de $n$ parcelas em um produto de $n$ fatores.

O ganho prático é o seguinte. Para identificar a distribuição de uma soma, podemos calcular o produto das geradoras, que é uma operação algébrica, e depois reconhecer a distribuição cuja geradora é aquele produto. Nada de convoluções: o preço a pagar está na última etapa, que exige saber ler a tabela de geradoras no sentido inverso. Voltaremos a esse ponto quando discutirmos a unicidade da função geradora de momentos.

## 3.3. O Teorema de Unicidade e a Soma de Duas Normais

O desenvolvimento formal da seção anterior produziu a expansão (32) da função geradora de momentos em série de potências. Para que esse cálculo signifique algo, precisamos saber que a função geradora de momentos determina a distribuição da variável aleatória. O resultado a seguir garante exatamente isso.

**Teorema 23** (Unicidade das funções geradoras de momentos)**.** Seja $X$ variável aleatória cuja função geradora de momentos $M_X(t) = \mathbb{E}[e^{tX}]$ satisfaz $M_X(t) < \infty$ para todo $t \in (-\delta, \delta)$, para algum $\delta > 0$. Então valem as duas conclusões a seguir. (i) Existe uma única distribuição com função geradora de momentos $M_X$. (ii) Todos os momentos $\mathbb{E}[X^k]$ são finitos, para $k = 1, 2, \ldots$, e vale a expansão em série de potências

$$M_X(t) = \sum_{k \geqslant 0} \frac{t^k}{k!}\,\mathbb{E}[X^k], \qquad |t| < \delta. \tag{39}$$

O teorema reúne duas afirmações: a função geradora determina a distribuição e, perto de 0, pode ser escrita como uma série cujos coeficientes são os momentos. Vamos explicar separadamente essas duas afirmações e sua relação com o teorema de unicidade dos momentos.

**Por que a geradora determina a distribuição?** A conclusão de unicidade significa o seguinte: se outra variável aleatória $Y$ satisfaz $M_Y(t) = M_X(t)$ para todo $t$ em algum intervalo aberto contendo 0, então $X$ e $Y$ têm a mesma distribuição. Não precisamos conhecer a geradora em toda a reta; seus valores nesse intervalo já bastam. Usaremos esse resultado sem demonstração, pois sua prova exige ferramentas de análise que não desenvolveremos nestas notas.

Uma conexão com as transformadas integrais ajuda a situar o resultado. Se $X$ tem densidade $f_X$, então

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx.$$

Essa expressão é uma transformada de Laplace *bilateral*, com troca de sinal no parâmetro. Mais precisamente, se adotamos a convenção

$$\mathcal{L}_{\text{bil}} f_X(s) = \int_{-\infty}^{\infty} e^{-sx} f_X(x)\, dx,$$

então $M_X(t) = \mathcal{L}_{\text{bil}} f_X(-t)$ nos pontos em que a integral é finita. A palavra "bilateral" indica que a integração é feita em toda a reta ao invés de apenas sobre o intervalo $[0, \infty)$. A unicidade da geradora é uma manifestação da propriedade de unicidade dessa transformada em um contexto apropriado.

**Por que a expansão em série é válida?** Aqui podemos justificar a passagem que antes fizemos formalmente. Fixe $t$ com $|t| < \delta$ e escolha $r$ tal que $|t| < r < \delta$. Como $r$ e $-r$ estão no intervalo onde a geradora é finita, temos

$$e^{r|X|} \leqslant e^{rX} + e^{-rX}, \qquad \mathbb{E}[e^{r|X|}] \leqslant M_X(r) + M_X(-r) < \infty.$$

Considere agora as somas parciais da série da exponencial,

$$S_n = \sum_{k=0}^{n} \frac{t^k X^k}{k!}.$$

Para cada valor de $X$, temos $S_n \to e^{tX}$. Além disso, pela desigualdade triangular e pela série da exponencial,

$$|S_n| \leqslant \sum_{k=0}^{n} \frac{|t|^k |X|^k}{k!} \leqslant e^{|t||X|} \leqslant e^{r|X|}.$$

Encontramos, assim, um majorante integrável que não depende de $n$. Pelo Teorema da Convergência Dominada, podemos justificar a troca de ordem do limite com a esperança, obtendo as seguintes igualdades:

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tX}] &= \lim_{n\to\infty} \mathbb{E}[S_n] \\ &= \lim_{n\to\infty} \sum_{k=0}^{n} \frac{t^k}{k!} \mathbb{E}[X^k] = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbb{E}[X^k]. \end{aligned}$$

Isso prova (39) e justifica a troca entre a esperança e a soma em (32).

**Qual é a relação com a unicidade dos momentos?** A hipótese de que $M_X$ é finita numa vizinhança de 0 equivale à hipótese de Teorema 5: todos os momentos existem e a série (7) converge absolutamente para algum parâmetro positivo. Vejamos as duas implicações.

Primeiro, suponha que $M_X$ seja finita em $(-\delta, \delta)$. Já provamos que todos os momentos absolutos são finitos. Para verificar a convergência absoluta da série, fixe $|t| < r < \delta$ e observe que

$$e^{r|X|} \geqslant \frac{r^k |X|^k}{k!}, \qquad \mathbb{E}[|X|^k] \leqslant k!\, r^{-k} \mathbb{E}[e^{r|X|}].$$

Como $|\mathbb{E}[X^k]| \leqslant \mathbb{E}[|X|^k]$, segue que

$$\sum_{k=0}^{\infty} \frac{|t|^k}{k!} \left|\mathbb{E}[X^k]\right| \leqslant \mathbb{E}[e^{r|X|}] \sum_{k=0}^{\infty} \left(\frac{|t|}{r}\right)^k$$

$$= \frac{\mathbb{E}[e^{r|X|}]}{1 - |t|/r} < \infty.$$

A série geométrica converge porque $|t| < r$. É importante usar um ponto $r$ estritamente menor que $\delta$: a hipótese não garante finitude da geradora nos extremos $\delta$ e $-\delta$.

Reciprocamente, suponha que todos os momentos existam e que a série (7) convirja absolutamente para algum $t_0 > 0$. Sua subsérie de índices pares também converge, de modo que

$$A = \sum_{j=0}^{\infty} \frac{t_0^{2j}}{(2j)!} \mathbb{E}[X^{2j}] < \infty.$$

Vamos mostrar que essa informação já controla uma exponencial de $|X|$. Somando as séries de $e^u$ e $e^{-u}$, os termos ímpares se cancelam e os termos pares se somam. Assim, para todo $x \in \mathbb{R}$,

$$e^{t_0|x|} \leqslant e^{t_0 x} + e^{-t_0 x} = 2 \sum_{j=0}^{\infty} \frac{t_0^{2j} x^{2j}}{(2j)!}.$$

Todos os termos da última série são não negativos. Por isso, podemos trocar a esperança e a soma pelo Teorema da Convergência Monótona, obtendo

$$\mathbb{E}[e^{t_0|X|}] \leqslant 2 \sum_{j=0}^{\infty} \frac{t_0^{2j}}{(2j)!} \mathbb{E}[X^{2j}] = 2A < \infty.$$

Finalmente, se $|s| \leqslant t_0$, então $sX \leqslant t_0|X|$ e, portanto,

$$M_X(s) = \mathbb{E}[e^{sX}] \leqslant \mathbb{E}[e^{t_0|X|}] < \infty.$$

Logo, a geradora é finita numa vizinhança de 0, como queríamos mostrar. As duas hipóteses expressam, assim, a mesma condição de controle dos momentos: podemos apresentá-la pela convergência da série ou pela finitude da função geradora perto de 0.

Vamos aplicar a parte de unicidade do teorema em um exemplo em que ela é indispensável.

**Exemplo 24** (Soma de duas normais independentes)**.** Sejam $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ e $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$ variáveis aleatórias independentes. Mostre que a variável aleatória $Z = X + Y$ possui distribuição $\mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

**Solução.** Para usar a propriedade da geradora de uma soma de variáveis independentes, precisamos primeiro conhecer a geradora de cada parcela. Vamos, portanto, calcular a função geradora de momentos de uma normal com parâmetros gerais. Seja $U \sim \mathrm{N}(\mu, \sigma^2)$, com $\mu \in \mathbb{R}$ e $\sigma > 0$, e fixe $t \in \mathbb{R}$. Pela expressão da densidade de $U$, temos que

$$\mathbb{E}[e^{tU}] = \int_{-\infty}^{+\infty} e^{tu} \, \frac{1}{\sigma\sqrt{2\pi}} \, \exp\Big(-\frac{(u-\mu)^2}{2\sigma^2}\Big) \, du.$$

Como o integrando é não negativo, essa integral está bem definida, embora ainda possa valer $+\infty$. As mudanças de variável abaixo são válidas também nessa situação; o próprio cálculo mostrará que a integral é finita.

Começamos pela mudança de variável que aparece na padronização de uma normal:

$$x = \frac{u-\mu}{\sigma}, \qquad u = \mu + \sigma x, \qquad du = \sigma \, dx.$$

Ela transforma $(u-\mu)^2/(2\sigma^2)$ em $x^2/2$. Além disso, $e^{tu} = e^{t(\mu+\sigma x)} = e^{\mu t} e^{\sigma t x}$ e, como $\sigma > 0$, os limites de integração continuam sendo $-\infty$ e $+\infty$. Substituindo essas expressões na integral, obtemos

$$\begin{aligned}
\mathbb{E}[e^{tU}] &= \int_{-\infty}^{+\infty} e^{\mu t} e^{\sigma t x} \, \frac{1}{\sigma\sqrt{2\pi}} \, e^{-x^2/2} \, \sigma \, dx \\
&= e^{\mu t} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \, \exp\Big(\sigma t x - \frac{x^2}{2}\Big) \, dx.
\end{aligned}$$

Na última passagem, cancelamos os fatores $\sigma$ e retiramos da integral o fator $e^{\mu t}$, que não depende de $x$.

Queremos agora reconhecer uma densidade normal na integral restante. Para isso, vamos escrever o expoente como um quadrado com sinal negativo, mais um termo que não depende de $x$. A identidade

$$(x - \sigma t)^2 = x^2 - 2\sigma t x + \sigma^2 t^2$$

mostra qual termo precisamos acrescentar e subtrair:

$$\begin{aligned}
\sigma t x - \frac{x^2}{2} &= -\frac{1}{2}\big(x^2 - 2\sigma t x\big) \\
&= -\frac{1}{2}\big[(x-\sigma t)^2 - \sigma^2 t^2\big] \\
&= -\frac{(x-\sigma t)^2}{2} + \frac{\sigma^2 t^2}{2}.
\end{aligned}$$

Portanto, podemos retirar também o fator $e^{\sigma^2 t^2/2}$ da integral:

$$\mathbb{E}[e^{tU}] = e^{\mu t + \sigma^2 t^2/2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \, e^{-(x-\sigma t)^2/2} \, dx.$$

A função que aparece nessa última integral é a densidade de uma $\mathrm{N}(\sigma t, 1)$, cuja integral sobre toda a reta vale 1. Podemos verificar isso diretamente com a translação $v = x - \sigma t$: nesse caso, $dv = dx$, os limites continuam sendo $-\infty$ e $+\infty$, e

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}\, e^{-(x-\sigma t)^2/2}\, dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}\, e^{-v^2/2}\, dv = 1,$$

pois reconhecemos a densidade da normal padrão. Concluímos que a esperança calculada é finita para qualquer $t \in \mathbb{R}$. Assim, a geradora de $U$ está definida em toda a reta e é dada por

$$M_U(t) = \exp\!\left(\mu t + \frac{\sigma^2 t^2}{2}\right), \qquad t \in \mathbb{R}. \tag{40}$$

Vamos voltar ao problema de calcular a função geradora de $Z \equiv X + Y$. Como $X$ e $Y$ são variáveis aleatórias independentes, o Teorema 22 nos dá $M_Z(t) = M_X(t)\, M_Y(t)$ para todo $t$ em que as duas geradoras são finitas. Aplicando (40) a cada parcela, obtemos

$$\begin{aligned} M_Z(t) &= M_X(t)\, M_Y(t) \\ &= \exp\!\left(\mu_1 t + \frac{\sigma_1^2 t^2}{2}\right) \exp\!\left(\mu_2 t + \frac{\sigma_2^2 t^2}{2}\right) \\ &= \exp\!\left[(\mu_1 + \mu_2)t + \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}\right]. \end{aligned}$$

O lado direito é exatamente (40) com $\mu = \mu_1 + \mu_2$ e $\sigma^2 = \sigma_1^2 + \sigma_2^2$ e, portanto, trata-se da função geradora de momentos de uma variável aleatória com distribuição $\mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, que é finita para todo $t \in \mathbb{R}$. Assim, vale a hipótese do Teorema 23 para qualquer $\delta > 0$, e a conclusão de unicidade assegura que $Z$ possui essa distribuição, isto é,

$$Z \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Nas notas sobre somas de variáveis aleatórias contínuas, mostramos, por meio da convolução das densidades, que a soma de duas variáveis aleatórias independentes com distribuições gama de mesma taxa é ainda uma gama. Mais precisamente, se $X \sim \mathrm{Gama}(w_1, \lambda)$ e $Y \sim \mathrm{Gama}(w_2, \lambda)$ são independentes, com $w_1, w_2, \lambda > 0$, então $X + Y \sim \mathrm{Gama}(w_1 + w_2, \lambda)$. Esse resultado se obtém muito mais rapidamente com funções geradoras de momentos. Para $t < \lambda$,

$$M_{X+Y}(t) = M_X(t) M_Y(t) = \left(\frac{\lambda}{\lambda - t}\right)^{w_1} \left(\frac{\lambda}{\lambda - t}\right)^{w_2} = \left(\frac{\lambda}{\lambda - t}\right)^{w_1 + w_2},$$

que é a geradora de $\mathrm{Gama}(w_1 + w_2, \lambda)$; a unicidade, Teorema 23, converte essa identidade algébrica na igualdade de distribuições. O exemplo anterior, com a soma de duas normais, mostra o mesmo mecanismo: o produto das geradoras resolve a soma, e a unicidade traduz o resultado de volta em distribuição. No caso gama a economia é ainda mais visível, pois a convolução exige a relação beta–gama.

# 4. Uma Aplicação: Faulhaber e Bernoulli

As fórmulas

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}, \qquad \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6} \qquad \text{e} \qquad \sum_{k=1}^{n} k^3 = \left(\frac{n(n+1)}{2}\right)^2$$

costumam ser demonstradas uma a uma, por indução ou por argumentos telescópicos, e cada nova potência parece exigir uma ideia nova. Nesta seção mostramos que todas elas podem ser obtidas pelo mesmo procedimento. Para isso, vamos interpretar uma soma de potências como um momento de uma variável aleatória uniforme. A função geradora de momentos reunirá então todas essas somas em uma única função: sua derivada de ordem $m$ em 0 fornecerá, a menos de um fator constante, a soma das potências de ordem $m$. A fórmula geral, que aparece na obra de Bernoulli de 1713 e usa os números que hoje levam seu nome, sai então de uma única expansão de Taylor. Faulhaber já conhecia muitos casos particulares.

Fixamos um inteiro $n \geqslant 1$ e tomamos $X$ uniforme em $\{0, 1, \ldots, n-1\}$, isto é,

$$\mathbb{P}(X = k) = \frac{1}{n}, \qquad k = 0, 1, \ldots, n-1.$$

Essa escolha é útil porque a esperança de uma função de $X$ é a média dos valores dessa função nos pontos $0, 1, \ldots, n-1$. Em particular, para cada inteiro $m \geqslant 0$,

$$\mathbb{E}[X^m] = \sum_{k=0}^{n-1} k^m \, \mathbb{P}(X = k) = \frac{1}{n} \sum_{k=0}^{n-1} k^m.$$

Assim, basta calcular esse momento e multiplicá-lo por $n$ para obter a soma desejada. Quando $m = 0$, adotamos aqui a convenção $k^0 = 1$ também para $k = 0$: a soma tem $n$ parcelas iguais a 1 e $\mathbb{E}[X^0] = 1$.

Como $X$ assume apenas uma quantidade finita de valores, todos os seus momentos são finitos. Sua geradora também é finita para todo $t$ real e vale

$$M_X(t) = \mathbb{E}[e^{tX}] = \frac{1}{n} \sum_{k=0}^{n-1} e^{tk} = \frac{1}{n}\Big(1 + e^t + \cdots + e^{(n-1)t}\Big).$$

Para $t \neq 0$, a expressão entre parênteses é uma soma geométrica de razão $e^t \neq 1$. De fato, escrevendo $S = 1 + e^t + \cdots + e^{(n-1)t}$, a subtração $e^t S - S$ cancela todos os termos intermediários e dá $(e^t - 1)S = e^{nt} - 1$. Portanto,

$$M_X(t) = \frac{1}{n} \sum_{k=0}^{n-1} e^{tk} = \frac{1}{n} \, \frac{e^{nt} - 1}{e^t - 1}, \qquad t \neq 0. \tag{41}$$

Em $t = 0$, usamos a soma finita original e obtemos $M_X(0) = 1$. Essa soma mostra ainda que $M_X$ é contínua e possui derivadas de todas as ordens. Como a geradora é finita em toda a reta, podemos aplicar Teorema 21. Esse teorema dá, para todo inteiro $m \geqslant 0$,

$$\mathbb{E}[X^m] = M_X^{(m)}(0), \qquad \text{e portanto} \qquad \sum_{k=0}^{n-1} k^m = n \, M_X^{(m)}(0). \tag{42}$$

Resta calcular as derivadas de $M_X$ em 0. Para isso, basta conhecer a expansão de Taylor do quociente em (41). De fato, a expansão da geradora é

$$n M_X(t) = \sum_{m=0}^{\infty} \left( \sum_{k=0}^{n-1} k^m \right) \frac{t^m}{m!}.$$

Uma série da forma $\sum_{m\geqslant 0} a_m t^m/m!$ é chamada de *função geradora exponencial* da sequência $(a_m)$. Neste caso, os coeficientes $a_m$ são precisamente as somas que queremos calcular.

O denominador $e^t - 1$ se anula em 0. Entretanto, quando $t \to 0$, temos $(e^t - 1)/t \to 1$, de modo que seu recíproco $t/(e^t - 1)$ também tende a 1. Isso sugere separar o quociente em dois fatores que têm limite quando $t \to 0$:

$$\frac{e^{nt} - 1}{e^t - 1} = \frac{e^{nt} - 1}{t}\,\frac{t}{e^t - 1}.$$

Com $n$ fixado, o primeiro fator tende a $n$ e o segundo tende a 1 quando $t \to 0$. Para aproveitar a expansão do segundo fator, distribuímos o numerador $e^{nt} - 1$ e escrevemos, para $t \neq 0$,

$$\frac{e^{nt} - 1}{e^t - 1} = \frac{1}{t}\left(\frac{t\,e^{nt}}{e^t - 1} - \frac{t}{e^t - 1}\right).$$

As duas parcelas entre parênteses têm agora a mesma forma: $t\,e^{xt}/(e^t - 1)$, nos pontos $x = n$ e $x = 0$. É essa função que define os polinômios de Bernoulli.

**Definição 25** (Números e polinômios de Bernoulli)**.** Os *números de Bernoulli* $B_j$ são definidos pela expansão

$$\frac{t}{e^t - 1} = \sum_{j=0}^{\infty} B_j\,\frac{t^j}{j!}, \qquad |t| < 2\pi, \tag{43}$$

e os *polinômios de Bernoulli* $B_j(x)$ pela expansão

$$\frac{t\,e^{xt}}{e^t - 1} = \sum_{j=0}^{\infty} B_j(x)\,\frac{t^j}{j!}, \qquad |t| < 2\pi, \tag{44}$$

para cada $x \in \mathbb{R}$ fixado. Em $t = 0$, os quocientes são entendidos pelo seu limite quando $t \to 0$, igual a 1.

Usaremos essas séries apenas perto de $t = 0$, onde convergem absolutamente. O raio de convergência $2\pi$ pode ser justificado por análise complexa: após preencher o valor em 0 por continuidade, as singularidades mais próximas são os polos $t = \pm 2\pi i$. Essa justificativa não será necessária nos cálculos abaixo.

Vamos mostrar como a definição permite calcular os primeiros números. Como $t/(e^t - 1) \to 1$ quando $t \to 0$, temos $B_0 = 1$. Além disso, multiplicando (43) por $(e^t - 1)/t$, obtemos

$$\left(1 + B_1 t + \frac{B_2}{2!}t^2 + \cdots\right)\left(1 + \frac{t}{2!} + \frac{t^2}{3!} + \cdots\right) = 1.$$

O lado direito tem coeficientes nulos em todas as potências positivas de $t$. Comparando os coeficientes de $t$ e de $t^2$, respectivamente,

$$B_1 + \frac{1}{2} = 0, \qquad \frac{B_2}{2} + \frac{B_1}{2} + \frac{1}{6} = 0.$$

A primeira equação dá $B_1 = -1/2$; substituindo esse valor na segunda, obtemos $B_2/2 = 1/4 - 1/6 = 1/12$, isto é, $B_2 = 1/6$. Continuando a comparação, calculamos os demais valores. Os primeiros são:

| $j$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| $B_j$ | 1 | $-\frac{1}{2}$ | $\frac{1}{6}$ | 0 | $-\frac{1}{30}$ | 0 | $\frac{1}{42}$ | 0 | $-\frac{1}{30}$ |

**Proposição 26.** Para todo $j \geqslant 0$ e todo $x \in \mathbb{R}$,

$$B_j(x) = \sum_{i=0}^{j} \binom{j}{i} B_i \, x^{j-i}. \tag{45}$$

Em particular, $B_j(0) = B_j$ e $B_1(x) = x - \frac{1}{2}$. Além disso, $B_{2k+1} = 0$ para todo $k \geqslant 1$.

**Prova**. Para encontrar $B_j(x)$, precisamos identificar o coeficiente de $t^j$ no produto que define sua geradora. Para $|t| < 2\pi$, as duas séries abaixo convergem absolutamente, o que permite multiplicá-las e agrupar os termos de mesma potência:

$$\frac{t\, e^{xt}}{e^t - 1} = \left( \sum_{r=0}^{\infty} \frac{x^r t^r}{r!} \right) \left( \sum_{i=0}^{\infty} B_i \, \frac{t^i}{i!} \right).$$

Um termo de grau $r$ da primeira série e um termo de grau $i$ da segunda contribuem para $t^j$ exatamente quando $r + i = j$, ou seja, $r = j - i$. Assim, o coeficiente de $t^j$ nesse produto é

$$\sum_{i=0}^{j} \frac{B_i x^{j-i}}{i!(j-i)!} = \frac{1}{j!} \sum_{i=0}^{j} \binom{j}{i} B_i x^{j-i}.$$

Na série (44), esse mesmo coeficiente é $B_j(x)/j!$. Igualando as duas expressões e multiplicando por $j!$, obtemos (45). Em particular, essa expressão mostra que $B_j(x)$ é de fato um polinômio. Com $x = 0$, sobra apenas o termo $i = j$, dando $B_j(0) = B_j$; para $j = 1$, $B_1(x) = B_0 x + B_1 = x - 1/2$.

Resta explicar por que os números de índice ímpar maior que 1 se anulam. Como $B_1 = -1/2$, somar $t/2$ à geradora cancela justamente seu termo de grau 1. A função resultante é par: para $t \neq 0$,

$$F(t) := \frac{t}{e^t - 1} + \frac{t}{2} = \frac{t(e^t + 1)}{2(e^t - 1)},$$

e uma substituição direta dá

$$F(-t) = \frac{(-t)(e^{-t} + 1)}{2(e^{-t} - 1)} = \frac{(-t)(1 + e^t)}{2(1 - e^t)} = F(t).$$

A igualdade vale também em 0 por continuidade. Na série de Taylor de uma função par, os coeficientes das potências ímpares são nulos: substituir $t$ por $-t$ troca o sinal apenas dessas potências, mas deixa a função inalterada. Para os graus ímpares $j \geqslant 3$, os coeficientes de $F$ continuam sendo $B_j/j!$, pois só alteramos o termo de grau 1. Portanto, $B_{2k+1} = 0$ para $k \geqslant 1$. ■

**Teorema 27** (Fórmula de Faulhaber)**.** Para quaisquer inteiros $n \geqslant 1$ e $m \geqslant 0$, com a convenção $0^0 = 1$,

$$\sum_{k=0}^{n-1} k^m = \frac{B_{m+1}(n) - B_{m+1}}{m+1}. \tag{46}$$

Para $m \geqslant 1$, equivalentemente,

$$\sum_{k=1}^{n} k^m = \frac{B_{m+1}(n+1) - B_{m+1}}{m+1}. \tag{47}$$

**Prova**. Pela relação (42), basta encontrar a derivada $M_X^{(m)}(0)$. Vamos obtê-la identificando o coeficiente de $t^m$ na série da geradora: esse coeficiente é $M_X^{(m)}(0)/m!$. Usando a decomposição que motivou os polinômios de Bernoulli, para $0 < |t| < 2\pi$ temos

$$M_X(t) = \frac{1}{n\,t}\left(\frac{t\,e^{nt}}{e^t-1} - \frac{t}{e^t-1}\right).$$

Expandimos a primeira parcela com (44), tomando $x = n$, e a segunda com (43):

$$M_X(t) = \frac{1}{n\,t}\sum_{j=0}^{\infty}\Big(B_j(n) - B_j\Big)\frac{t^j}{j!}.$$

O termo $j = 0$ é nulo, pois $B_0(x) = B_0 = 1$. Por isso, todos os termos restantes têm um fator $t$, que cancela o $t$ do denominador:

$$M_X(t) = \frac{1}{n}\sum_{j=1}^{\infty}\Big(B_j(n) - B_j\Big)\frac{t^{j-1}}{j!}.$$

Agora as potências aparecem como $t^{j-1}$. Para escrevê-las como $t^m$, fazemos a mudança de índice $m = j-1$, isto é, $j = m+1$. Quando $j$ começa em 1, o novo índice $m$ começa em 0, e $j! = (m+1)!$. Logo,

$$M_X(t) = \frac{1}{n}\sum_{m=0}^{\infty}\frac{B_{m+1}(n) - B_{m+1}}{(m+1)!}\,t^m,$$

inicialmente para $0 < |t| < 2\pi$ e também em $t = 0$ por continuidade. Comparando essa série com a série de Taylor de $M_X$, obtemos

$$\frac{M_X^{(m)}(0)}{m!} = \frac{1}{n}\,\frac{B_{m+1}(n) - B_{m+1}}{(m+1)!}.$$

Multiplicando por $n\,m!$ e usando $(m+1)! = (m+1)m!$, chegamos a

$$\sum_{k=0}^{n-1} k^m = nM_X^{(m)}(0) = \frac{m!}{(m+1)!}\Big(B_{m+1}(n) - B_{m+1}\Big) = \frac{B_{m+1}(n) - B_{m+1}}{m+1}.$$

Isso prova (46). Para obter a soma de 1 a $n$, substituímos $n$ por $n+1$, obtendo uma soma de 0 a $n$. Quando $m \geqslant 1$, seu termo $k = 0$ vale $0^m = 0$ e pode ser retirado. Essa é precisamente (47). Para $m = 0$, o termo $k = 0$ vale 1 pela convenção adotada, razão pela qual essa segunda forma foi enunciada apenas para $m \geqslant 1$. ■

**Exemplos.** Para aplicar a fórmula, calculamos primeiro os polinômios necessários usando (45) e os valores da tabela:

$$B_1(x) = x - \frac{1}{2},$$

$$B_2(x) = x^2 + 2\left(-\frac{1}{2}\right)x + \frac{1}{6} = x^2 - x + \frac{1}{6},$$

$$B_3(x) = x^3 + 3\left(-\frac{1}{2}\right)x^2 + 3\left(\frac{1}{6}\right)x = x^3 - \frac{3}{2}x^2 + \frac{1}{2}x,$$

$$B_4(x) = x^4 + 4\left(-\frac{1}{2}\right)x^3 + 6\left(\frac{1}{6}\right)x^2 - \frac{1}{30} = x^4 - 2x^3 + x^2 - \frac{1}{30}.$$

Para $m=0$, (46) dá $B_1(n)-B_1=(n-1/2)-(-1/2)=n$, como esperado. Para $m=1,2,3$, a mesma fórmula fornece

$$\sum_{k=0}^{n-1} k = \frac{B_2(n)-B_2}{2} = \frac{n^2-n}{2} = \frac{n(n-1)}{2},$$

$$\sum_{k=0}^{n-1} k^2 = \frac{B_3(n)-B_3}{3} = \frac{n^3-\frac{3}{2}n^2+\frac{1}{2}n}{3} = \frac{n(n-1)(2n-1)}{6},$$

$$\sum_{k=0}^{n-1} k^3 = \frac{B_4(n)-B_4}{4} = \frac{n^4-2n^3+n^2}{4} = \frac{n^2(n-1)^2}{4}.$$

Substituindo $n$ por $n+1$ nessas três expressões, recuperamos as fórmulas de somas de 1 a $n$ apresentadas no início da seção.

Por fim, em (45) o termo $i=0$ é $B_0x^j=x^j$ e os demais têm grau menor que $j$. Assim, $B_{m+1}(x)$ tem grau $m+1$ e coeficiente líder 1. Subtrair a constante $B_{m+1}$ não altera esse termo de maior grau; dividir por $m+1$ mostra que a soma em (46) é um polinômio em $n$ de grau $m+1$, cujo termo líder é $n^{m+1}/(m+1)$.

**Observação 28** (Bernoulli e a função zeta)**.** Os números de Bernoulli também aparecem em um problema que, à primeira vista, parece distante das somas finitas de potências: calcular as séries infinitas

$$\zeta(2k) = \sum_{r=1}^{\infty} \frac{1}{r^{2k}}, \qquad k \geqslant 1.$$

Euler obteve

$$\zeta(2k) = (-1)^{k+1} \frac{(2\pi)^{2k}}{2\,(2k)!}\, B_{2k}, \qquad k \geqslant 1, \tag{48}$$

fórmula que não demonstraremos aqui. Por exemplo, com $k=1$ e $B_2=1/6$, temos

$$\zeta(2) = \frac{(2\pi)^2}{2\cdot 2!}\,\frac{1}{6} = \frac{4\pi^2}{4}\,\frac{1}{6} = \frac{\pi^2}{6}.$$

Com $k=2$ e $k=3$, usando $B_4=-1/30$ e $B_6=1/42$, obtemos

$$\zeta(4) = \frac{16\pi^4}{2\cdot 24\cdot 30} = \frac{\pi^4}{90}, \qquad \zeta(6) = \frac{64\pi^6}{2\cdot 720\cdot 42} = \frac{\pi^6}{945}.$$

Assim, $\zeta(2k)$ é sempre um múltiplo racional de $\pi^{2k}$; e, como $\zeta(2k)>0$, podemos tomar módulos em (48) e isolar $|B_{2k}|$:

$$|B_{2k}| = \frac{2(2k)!}{(2\pi)^{2k}}\,\zeta(2k).$$

Além disso, $\zeta(2k)\to 1$ quando $k\to\infty$. Para verificar esse limite, separamos o termo $r=1$ e estimamos a cauda da série por uma integral:

$$0 \leqslant \zeta(2k)-1 = \sum_{r=2}^{\infty} \frac{1}{r^{2k}} \leqslant \frac{1}{2^{2k}} + \int_2^{\infty} x^{-2k}\,dx = \frac{1}{2^{2k}} + \frac{2^{1-2k}}{2k-1} \xrightarrow{k\to\infty} 0.$$

Aqui usamos que $x^{-2k}$ é decrescente: para cada $r \geqslant 3$, $r^{-2k} \leqslant \int_{r-1}^{r} x^{-2k}\,dx$. Voltando à expressão de $|B_{2k}|$, obtemos, quando $k \to \infty$,

$$\frac{|B_{2k}|}{2(2k)!/(2\pi)^{2k}} \longrightarrow 1, \qquad \text{isto é,} \qquad |B_{2k}| \sim \frac{2(2k)!}{(2\pi)^{2k}}.$$

A fórmula de Euler acima trata dos argumentos pares. Os valores nos inteiros ímpares maiores que 1 são menos compreendidos: $\zeta(3)$ é irracional (Apéry, 1978), mas não se sabe se é transcendente. Nas notas sobre esperança de variáveis discretas, $\zeta(2)$ e $\zeta(3)$ já haviam aparecido no cálculo dos momentos da distribuição zeta.

# 5. Exercícios propostos

**1.** Se $X$ é uniformemente distribuída no intervalo $(a, b)$, mostre que

$$\mathbb{E}[X^k] = \frac{b^{k+1} - a^{k+1}}{(b-a)(k+1)} \qquad \text{para } k = 1, 2, \ldots.$$

**2.** Se $X$ tem distribuição gama com parâmetros $w$ e $\lambda$, mostre que

$$\mathbb{E}[X^k] = \frac{\Gamma(w+k)}{\lambda^k\,\Gamma(w)} \qquad \text{para } k = 1, 2, \ldots.$$

*Nota:* usamos a convenção de densidade adotada nas notas sobre somas de variáveis aleatórias, em que

$$f_X(x) = \frac{\lambda^w}{\Gamma(w)}\, x^{w-1} e^{-\lambda x}, \qquad x > 0,$$

com $w > 0$ o parâmetro de forma e $\lambda > 0$ a taxa. Verifique, antes de começar, que a integral que define $\mathbb{E}[X^k]$ converge para todo $k \geqslant 1$.

**3.** Se $X$ tem distribuição $\chi^2$ com $n$ graus de liberdade, mostre que

$$\mathbb{E}[X^k] = 2^k\, \frac{\Gamma\big(k + \frac{1}{2}n\big)}{\Gamma\big(\frac{1}{2}n\big)} \qquad \text{para } k = 1, 2, \ldots.$$

*Sugestão:* observe que a distribuição $\chi^2$ com $n$ graus de liberdade é o caso particular $w = n/2$ e $\lambda = 1/2$ da distribuição gama e aplique o exercício anterior.

$*$ **4.** Sejam $X$ e $Y$ variáveis aleatórias com distribuição normal bivariada de parâmetros $\mu_1$, $\mu_2$, $\sigma_1$, $\sigma_2$ e $\rho$. Mostre que

$$\mathrm{Cov}(X, Y) = \rho\,\sigma_1\sigma_2 \qquad \text{e} \qquad \rho(X, Y) = \rho.$$

*Nota:* a densidade normal bivariada já foi estudada nas notas sobre densidades condicionais. Use a expressão da densidade lá obtida e, para calcular $\mathrm{Cov}(X, Y)$, comece determinando $\mathbb{E}[X \mid Y]$, tal como fizemos naquelas notas. Este exercício justifica o nome do parâmetro $\rho$: ele é, de fato, o coeficiente de correlação entre as coordenadas.

**5.** Sejam $X_1, X_2, \ldots$ variáveis aleatórias não correlacionadas, todas com variância $\sigma^2$. Escreva $S_n = X_1 + X_2 + \cdots + X_n$ para $n \geqslant 1$. Mostre que

$$\mathrm{Cov}(S_m, S_n) = \mathrm{Var}(S_m) = m\sigma^2 \qquad \text{se } m \leqslant n.$$

*Sugestão:* use a bilinearidade da covariância e a hipótese de que $\operatorname{Cov}(X_i, X_j) = 0$ para $i \neq j$. A condição $m \leqslant n$ indica que $S_m$ contém as primeiras $m$ parcelas de $S_n$. Para $m > n$ a mesma conta dá $\operatorname{Cov}(S_m, S_n) = n\sigma^2$, pois cada parcela de $S_n$ aparece também em $S_m$.

∗ **6.** Mostre que $\operatorname{Cov}(X, Y) = 1$ quando $X$ e $Y$ têm densidade conjunta

$$f(x,y) = \begin{cases} \dfrac{1}{y}\, e^{-y-x/y}, & \text{se } x, y > 0, \\ 0, & \text{caso contrário.} \end{cases}$$

*Sugestão:* condicione em $Y$ para calcular $\mathbb{E}[X \mid Y]$ e, em seguida, $\mathbb{E}[X]$ e $\mathbb{E}[XY]$; ou trabalhe diretamente com a densidade conjunta, calculando as integrais iteradas na ordem mais conveniente.

**7.** Determine a função geradora de momentos de uma variável aleatória com

(a) distribuição gama de parâmetros $w$ e $\lambda$;

(b) distribuição de Poisson de parâmetro $\lambda$.

*Nota:* em (a), use a mesma convenção de densidade do exercício 2, com $\lambda$ como taxa. Em ambos os itens, indique o intervalo de valores de $t$ para os quais a FGM está definida.

**8.** Se $X$ tem distribuição normal com média $\mu$ e variância $\sigma^2$, determine $\mathbb{E}[X^3]$.
*Sugestão:* use a expressão analítica da FGM da normal obtida em (40). Pelo Teorema 21, basta derivar $M_X(t)$ três vezes em relação a $t$ e avaliar a derivada em $t = 0$.

**9.** Mostre que, se $X$ tem distribuição normal, então $aX + b$ também tem distribuição normal, quaisquer que sejam $a, b \in \mathbb{R}$ com $a \neq 0$.
*Sugestão:* use a unicidade das funções geradoras de momentos, vista nesta nota, e a expressão da FGM da distribuição normal para calcular a FGM de $aX + b$ e reconhecê-la como a FGM de uma normal.

∗ **10.** Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas, com função geradora de momentos comum $M$. Seja $N$ uma variável aleatória que assume valores inteiros não negativos, com função geradora de probabilidade $G$, e suponha que $N$ seja independente da sequência $(X_i)$. Mostre que a soma aleatória $S = X_1 + X_2 + \cdots + X_N$, com $S = 0$ se $N = 0$, satisfaz

$$M_S(t) = G\Big(M(t)\Big).$$

Considere os valores de $t$ para os quais $M(t) < \infty$ e $G(M(t)) < \infty$. Aqui, a função $G$ é dada pela série $G(s) = \sum_{n=0}^{\infty} \mathbb{P}(N = n) s^n$ também nos pontos $s > 1$ em que essa série converge.
*Sugestão:* condicione no valor de $N$ e use a independência para fatorar a esperança. O resultado é a versão, na linguagem das funções geradoras de momentos, da composição de funções geradoras de probabilidade estudada nas notas sobre funções geradoras de probabilidade, em que a soma de um número aleatório de parcelas foi examinada no caso discreto. A composição obtida aqui envolve $G$ no ponto $M(t) > 0$, que pode ser maior que 1; por isso, é necessário verificar a convergência da série nesse ponto para obter uma geradora finita.

*Nota:* além da independência entre $N$ e a sequência, usamos a independência mútua das parcelas $X_i$ para garantir a identidade $M_{X_1+\cdots+X_n}(t) = M(t)^n$. Sem essa hipótese, a identidade pode falhar e a fórmula para $M_S(t)$ pode deixar de valer. O enunciado impresso do livro que seguimos menciona apenas que as variáveis são identicamente distribuídas; acrescentamos a hipótese que falta para que o enunciado seja verdadeiro.

# Referências

[1] Geoffrey Grimmett and Dominic Welsh. *Probability—an introduction*. Oxford University Press, Oxford, second edition, 2014.