UnB | Departamento de Matemática

A Lei Forte dos Grandes Números

L. Cioletti

Outubro de 2026

Resumo

Nestas notas, estudamos a Lei Forte dos Grandes Números e suas aplicações. Apresentamos inicialmente uma demonstração para variáveis aleatórias independentes e identicamente distribuídas com quarto momento finito, baseada na desigualdade de Markov e no critério de Borel-Cantelli. Em seguida, demonstramos a versão de Etemadi para variáveis integráveis, identicamente distribuídas e independentes dois a dois. Mostramos também que, no caso independente e identicamente distribuído, a integrabilidade é necessária para a convergência quase certa das médias a um limite finito. Analisamos o comportamento das médias amostrais de variáveis independentes com distribuição de Cauchy. A família de Pareto exemplifica a aplicação da Lei Forte a sequências de variáveis integráveis com variância infinita. Entre as aplicações, incluímos a normalidade dos números reais, a integração de Monte Carlo e a convergência da função de distribuição empírica. As notas contêm exemplos e exercícios propostos. A exposição segue as Seções 2.3 e 2.4 de [2], as Seções 1, 2 e 6 do Capítulo 1 de [1] e as Seções 8.1 e 8.2 do Capítulo 8 de [3].

1. Introdução

Nas notas anteriores sobre modos de convergência, aprendemos a distinguir cinco noções de convergência para sequências de variáveis aleatórias. A lei fraca dos grandes números, estudada nas notas anteriores afirma que, sob determinadas condições, a média empírica S_n/n converge em probabilidade para seu valor esperado μ, isto é,

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{\mathbb{P}} \mu.$$

Ou seja, para $n \in \mathbb{N}$ suficientemente grande, a probabilidade de que a média se afaste de μ por uma magnitude maior que ε é pequena. A lei forte faz uma afirmação ainda mais precisa, ela garante que para quase todo ponto $\omega \in \Omega$, a sequência numérica $S_n(\omega)/n$ converge para μ no sentido usual de sequências numéricas.

A diferença entre as duas leis não é um refinamento técnico. Considere o lançamento repetido de um dado honesto e seja S_n a soma dos n primeiros resultados. A lei fraca diz que, para n grande, é improvável que S_n/n se afaste de $7/2$. A lei forte diz que, com probabilidade 1, existe um índice a partir do qual $S_n(\omega)/n$ permanece próximo de $7/2$ para

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

sempre. A segunda afirmação implica a primeira, mas não é consequência dela: há sequências que convergem em probabilidade e, ainda assim, oscilam infinitas vezes em quase todos os pontos, como a sequência que varre o intervalo discutida nas notas anteriores. A lei forte exclui esse comportamento para as médias de variáveis independentes e identicamente distribuídas com $\mathbb{E}[|X_1|] < \infty$.

O exemplo mais usado nestas notas para ilustrar os resultados principais será o das frequências de dígitos. Mais precisamente, suponha que $\omega \in [0,1]$ é um número real sorteado ao acaso, de maneira uniforme, e considere sua expansão decimal $\omega = 0{,}d_1d_2d_3\ldots$. Qual é a frequência relativa do dígito 7 entre os n primeiros dígitos, quando n cresce? A intuição diz que a frequência deve se aproximar de $1/10$, e o teorema de Borel, que demonstraremos, confirma a intuição para quase todo ω. O mesmo problema pode ser formulado de maneira mais clara usando o espaço produto $(\Omega, \mathcal{F}, \mathbb{P})$, onde

$$\Omega \equiv \{0,1,\ldots,9\}^{\mathbb{N}},$$

$\mathcal{F}$ é uma σ-álgebra apropriada e a medida de probabilidade $\mathbb{P}$ sendo a única medida de probabilidade sobre $\mathcal{F}$ satisfazendo $\mathbb{P}(X_n = k) = 1/10$, onde $X_n(\omega) = \omega_n$ é a n-ésima projeção $X_n : \Omega \to \{0,1,2,\ldots,9\}$ (que sobre o espaço $(\Omega, \mathcal{F}, \mathbb{P})$ é uma variável aleatória) e $k \in \{0,1,2,\ldots,9\}$ é um dígito fixado. Neste modelo, os dígitos são representados pelas variáveis aleatórias X_n's que são, por construção, independentes e identicamente distribuídas. Além do mais, neste modelo a frequência relativa de um dígito k fixado, é simplesmente a média empírica da sequência de indicadoras $\mathbb{1}_{\{X_j=k\}}$, isto é,

$$\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{X_j=k\}}$$

Neste modelo o problema da determinação da frequência relativa dos dígitos pode ser obtido como uma aplicação da Lei Forte dos Grandes Números.

A estratégia da prova segue de perto a adotada na referência [2]. Primeiro, apresentamos uma belíssima prova da Lei Forte para o caso em que as variáveis têm quarto momento finito. Além da prova neste caso ser muito clara e curta ela ilustra como as principais ferramentas são utilizadas no caso geral. Veremos que se assumimos a existência do quarto momento, então a desigualdade de Markov aplicada à quarta potência da soma centrada, $(S_n - n\mu)^4$, fornece uma cota de ordem $1/n^2$, cuja soma converge. Este fato irá permitir aplicar o critério de Borel-Cantelli, apresentado nas notas anteriores, para concluir a ocorrência da convergência quase certa. Embora esta versão seja ainda um caso particular ela é suficiente para resolver o problema da frequência relativa do dígitos, porque as variáveis indicadoras são limitadas e portanto possuem todos os momentos finitos.

Depois, voltamos nossas atenções para o caso geral. Vamos então substituir a hipótese de quarto momento por $\mathbb{E}[|X_1|] < \infty$. A prova apresentada nestas notas para este caso é a prova inspirada no argumento de Etemadi. Como veremos, este argumento contorna a ausência de momentos de ordem superior por meio de um truncamento e de uma estimativa de momentos para as variáveis truncadas. O truncamento faz com que cada variável seja limitada por uma escala que cresce com o índice, e a estimativa mostra que a soma das variâncias truncadas, dividida pelos quadrados dos índices, é finita. A convergência das esperanças truncadas identifica o centro dessas somas com a média original. Com isso, a cota de Chebyshev tem somabilidade garantida ao longo de uma subsequência geométrica de índices, mesmo quando não a temos ao longo de todos os índices. O preenchimento dos

índices restantes usa a monotonicidade da soma truncada, e é aqui que a redução a variáveis não negativas se torna necessária.

A prova de Etemadi é um bom exemplo de argumento em que cada etapa tem uma função clara: o truncamento elimina a cauda das variáveis, a estimativa de momentos controla a variância acumulada, a subsequência geométrica torna a série de probabilidades convergente e a monotonicidade transporta a conclusão da subsequência para todos os índices. Ao longo da exposição, procuraremos explicitar qual dessas funções cada parte do argumento desempenha, porque a prova é frequentemente apresentada como uma sequência de estimativas sem que o leitor veja por que cada passo foi dado.

O texto está organizado da seguinte maneira. Na Seção 2 recordamos as ferramentas de cálculo de probabilidades que serão usadas: as desigualdades de Markov e Chebyshev, os dois lemas de Borel-Cantelli, o critério de convergência quase certa por somabilidade e a representação da esperança pela integral da cauda. Na seção Seção 3 apresentamos mais alguns detalhes sobre a construção do espaço produto e sobre o problema das frequências relativas. Na Seção 4 demonstramos a lei forte sob a hipótese de quarto momento finito e em seguida, mostramos como este resultado é aplicado para resolver o problema da frequência relativa dos dígitos. Na Seção 5 apresentamos a prova de Etemadi. A Seção 6 mostra que a hipótese $\mathbb{E}[|X_1|] < \infty$ é necessária no caso independente e identicamente distribuído. A Seção 7 reúne aplicações, e a Seção 8 propõe exercícios.

A exposição da prova de Etemadi segue de perto a Seção 2.4 de [2], com detalhamento maior das passagens que o livro deixa para o leitor. O exemplo dos dígitos e o enunciado do Teorema de Borel seguem as Seções 1 e 2 do Capítulo 1 de [1], adaptadas ao espaço produto $\{0, 1, \ldots, 9\}^{\mathbb{N}}$. A prova da Lei Forte no caso de quarto momento é inspirada na Seção 2.3 da referência [2].

2. Preliminares

Nesta seção, reunimos as ferramentas que serão usadas nas provas da lei forte e de sua recíproca. Recordamos as desigualdades e os lemas das notas anteriores e explicitamos a relação entre integrabilidade e probabilidades de cauda.

2.1. Desigualdades de Markov e Chebyshev e a Lei Fraca

As desigualdades de Markov e Chebyshev, estudadas em notas de aula anteriores, constituem o ponto de partida natural para estimar probabilidades de cauda e desvios a partir de momentos. A pergunta fundamental que orienta essas estimativas é simples: conhecendo apenas a esperança ou a variância de uma variável aleatória, o que podemos garantir sobre a probabilidade de ela assumir valores excessivamente grandes ou afastar-se de sua média?

A resposta mais elementar é dada pela desigualdade de Markov: se Z é uma variável aleatória não negativa e $a > 0$, então

$$\mathbb{P}(Z \geqslant a) \leqslant \frac{\mathbb{E}[Z]}{a}. \tag{1}$$

A intuição por trás de (1) é imediata: se a probabilidade de Z superar um patamar a fosse estritamente maior que $\mathbb{E}[Z]/a$, a contribuição desse evento já excederia a própria média global de Z, o que é impossível.

Quando dispomos de mais informação sobre a distribuição, como o segundo momento finito, podemos refinar substancialmente o controle das flutuações. Seja X uma variável aleatória com $\mathbb{E}[X^2] < \infty$. Fixado $\varepsilon > 0$, a função quadrática nos permite transformar o evento de desvio $\{|X - \mathbb{E}[X]| \geqslant \varepsilon\}$ na condição equivalente $\{(X - \mathbb{E}[X])^2 \geqslant \varepsilon^2\}$. Aplicando a desigualdade de Markov (1) à variável não negativa $Z = (X - \mathbb{E}[X])^2$ com o patamar $a = \varepsilon^2$, obtemos a desigualdade de Chebyshev:

$$\mathbb{P}\Big(|X - \mathbb{E}[X]| \geqslant \varepsilon\Big) \leqslant \frac{\mathrm{Var}(X)}{\varepsilon^2}. \tag{2}$$

A desigualdade (2) mostra com clareza como a variância mede a concentração de X em torno de sua esperança: quanto menor a dispersão, menor a probabilidade de observarmos desvios de ordem ε.

Vejamos agora como essa estimativa se comporta ao avaliarmos o comportamento das médias amostrais. Considere uma sequência de variáveis aleatórias $X_1, X_2, \ldots$ com variâncias finitas e suponha que sejam duas a duas independentes. Definindo a soma parcial $S_n = X_1 + \cdots + X_n$, a independência aos pares assegura que as covariâncias se anulam, de modo que a variância da soma se reduz à soma das variâncias:

$$\mathrm{Var}(S_n) = \sum_{k=1}^{n} \mathrm{Var}(X_k).$$

Consequentemente, a variância da média amostral S_n/n decai com o quadrado de n no denominador:

$$\mathrm{Var}\Big(\frac{S_n}{n}\Big) = \frac{1}{n^2} \sum_{k=1}^{n} \mathrm{Var}(X_k).$$

Aplicando a desigualdade de Chebyshev (2) à variável S_n/n, obtemos, para cada $\varepsilon > 0$, a estimativa

$$\mathbb{P}\Big(\Big|\frac{S_n}{n} - \bar{\mu}_n\Big| \geqslant \varepsilon\Big) \leqslant \frac{1}{\varepsilon^2}\,\mathrm{Var}\Big(\frac{S_n}{n}\Big), \qquad \text{onde} \qquad \bar{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}[X_k]. \tag{3}$$

No cenário particular em que as variáveis são identicamente distribuídas, com média comum μ e variância finita σ^2, temos $\bar{\mu}_n = \mu$ e o lado direito de (3) assume a forma explícita

$$\mathbb{P}\Big(\Big|\frac{S_n}{n} - \mu\Big| \geqslant \varepsilon\Big) \leqslant \frac{\sigma^2}{n\varepsilon^2}.$$

Como essa cota tende a zero quando $n \to \infty$, concluímos que a média amostral S_n/n converge em probabilidade para a média teórica μ. Esse resultado é precisamente a Lei Fraca dos Grandes Números sob a hipótese de variância finita.

No entanto, há uma limitação crucial nessa estimativa que motiva todo o estudo que faremos a seguir. Sempre que $\sigma^2 > 0$, a cota superior fornecida por Chebyshev decai apenas com taxa $1/n$. Como a série harmônica $\sum_{n=1}^{\infty} 1/n$ diverge, essas probabilidades de desvio não são somáveis em n. Isso nos impede de usar diretamente o primeiro lema de Borel-Cantelli para deduzir que a convergência ocorre quase certamente. Essa constatação é a motivação central destas notas de aula: para demonstrar a Lei Forte dos Grandes Números, precisaremos ou de momentos superiores que produzam cotas somáveis (como $1/n^2$), ou de estratégias mais refinadas que explorem truncamentos e subsequências ao longo das quais a somabilidade possa ser recuperada.

2.2. Borel-Cantelli e o Critério de Convergência Quase Certa

Os resultados a seguir foram discutidos nas notas *Modos de Convergência de Sequências de Variáveis Aleatórias*. Aqui recordamos os enunciados na forma em que serão usados nas provas da lei forte e de sua recíproca.

Para uma sequência de eventos (A_n), o limite superior

$$\limsup_{n\to\infty} A_n = \bigcap_{N=1}^{\infty} \bigcup_{n\geqslant N} A_n$$

é o evento formado pelos pontos ω que pertencem a A_n para infinitos índices n. Escrevemos também $\{A_n \text{ i.v.}\}$ para esse evento. O primeiro lema de Borel-Cantelli afirma que, se $\sum_n \mathbb{P}(A_n) < \infty$, então

$$\mathbb{P}\Big(\limsup_{n\to\infty} A_n\Big) = 0,$$

isto é, com probabilidade um apenas finitos dos eventos A_n ocorrem. Note que este resultado não exige a independência dos eventos.

Aplicado aos eventos de desvio $\{|X_n - X| \geqslant \varepsilon\}$, o primeiro lema fornece o *critério de Borel-Cantelli para a convergência quase certa*: se $X, X_1, X_2, \ldots$ são variáveis aleatórias e, para todo $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}\Big(|X_n - X| \geqslant \varepsilon\Big) < \infty, \tag{4}$$

então $X_n \to X$ quase certamente. A demonstração detalhada desse critério foi apresentada na Proposição 8 da Subseção 3.5 das notas de aula *Modos de Convergência de Sequências de Variáveis Aleatórias*.

A convergência em probabilidade pede apenas que a probabilidade de cada desvio tenda a zero; a somabilidade em (4) permite controlar sua repetição ao longo da sequência. Essa condição é suficiente para a convergência quase certa, mas não necessária, como mostram os exemplos da Subseção 3.6, *O Critério não é Necessário*, das notas anteriores. Nas duas provas da lei forte que apresentaremos, procuraremos estimativas que permitam aplicar o critério.

Na prova da recíproca, usaremos também o segundo lema de Borel-Cantelli: se os eventos A_n são independentes e $\sum_n \mathbb{P}(A_n) = \infty$, então

$$\mathbb{P}\Big(\limsup_{n\to\infty} A_n\Big) = 1.$$

Nesse caso, quase certamente infinitos dos eventos A_n ocorrem. Os dois lemas estão enunciados na Seção 2 das notas anteriores.

2.3. Esperança e Probabilidades de Cauda

Se W é uma variável aleatória não negativa, a fórmula da esperança por integração da cauda afirma que

$$\mathbb{E}[W] = \int_0^{\infty} \mathbb{P}(W > t)\, dt, \tag{5}$$

com ambos os lados podendo ser infinitos. De fato,

$$W(\omega) = \int_0^{\infty} \mathbb{1}_{(0,W(\omega))}(t)\, dt,$$

e o teorema de Tonelli permite trocar a esperança com a integral, pois o integrando é não negativo.

Como $t \mapsto \mathbb{P}(W > t)$ é não crescente, para todo inteiro $k \geqslant 0$ temos

$$\mathbb{P}(W > k+1) \leqslant \int_k^{k+1} \mathbb{P}(W > t)\, dt \leqslant \mathbb{P}(W > k).$$

Somando essas desigualdades e usando (5), obtemos

$$\sum_{k=1}^{\infty} \mathbb{P}(W > k) \leqslant \mathbb{E}[W] \leqslant \sum_{k=0}^{\infty} \mathbb{P}(W > k). \tag{6}$$

As séries dos extremos diferem apenas pelo termo $\mathbb{P}(W > 0) \leqslant 1$. Portanto, $\mathbb{E}[W] < \infty$ se, e somente se, $\sum_{k=1}^{\infty} \mathbb{P}(W > k) < \infty$. Aplicaremos essa equivalência a variável aleatória $W \equiv |X_1|$. A integrabilidade tornará raros os truncamentos na prova de Etemadi, enquanto sua ausência produzirá termos grandes infinitas vezes na prova da recíproca.

3. O Espaço Produto e as Frequências dos Dígitos

3.1. A Construção do Modelo

Trabalharemos no espaço de todas as sequências infinitas de dígitos,

$$\Omega = \{0, 1, \ldots, 9\}^{\mathbb{N}} = \Big\{\omega = (\omega_1, \omega_2, \omega_3, \ldots) : \omega_n \in \{0, 1, \ldots, 9\} \text{ para todo } n\Big\}.$$

As funções $X_n : \Omega \to \mathbb{R}$ dadas por

$$X_n(\omega) = \omega_n, \qquad n \geqslant 1,$$

podem ser pensadas como variáveis aleatórias sobre o espaço mensurável $(\Omega, \mathcal{F})$, onde $\mathcal{F}$ é a menor σ-álgebra, onde para todo $n \in \mathbb{N}$, cada função X_n é mensurável.

Um *cilindro de posto n* é um subconjunto de Ω obtido fixando as n primeiras coordenadas:

$$C(u_1, \ldots, u_n) = \Big\{\omega \in \Omega : X_1(\omega) = u_1, \ldots, X_n(\omega) = u_n\Big\},$$

com $u_1, \ldots, u_n \in \{0, 1, \ldots, 9\}$. É claro pela definição de $\mathcal{F}$ que qualquer cilindro da forma $C(u_1, \ldots, u_n) \in \mathcal{F}$. Basta observar que este cilindro corresponde ao evento

$$\{X_1 = u_1\} \cap \ldots \cap \{X_n = u_n\} \in \mathcal{F}.$$

A medida de probabilidade $\mathbb{P}$ que nos interessa sobre $(\Omega, \mathcal{F})$ é a única medida de probabilidade definida sobre o espaço mensurável $(\Omega, \mathcal{F})$ satisfazendo a seguinte propriedade

$$\mathbb{P}\Big(C(u_1, \ldots, u_n)\Big) = \frac{1}{10^n} = \prod_{j=1}^{n} \frac{1}{10} \tag{7}$$

Proposição 1. Para todo $n \geqslant 1$ e todo $k \in \{0, 1, \ldots, 9\}$,

$$\mathbb{P}(X_n = k) = \frac{1}{10}.$$

Prova. O evento $\{X_n = k\}$ é a união disjunta dos 10^{n-1} cilindros de posto n cuja n-ésima coordenada vale k, um para cada escolha das $n-1$ primeiras coordenadas. Por (7), cada um desses cilindros tem probabilidade 10^{-n}. Logo

$$\mathbb{P}(X_n = k) = 10^{n-1} \cdot 10^{-n} = \frac{1}{10}.$$

■

Proposição 2. A sequência $(X_n)_{n\geqslant 1}$ é independente. Mais precisamente, para todos $1 \leqslant n_1 < n_2 < \cdots < n_j$ e todos $k_1, \ldots, k_j \in \{0, 1, \ldots, 9\}$,

$$\mathbb{P}\Big(X_{n_1} = k_1, \ldots, X_{n_j} = k_j\Big) = \prod_{i=1}^{j} \mathbb{P}(X_{n_i} = k_i).$$

Prova. O evento $\{X_{n_1} = k_1, \ldots, X_{n_j} = k_j\}$ é a união disjunta dos cilindros de posto n_j compatíveis com as j condições. As coordenadas fora do conjunto $\{n_1, \ldots, n_j\}$ são livres, em número de $n_j - j$, de modo que há $10^{\,n_j - j}$ cilindros. Por (7),

$$\mathbb{P}\Big(X_{n_1} = k_1, \ldots, X_{n_j} = k_j\Big) = \frac{10^{\,n_j - j}}{10^{\,n_j}} = 10^{-j} = \prod_{i=1}^{j} \frac{1}{10},$$

e a última expressão é o produto das probabilidades marginais pela Proposição 1. ■

Cada evento $\{X_n \in B\}$, com $B \subseteq \{0, 1, \ldots, 9\}$, é uma união finita disjunta de eventos $\{X_n = k\}$. Somando a fatoração demonstrada acima, obtemos a independência para subconjuntos arbitrários e, portanto, para toda a sequência. O modelo é o de uma sequência de sorteios independentes de um dígito, cada um com probabilidade $1/10$. Ele é a versão de dez alternativas do modelo de lançamentos de uma moeda honesta e também pode ser obtido a partir da expansão decimal de um ponto escolhido ao acaso em $[0, 1]$, como veremos na Seção 4.3.

3.2. O Problema das Frequências

Para $n \geqslant 1$ e $k \in \{0, 1, \ldots, 9\}$, definimos a frequência relativa do dígito k entre os n primeiros dígitos:

$$\nu_{n,k}(\omega) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{\{X_j = k\}}(\omega). \tag{8}$$

A variável $\nu_{n,k}$ é a proporção das n primeiras posições ocupadas pelo dígito k. Pela Proposição 1, cada indicadora tem esperança $1/10$, logo

$$\mathbb{E}[\nu_{n,k}] = \frac{1}{10} \qquad \text{para todo } n.$$

Resta saber se a frequência relativa converge para esse valor ao longo da sequência, e em que sentido.

Teorema 3 (Teorema de Borel para a base 10)**.** Para cada $k \in \{0, 1, \ldots, 9\}$,

$$\lim_{n\to\infty} \nu_{n,k}(\omega) = \frac{1}{10} \qquad \text{para quase todo } \omega \in \Omega.$$

Além disso, a interseção dos conjuntos de convergência sobre os dez valores de k tem probabilidade 1, de modo que, quase certamente, todos os dez dígitos têm frequência assintótica $1/10$ simultaneamente.

A segunda afirmação é consequência da primeira e da subaditividade da probabilidade: a interseção de dez eventos de probabilidade 1 tem probabilidade 1. Registramos o resultado como um único teorema porque é assim que ele é usado.

Definição 4 (Sequências simplesmente normais e normais)**.** Um ponto $\omega \in \Omega$ é dito *simplesmente normal na base* 10 se, para todo $k \in \{0, 1, \ldots, 9\}$,

$$\lim_{n\to\infty} \nu_{n,k}(\omega) = \frac{1}{10}.$$

Dizemos que ω é *normal na base* 10 se, para toda palavra finita $u = u_1 \ldots u_\ell$ de dígitos com $\ell \geqslant 1$, a frequência assintótica de ocorrências do bloco u nas posições consecutivas de ω é igual a $10^{-\ell}$.

Na definição de normalidade, todas as posições iniciais são levadas em conta, permitindo ocorrências sobrepostas do bloco u. Enquanto a normalidade simples garante apenas o equilíbrio assintótico de cada dígito isolado, a normalidade plena impõe que qualquer sequência finita de dígitos ocorra exatamente com a frequência ditada pela probabilidade uniforme.

Com essa terminologia, o Teorema 3 afirma precisamente que o conjunto dos pontos simplesmente normais na base 10 tem probabilidade um, ou seja, uma sequência sorteada ao acaso satisfaz essa propriedade de equidistribuição.

O Teorema 3 decorre naturalmente da Lei Forte dos Grandes Números. Na Seção 4, demonstramos a lei forte sob a hipótese de quarto momento finito, que pode portanto ser aplicado ao caso de sequências de variáveis aleatórias limitadas, como é o caso das indicadoras dos dígitos. No entanto, a exigência de momentos de ordem quatro pode ser bastante relaxada. Na Seção 5, apresentamos a demonstração da lei forte geral sob a hipótese mínima de primeiro momento finito ($\mathbb{E}[|X_1|] < \infty$), devida a Etemadi. Mais adiante, na Seção 7, exploramos a força dessa generalidade ao estudar distribuições de cauda pesada, como a distribuição de Pareto com expoente em $(1, 2]$, cujas médias amostrais convergem quase certamente embora todos os momentos de ordem superior sejam infinitos.

Antes de prosseguir, vale registrar uma sutileza do modelo. A expansão decimal de um número real não é única: $0{,}4999\ldots = 0{,}5000\ldots$. O conjunto dos pontos com duas expansões é enumerável, logo tem probabilidade zero para a medida uniforme em $[0, 1]$, e a ambiguidade não afeta nenhuma afirmação quase certa. No espaço produto não há ambiguidade alguma, pois cada ω é uma sequência e não um número. Essa é uma das razões para trabalhar no espaço produto: ele elimina uma complicação que é irrelevante para o problema.

4. A Lei Forte com Quarto Momento

4.1. Enunciado e Prova do Teorema

Iniciamos nosso estudo da Lei Forte dos Grandes Números analisando o caso em que as variáveis aleatórias possuem quarto momento finito. Essa demonstração é particularmente instrutiva porque ilustra, em sua forma mais transparente e sem as dificuldades técnicas de truncamento, como o critério de Borel-Cantelli transforma estimativas de momentos em convergência quase certa.

Para compreender a estratégia que adotaremos, recordemos a dificuldade central encontrada na Seção 2.1: ao tentar estimar as probabilidades de desvio pela desigualdade de

Chebyshev (que se baseia no segundo momento e na variância), obtivemos uma cota proporcional a $1/n$. Como a série harmônica $\sum_{n=1}^{\infty} 1/n$ diverge, essa cota baseada no segundo momento não permite aplicar diretamente o primeiro lema de Borel-Cantelli.

A ideia fundamental para contornar essa limitação consiste em explorar momentos superiores. Aplicando a desigualdade de Markov à quarta potência do desvio, $(S_n - n\mu)^4$, mostraremos que a esperança correspondente cresce apenas com n^2. Dividindo por n^4, obteremos para a probabilidade de desvio uma cota superior proporcional a $1/n^2$. Como a série $\sum_{n=1}^{\infty} 1/n^2$ converge, o critério de Borel-Cantelli garantirá de imediato que os desvios ocorrem apenas um número finito de vezes, entregando a convergência quase certa.

Teorema 5 (Lei forte com quarto momento)**.** Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas com $\mu = \mathbb{E}[X_i]$ e $\mathbb{E}[X_i^4] < \infty$. Se $S_n = X_1 + \cdots + X_n$, então

$$\frac{S_n}{n} \longrightarrow \mu \qquad \text{quase certamente.}$$

Prova. Substituindo X_i por $X_i - \mu$, podemos supor sem perda de generalidade que $\mu = 0$. De fato, se $X_i' = X_i - \mu$, então $S_n' = S_n - n\mu$ e a convergência $S_n'/n \to 0$ é equivalente a $S_n/n \to \mu$; além disso, as variáveis X_i' continuam independentes, identicamente distribuídas e com quarto momento finito.

Suponha então $\mu = 0$. Expandindo o quarto momento da soma e usando a propriedade de linearidade da esperança ficamos com

$$\mathbb{E}[S_n^4] = \mathbb{E}\Big[\Big(\sum_{i=1}^{n} X_i\Big)^4\Big] = \sum_{i,j,k,\ell=1}^{n} \mathbb{E}[X_i X_j X_k X_\ell]. \tag{9}$$

Vamos classificar as parcelas de acordo com o número de índices distintos.

Se algum índice aparece uma única vez entre i, j, k, ℓ, a parcela se anula. Por exemplo, se i, j, k são distintos e $\ell = i$, então $\mathbb{E}[X_i^2 X_j X_k] = \mathbb{E}[X_i^2]\,\mathbb{E}[X_j]\,\mathbb{E}[X_k] = 0$, porque $\mathbb{E}[X_j] = 0$ e as variáveis são independentes. O mesmo argumento se aplica a $\mathbb{E}[X_i^3 X_j]$ e a $\mathbb{E}[X_i X_j X_k X_\ell]$ com os quatro índices distintos.

Restam duas formas. A primeira é a dos termos com os quatro índices iguais, $\mathbb{E}[X_i^4]$, que ocorrem n vezes, uma para cada escolha de i. A segunda é a dos termos com dois pares de índices iguais, $\mathbb{E}[X_i^2 X_j^2] = (\mathbb{E}[X_i^2])^2$ com $i \neq j$. Para contá-los, escolhemos o par $\{i, j\}$ de $\binom{n}{2}$ maneiras e, uma vez fixado o par, o termo $X_i^2 X_j^2$ aparece em 6 posições distintas, correspondentes às permutações de i, i, j, j. Logo há

$$6\binom{n}{2} = 3n(n-1)$$

tais termos. Como as variáveis são identicamente distribuídas,

$$\mathbb{E}[S_n^4] = n\,\mathbb{E}[X_1^4] + 3n(n-1)\Big(\mathbb{E}[X_1^2]\Big)^2 \leqslant Cn^2, \tag{10}$$

com

$$C = \mathbb{E}[X_1^4] + 3\Big(\mathbb{E}[X_1^2]\Big)^2 < +\infty.$$

Observe que C não depende de n.

Fixe $\varepsilon > 0$. Aplicamos agora a desigualdade de Markov à variável não negativa S_n^4, com nível $\varepsilon^4 n^4$:

$$\mathbb{P}\Big(\Big|\frac{S_n}{n}\Big| \geqslant \varepsilon\Big) = \mathbb{P}\big(|S_n| \geqslant \varepsilon n\big) = \mathbb{P}\big(S_n^4 \geqslant \varepsilon^4 n^4\big) \leqslant \frac{\mathbb{E}[S_n^4]}{\varepsilon^4 n^4} \leqslant \frac{C}{\varepsilon^4 n^2}.$$

Somando em n,

$$\sum_{n=1}^{\infty} \mathbb{P}\Big(\Big|\frac{S_n}{n} - 0\Big| \geqslant \varepsilon\Big) \leqslant \frac{C}{\varepsilon^4} \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty. \tag{11}$$

Como essa série converge para todo $\varepsilon > 0$, segue diretamente do critério de Borel-Cantelli (Seção 2.2) que $S_n/n \to 0$ quase certamente. ■

Observação 6 (Por que o quarto momento?)**.** Poderíamos indagar se a mesma demonstração elementar funcionaria assumindo apenas a existência de momentos de ordem inferior, como o terceiro momento ($\mathbb{E}[|X_1|^3] < \infty$). A resposta é negativa, e a estratégia empregada quebra por duas razões estruturais:

(i) **Paridade da potência e não negatividade:** a função $x \mapsto x^4$ é par e não negativa, o que permitiu transformar diretamente o evento de desvio $\{|S_n| \geqslant \varepsilon n\}$ em $\{S_n^4 \geqslant \varepsilon^4 n^4\}$ e aplicar a desigualdade de Markov à variável não negativa S_n^4. Com o terceiro momento, a variável S_n^3 assume valores negativos, e para aplicar Markov seríamos forçados a considerar a variável não negativa $|S_n|^3$.

(ii) **Álgebra da independência:** para calcular $\mathbb{E}[S_n^4]$, expandimos a potência polinomialmente e usamos que a independência anula todas as parcelas com índices ímpares isolados. Em contrapartida, a expressão $|S_n|^3 = |\sum_{i=1}^n X_i|^3$ não admite expansão em produtos algébricos simples das variáveis X_i, destruindo o mecanismo combinatório elementar que zerou a maioria dos termos cruzados.

Portanto, o expoente $p = 4$ é o menor inteiro par em que esse método algébrico direto produz uma cota somável (de ordem $1/n^2$). Para $p = 2$, a cota de Chebyshev decai com taxa $1/n$ e não é somável. Já para contornar a exigência de momentos superiores e alcançar a hipótese ótima $\mathbb{E}[|X_1|] < \infty$, precisaremos de técnicas analíticas mais profundas, como o método de Etemadi que apresentaremos na Seção 5.

Corolário 7 (Variáveis aleatórias limitadas)**.** Se (X_n) é uma sequência de variáveis aleatórias independente e identicamente distribuída com $|X_n| \leqslant M$ quase certamente, e $\mu = \mathbb{E}[X_1]$, então

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{} \mu, \qquad \text{quase certamente.}$$

Prova. A hipótese de limitação implica imediatamente $\mathbb{E}[X_1^4] \leqslant M^4 < \infty$. Como as demais hipóteses do Teorema 5 também são satisfeitas o resultado segue. ■

4.2. Aplicação às Frequências dos Dígitos

Agora combinamos o Teorema 5 com o Corolário 7 para demonstrar o Teorema 3. Fixe $k \in \{0, 1, \ldots, 9\}$ e defina

$$Y_j = \mathbb{1}_{\{X_j = k\}} - \frac{1}{10}, \qquad j \geqslant 1.$$

Pela Proposição 2, as variáveis Y_j são independentes e identicamente distribuídas. Além disso,

$$\mathbb{E}[Y_j] = \frac{1}{10} - \frac{1}{10} = 0, \qquad |Y_j| \leqslant 1,$$

de modo que $\mathbb{E}[Y_j^4] \leqslant 1 < \infty$. Pelo Corolário 7,

$$\frac{1}{n}\sum_{j=1}^{n} Y_j \longrightarrow 0 \qquad \text{quase certamente.} \tag{12}$$

Mas

$$\frac{1}{n}\sum_{j=1}^{n} Y_j = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{X_j=k\}} - \frac{1}{10} = \nu_{n,k} - \frac{1}{10},$$

logo (12) é exatamente a afirmação $\nu_{n,k} \to 1/10$ quase certamente. Como k é arbitrário e a interseção sobre os dez valores é enumerável, o Teorema 3 segue.

Vale registrar o cálculo explícito da variância, porque ele mostra por que a aplicação direta de Chebyshev seguida de Borel-Cantelli não conclui. Para cada j,

$$\operatorname{Var}(Y_j) = \mathbb{P}(X_j = k)\big(1 - \mathbb{P}(X_j = k)\big) = \frac{1}{10}\cdot\frac{9}{10} = \frac{9}{100}.$$

Como as Y_j são independentes,

$$\operatorname{Var}\Big(\frac{1}{n}\sum_{j=1}^{n} Y_j\Big) = \frac{9}{100n}.$$

A cota de Chebyshev é, portanto,

$$\mathbb{P}\big(|\nu_{n,k} - \tfrac{1}{10}| \geqslant \varepsilon\big) \leqslant \frac{9}{100n\varepsilon^2},$$

que tende a zero e dá a convergência em probabilidade, mas não é somável. É precisamente a passagem ao quarto momento que produz a somabilidade, como no Teorema 5. Esse é o conteúdo do exercício sobre as frequências dos dígitos proposto nas notas de aula sobre modos de convergência.

O argumento acima demonstrou o Teorema 3 sem usar a versão geral da lei forte. A prova vale porque as indicadoras são limitadas. Para uma sequência geral de variáveis independentes e identicamente distribuídas com $\mathbb{E}[|X_1|] < \infty$ e sem quarto momento, o argumento não se aplica; para tratar esse caso, usaremos o método da Seção 5.

4.3. Dos Dígitos aos Números Reais

Na subseção anterior, demonstramos que uma sequência de dígitos sorteados de forma independente, com os dez dígitos equiprováveis, tem, quase certamente, frequência $1/10$ para cada dígito. Queremos agora obter uma afirmação sobre os dígitos de um número real escolhido ao acaso. Para isso, precisamos justificar a ligação entre os dois experimentos: sortear os dígitos e escolher um ponto do intervalo.

A ligação vem da seguinte observação: *fixar os primeiros n dígitos decimais de um número equivale a localizá-lo em um intervalo de comprimento* 10^{-n}. Vamos primeiro entender essa observação. Em seguida, usaremos comprimentos de intervalos para verificar a independência dos dígitos e aplicar a lei forte.

4.3.1. Um número escolhido uniformemente

Denotaremos por λ a medida de Lebesgue, que associa a cada intervalo seu comprimento. Como $\lambda([0,1)) = 1$, sua restrição a $[0,1)$ é uma medida de probabilidade. Escolher um ponto uniformemente nesse intervalo significa justamente usar essa probabilidade: a chance de o ponto escolhido pertencer a $[a,b) \subseteq [0,1)$ é

$$\lambda([a,b)) = b - a.$$

Trabalharemos primeiro em $[0,1)$; ao final, incluiremos o extremo 1.

Para cada $x \in [0,1)$, escrevemos

$$x = 0{,}d_1(x)d_2(x)d_3(x)\ldots = \sum_{j=1}^{\infty} \frac{d_j(x)}{10^j}, \qquad d_j(x) \in \{0,1,\ldots,9\}.$$

Aqui, $d_j(x)$ é o dígito que ocupa a j-ésima posição depois da vírgula. Adotamos sempre a expansão que não tem uma cauda infinita de noves. Por exemplo, entre as duas representações

$$0{,}37000\ldots = 0{,}36999\ldots,$$

usaremos a primeira. Essa convenção determina os dígitos de cada ponto sem ambiguidade.

4.3.2. Dígitos prescritos e intervalos decimais

Comecemos com o primeiro dígito. Dividindo $[0,1)$ em dez intervalos de mesmo comprimento, obtemos

$$\left[0, \frac{1}{10}\right), \quad \left[\frac{1}{10}, \frac{2}{10}\right), \quad \ldots, \quad \left[\frac{9}{10}, 1\right).$$

Esses intervalos correspondem, nessa ordem, aos dez valores possíveis de $d_1(x)$. Em particular, $d_1(x) = 3$ se, e somente se, $x \in \left[\frac{3}{10}, \frac{4}{10}\right)$. Para determinar o segundo dígito, dividimos cada um desses intervalos em dez partes iguais, e repetimos o procedimento para os dígitos seguintes. Por exemplo,

$$\{x : d_1(x) = 3\} = \left[\frac{3}{10}, \frac{4}{10}\right),$$

$$\{x : d_1(x) = 3,\ d_2(x) = 7\} = \left[\frac{37}{100}, \frac{38}{100}\right),$$

$$\{x : d_1(x) = 3,\ d_2(x) = 7,\ d_3(x) = 2\} = \left[\frac{372}{1000}, \frac{373}{1000}\right).$$

Cada novo dígito especificado reduz o comprimento do intervalo por um fator 10, como ilustra a Figura 1. Os extremos esquerdos estão incluídos e os direitos estão excluídos por causa da convenção adotada: por exemplo, $0{,}38 = 0{,}38000\ldots$ tem segundo dígito 8 e, portanto, não pertence ao intervalo dos números que começam por 0,37.

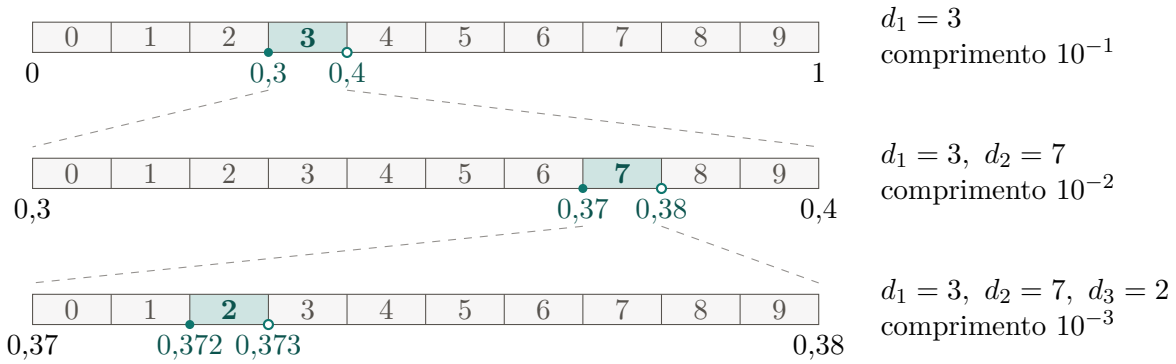


Figura 1: As linhas usam escalas distintas: o intervalo destacado em cada linha é ampliado na seguinte. Os números nas dez partes indicam o próximo dígito; os pontos cheios e vazios indicam se os extremos estão incluídos ou excluídos.

O mesmo raciocínio vale para quaisquer dígitos $u_1, \ldots, u_n$. O número

$$a = \frac{u_1}{10} + \frac{u_2}{10^2} + \cdots + \frac{u_n}{10^n}$$

é o extremo esquerdo do intervalo correspondente. Como fizemos n divisões sucessivas em dez partes iguais, esse intervalo tem comprimento 10^{-n}. Portanto,

$$I(u_1, \ldots, u_n) \equiv \{x \in [0,1) : d_1(x) = u_1, \ldots, d_n(x) = u_n\} = [a, a + 10^{-n}). \tag{13}$$

Em particular,

$$\lambda\Big(I(u_1, \ldots, u_n)\Big) = 10^{-n} = \mathbb{P}\Big(C(u_1, \ldots, u_n)\Big). \tag{14}$$

Assim, uma condição sobre os primeiros dígitos tem a mesma probabilidade nos dois experimentos: no espaço de sequências, ela determina um cilindro; no intervalo $[0,1)$, ela determina um intervalo decimal.

4.3.3. Da independência à frequência dos dígitos

Para aplicar a lei forte às funções d_j, precisamos verificar que elas são variáveis aleatórias independentes e identicamente distribuídas. A aleatoriedade está na escolha de x: depois de escolhido o ponto, todos os seus dígitos ficam determinados. A igualdade (14) já trata dos primeiros n dígitos. A independência exige também considerar posições que não sejam consecutivas. Podemos fazer isso deixando livres os dígitos das posições não especificadas.

Lema 8 (Os dígitos de um ponto uniforme)**.** No intervalo $[0,1)$ com a probabilidade λ, as funções $d_1, d_2, \ldots$ são variáveis aleatórias independentes e identicamente distribuídas, com

$$\lambda\Big(\{x : d_j(x) = k\}\Big) = \frac{1}{10} \qquad \text{para todo } j \geqslant 1 \text{ e } k \in \{0, 1, \ldots, 9\}.$$

Prova. Vejamos primeiro um exemplo com uma posição livre. A condição $d_1(x) = 3$, $d_3(x) = 2$ permite que o segundo dígito seja qualquer um dos dez dígitos. Logo, o evento correspondente é a união disjunta dos dez intervalos

$$I(3,0,2), \quad I(3,1,2), \quad \ldots, \quad I(3,9,2).$$

Cada um tem comprimento 10^{-3}. A probabilidade dessa união é, portanto, $10 \cdot 10^{-3} = 10^{-2}$.

Em geral, fixamos r posições $1 \leqslant n_1 < \cdots < n_r$ e os valores dos dígitos nessas posições, digamos $k_1, \ldots, k_r$. Agora, damos uma atenção especial para a última posição especificada, $N = n_r$. Entre as primeiras N posições, há r fixadas e $N - r$ livres. Como cada posição livre admite dez escolhas, há 10^{N-r} sequências de N dígitos compatíveis com as condições. Cada uma determina um intervalo de comprimento 10^{-N}, e esses intervalos são disjuntos. Somando seus comprimentos, obtemos

$$\lambda\Big(\{x : d_{n_1}(x) = k_1, \ldots, d_{n_r}(x) = k_r\}\Big) = 10^{N-r} \cdot 10^{-N} = 10^{-r}. \tag{15}$$

Essa descrição mostra também que os eventos considerados são mensuráveis, pois são uniões finitas de intervalos. Em particular, cada d_j é uma variável aleatória, já que assume apenas dez valores e cada conjunto $\{x : d_j(x) = k\}$ é mensurável.

Tomando $r = 1$ em (15), vemos que cada d_j tem distribuição uniforme em $\{0, 1, \ldots, 9\}$. Para r qualquer, a mesma igualdade fornece

$$\lambda\Big(\{d_{n_1} = k_1, \ldots, d_{n_r} = k_r\}\Big) = \prod_{i=1}^{r} \lambda\Big(\{d_{n_i} = k_i\}\Big).$$

Como cada d_j assume um número finito de valores, a fatoração para conjuntos de valores se obtém somando essas igualdades. Isso prova a independência. ■

O lema justifica o uso do modelo de sorteios independentes para estudar os dígitos de um ponto uniforme. Agora podemos repetir a aplicação da lei forte feita na subseção anterior, usando a medida λ como probabilidade. Nesse espaço, dizer que uma propriedade vale *quase certamente* significa que o conjunto dos pontos onde ela falha tem medida de Lebesgue zero. É exatamente esse o sentido de dizer que ela vale para *quase todo* ponto.

Teorema 9 (Borel)**.** Para quase todo $x \in [0, 1]$, a frequência relativa de cada dígito na expansão decimal de x tende a $1/10$. Em outras palavras, quase todo número real é simplesmente normal na base 10.

Prova. Fixe um dígito $k \in \{0, 1, \ldots, 9\}$. Para contar suas ocorrências, definimos, no espaço $[0, 1)$,

$$Z_j(x) = \mathbb{1}_{\{d_j = k\}}(x) = \begin{cases} 1, & \text{se } d_j(x) = k, \\ 0, & \text{se } d_j(x) \neq k. \end{cases}$$

Pelo Lema 8, as variáveis Z_j são independentes e identicamente distribuídas. Além disso, são limitadas por 1 e têm esperança

$$\int_{[0,1)} Z_j(x)\, d\lambda(x) = \lambda\Big(\{x : d_j(x) = k\}\Big) = \frac{1}{10}.$$

O Corolário 7 se aplica e fornece

$$\frac{\#\{1 \leqslant j \leqslant n : d_j(x) = k\}}{n} = \frac{1}{n} \sum_{j=1}^{n} Z_j(x) \longrightarrow \frac{1}{10} \qquad \text{para } \lambda\text{-quase todo } x \in [0, 1).$$

O numerador conta quantas vezes o dígito k aparece nas primeiras n posições; portanto, a razão é precisamente sua frequência relativa.

Para cada k, o conjunto excepcional tem medida zero. A união desses dez conjuntos também tem medida zero, de modo que, fora dela, as dez frequências convergem simultaneamente para $1/10$. Incluir o ponto 1 não altera a afirmação, pois $\lambda(\{1\}) = 0$. A escolha entre as duas expansões dos decimais finitos tampouco afeta o resultado: os pontos que admitem duas expansões formam um conjunto enumerável e, por isso, de medida zero.

Finalmente, para um número real x, consideramos os dígitos de sua parte fracionária $x - \lfloor x \rfloor \in [0, 1)$. Se $N \subseteq [0, 1)$ é o conjunto excepcional obtido acima, o conjunto excepcional em toda a reta é

$$\bigcup_{m \in \mathbb{Z}} (m + N).$$

Cada translação $m + N$ tem medida zero, e a união é enumerável. Isso prova a afirmação em $\mathbb{R}$. ■

5. A Prova de Etemadi da Lei Forte

5.1. Enunciado e Visão Geral

O resultado central destas notas é o seguinte teorema, devido a Etemadi. Ele estabelece a lei forte na hipótese mínima sobre os momentos e necessita apenas da hipótese mais fraca de independência dois a dois, e não de independência mútua.

Teorema 10 (Lei forte dos grandes números)**.** Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes dois a dois e identicamente distribuídas, com $\mathbb{E}[|X_i|] < \infty$. Seja $\mu = \mathbb{E}[X_i]$ e $S_n = X_1 + \cdots + X_n$. Então

$$\frac{S_n}{n} \longrightarrow \mu \qquad \text{quase certamente.}$$

Antes de apresentar os detalhes da prova, vale entender como ela amplia o resultado anterior. O Teorema 5 exige $\mathbb{E}[X_1^4] < \infty$, enquanto o Teorema 10 requer apenas $\mathbb{E}[|X_1|] < \infty$. Para ver concretamente a diferença entre essas hipóteses, considere uma variável aleatória X_1 com densidade

$$f(x) = \begin{cases} \dfrac{3}{2x^{5/2}}, & \text{se } x \geqslant 1, \\[2ex] 0, & \text{se } x < 1. \end{cases}$$

Essa é a distribuição de Pareto com escala 1 e parâmetro de forma $3/2$, que estudaremos na Seção 7.4. Abaixo mostramos o esboço do gráfico desta função densidade na Figura 2.

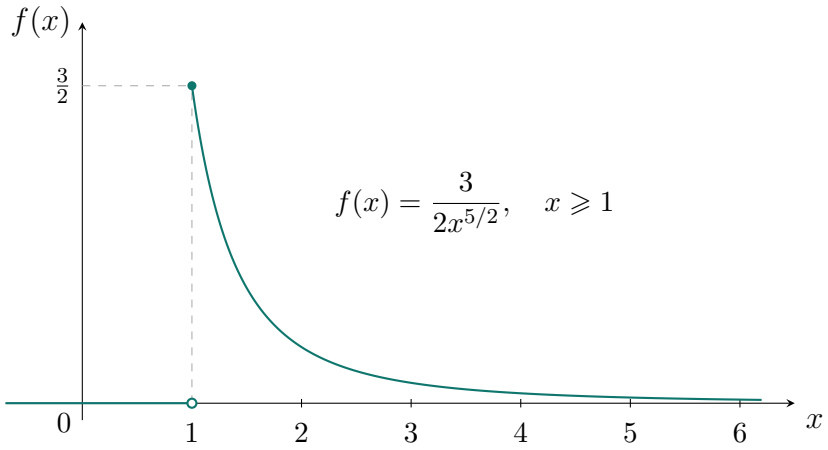


Figura 2: Densidade de Pareto com escala 1 e parâmetro de forma 3/2.

Como $X_1 \geqslant 1$ quase certamente, o cálculo dos dois momentos fornece

$$\mathbb{E}[|X_1|] = \int_1^\infty x \, \frac{3}{2x^{5/2}} \, dx = \frac{3}{2} \int_1^\infty x^{-3/2} \, dx = 3 < \infty,$$

$$\mathbb{E}[X_1^4] = \int_1^\infty x^4 \, \frac{3}{2x^{5/2}} \, dx = \frac{3}{2} \int_1^\infty x^{3/2} \, dx = \infty.$$

Portanto, para uma sequência independente e identicamente distribuída com essa distribuição, o Teorema 10 garante a convergência quase certa das médias amostrais para 3, embora a hipótese do Teorema 5 não seja satisfeita.

A prova de Etemadi substitui o quarto momento por uma estimativa de momentos das variáveis truncadas, cujo nível de truncamento cresce com o índice. Com apenas a integrabilidade, não temos garantia de que a cota de Chebyshev seja somável em todos os índices. A estimativa abaixo garante a somabilidade ao longo de índices de crescimento geométrico, e a monotonicidade permite controlar os índices restantes.

A prova tem quatro etapas, que enunciamos antes de executá-las:

(i) *Redução a variáveis não negativas.* Decompomos $X_n = X_n^+ - X_n^-$ e tratamos cada parte separadamente.

(ii) *Truncamento.* Substituímos X_k por $Y_k = X_k \mathbb{1}_{\{|X_k| \leqslant k\}}$ e mostramos que basta provar a lei forte para as variáveis truncadas; em seguida, identificamos o limite das médias de suas esperanças.

(iii) *Estimativa de momentos.* Mostramos que $\sum_k \operatorname{Var}(Y_k)/k^2 < \infty$.

(iv) *Subsequência geométrica e monotonicidade.* Aplicamos Chebyshev ao longo dos índices $k(n) = \lfloor \alpha^n \rfloor$, concluímos por Borel-Cantelli, e usamos a monotonicidade da soma truncada para controlar os índices intermediários.

No livro de Durrett, a redução a variáveis não negativas aparece depois das estimativas de momentos. Nestas notas, apresentamos essa redução primeiro para destacar a propriedade que será usada na etapa final: as somas parciais de variáveis não negativas são não decrescentes. Essa monotonicidade permitirá estender a convergência das médias, obtida na subsequência geométrica, a todos os índices.

5.2. Redução a Variáveis Não Negativas

Para $x \in \mathbb{R}$, escrevemos $x^+ = \max\{x, 0\}$ e $x^- = \max\{-x, 0\}$, de modo que $x = x^+ - x^-$ e $|x| = x^+ + x^-$. Definimos

$$X_n^+ = \max\{X_n, 0\}, \qquad X_n^- = \max\{-X_n, 0\}.$$

As duas famílias (X_n^+) e (X_n^-) são formadas por variáveis não negativas, independentes dois a dois e identicamente distribuídas. Além disso, $0 \leqslant X_n^{\pm} \leqslant |X_n|$, logo $\mathbb{E}[X_n^{\pm}] \leqslant \mathbb{E}[|X_n|] < \infty$.

Suponha que o Teorema 10 esteja demonstrado para variáveis não negativas. Aplicando-o a cada família, obtemos, quase certamente,

$$\frac{1}{n}\sum_{k=1}^{n} X_k^+ \longrightarrow \mathbb{E}[X_1^+], \qquad \frac{1}{n}\sum_{k=1}^{n} X_k^- \longrightarrow \mathbb{E}[X_1^-].$$

Os dois conjuntos de convergência têm probabilidade 1, logo a interseção também tem probabilidade 1. Em cada ponto dessa interseção, subtraindo as os dois limites, obtemos

$$\frac{S_n}{n} = \frac{1}{n}\sum_{k=1}^{n} X_k^+ - \frac{1}{n}\sum_{k=1}^{n} X_k^- \longrightarrow \mathbb{E}[X_1^+] - \mathbb{E}[X_1^-] = \mathbb{E}[X_1] = \mu.$$

Portanto, basta provar o teorema sob a hipótese adicional $X_n \geqslant 0$ para todo n. Nesse caso, $\mu = \mathbb{E}[X_1] \geqslant 0$ e as somas truncadas serão não decrescentes, o que será usado na Seção 5.7.

5.3. O Truncamento

Pelo argumento dado acima, podemos supor que $X_k \geqslant 0$ para todo k. É claro que a hipótese de integrabilidade não é suficiente para garantir que essas variáveis tenham segundo momento finito, uma condição necessária para obter uma cota finita pela desigualdade de Chebyshev. Por isto vamos substituir cada v.a. X_k por uma variável aleatória limitada que será denotada por Y_k que é definida a partir de um truncamento de X_k e as somas parciais destas novas variáveis serão denotadas por T_n. Mais precisamente, definimos

$$Y_k \equiv X_k \cdot \mathbb{1}_{\{|X_k| \leqslant k\}}, \qquad T_n \equiv Y_1 + \cdots + Y_n. \tag{16}$$

Assim, Y_k coincide com X_k ao longo de $\{X_k \leqslant k\}$ e vale zero no evento $\{X_k > k\}$. Como $0 \leqslant Y_k \leqslant k$, temos $\mathbb{E}[Y_k^2] \leqslant k^2 < \infty$. A independência dois a dois também é preservada, pois cada Y_k é uma função boreliana da respectiva variável X_k. Como o nível de truncamento varia com k, as variáveis Y_k não são, em geral, identicamente distribuídas.

A escolha do nível k é motivado pelo critério de comparação (6) que está relacionado a integrabilidade de X_1 via a convergência da série $\sum_{k=1}^{\infty} \mathbb{P}(X_1 > k)$. Como X_k e X_1 têm a mesma distribuição, temos

$$\mathbb{P}(X_k \neq Y_k) = \mathbb{P}(X_k > k) = \mathbb{P}(X_1 > k).$$

O primeiro lema de Borel-Cantelli permitirá mostrar que, quase certamente, o truncamento modifica apenas um número finito de termos. A contribuição dessas modificações para a diferença entre as médias tende a zero, como afirma o lema.

Lema 11 (Truncamento)**.** Assuma que são válidas as hipóteses do Teorema 10 e sejam Y_k e T_n as variáveis aleatórias definidos em (16). Então

$$\frac{S_n - T_n}{n} \longrightarrow 0 \qquad \text{quase certamente.}$$

Em particular, $S_n/n \to \mu$ quase certamente se, e somente se, $T_n/n \to \mu$ quase certamente.

Prova. Vamos mostrar que, quase certamente, o truncamento altera apenas um número finito de termos. Os eventos correspondentes são

$$A_k \equiv \{X_k \neq Y_k\} = \{|X_k| > k\}.$$

Como as variáveis são identicamente distribuídas, $\mathbb{P}(|X_k| > k) = \mathbb{P}(|X_1| > k)$. A fórmula de esperança por integração da cauda e a comparação (6) dão a seguinte estimativa (a soma é de termos não negativos):

$$\sum_{k=1}^{\infty} \mathbb{P}(|X_k| > k) = \sum_{k=1}^{\infty} \mathbb{P}(|X_1| > k) \leqslant \int_0^{\infty} \mathbb{P}(|X_1| > t)\, dt = \mathbb{E}[|X_1|]. \tag{17}$$

Por hipótese, o lado direito é finito.

Por (17) e pelo primeiro lema de Borel-Cantelli,

$$\mathbb{P}\Big(\limsup_{k\to\infty} A_k\Big) = 0.$$

Fora desse conjunto nulo, o conjunto dos índices k para os quais $X_k \neq Y_k$ é finito. Para tais ω, definimos

$$K(\omega) = \max\{k : X_k(\omega) \neq Y_k(\omega)\}, \qquad R(\omega) = \sum_{k=1}^{K(\omega)} \Big|X_k(\omega) - Y_k(\omega)\Big|,$$

com $K(\omega) = 0$ e $R(\omega) = 0$ quando o conjunto é vazio. Como cada diferença é finita e o máximo é tomado sobre um conjunto finito, $R(\omega) < \infty$.

Para $k > K(\omega)$ temos $Y_k(\omega) = X_k(\omega)$. Logo, para todo $n \geqslant 1$,

$$\Big|S_n(\omega) - T_n(\omega)\Big| = \Big|\sum_{k=1}^{n}\Big(X_k(\omega) - Y_k(\omega)\Big)\Big| \leqslant \sum_{k=1}^{K(\omega)} \Big|X_k(\omega) - Y_k(\omega)\Big| = R(\omega).$$

Dividindo por n,

$$\Big|\frac{S_n(\omega)}{n} - \frac{T_n(\omega)}{n}\Big| \leqslant \frac{R(\omega)}{n} \longrightarrow 0.$$

Isso prova que $(S_n - T_n)/n \to 0$ quase certamente. Consequentemente, $S_n/n \to \mu$ quase certamente se, e somente se, $T_n/n \to \mu$ quase certamente. ■

A cota obtida é uniforme em n, e é isso que permite dividir por n sem discutir subsequências. O ponto é que $K(\omega)$ é o *último* índice em que o truncamento interfere; depois dele, as diferenças são exatamente nulas e a soma parcial para de crescer. Uma cota dependente de n exigiria verificar adicionalmente que ela tende a zero após a divisão por n.

5.4. As Médias Truncadas

Antes de estimar as flutuações de T_n, identificamos o limite de suas médias. Como já reduzimos ao caso $X_k \geqslant 0$, a variável

$$Z_k = X_1 \mathbb{1}_{\{X_1 \leqslant k\}}$$

cresce para X_1 quando $k \to \infty$. Pelo Teorema da Convergência Monótona,

$$\mathbb{E}[Z_k] \longrightarrow \mathbb{E}[X_1] = \mu.$$

Para transferir essa convergência para as esperanças de Y_k, vamos mostrar que, para cada k fixado, as variáveis

$$Y_k = X_k \mathbb{1}_{\{X_k \leqslant k\}} \qquad \text{e} \qquad Z_k = X_1 \mathbb{1}_{\{X_1 \leqslant k\}}$$

têm a mesma distribuição. Nas duas definições, usamos o mesmo nível de truncamento k. Assim, aplicamos a mesma regra a X_k e a X_1, que têm a mesma distribuição. Para verificar que Y_k e Z_k também têm a mesma distribuição, basta comparar suas funções de distribuição.

Como $0 \leqslant Y_k \leqslant k$ e $0 \leqslant Z_k \leqslant k$, para $t < 0$ temos $\mathbb{P}(Y_k \leqslant t) = \mathbb{P}(Z_k \leqslant t) = 0$, e para $t \geqslant k$ essas duas probabilidades são iguais a 1. Resta considerar $0 \leqslant t < k$. Nesse caso, $Y_k \leqslant t$ ocorre quando $X_k \leqslant t$ ou quando $X_k > k$, pois, nesta última situação, o truncamento atribui a Y_k o valor zero. Portanto,

$$\{Y_k \leqslant t\} = \{X_k \leqslant t\} \cup \{X_k > k\}.$$

Os dois eventos à direita são disjuntos, e a mesma decomposição vale para Z_k, substituindo X_k por X_1. Como X_k e X_1 têm a mesma distribuição, obtemos

$$\begin{aligned} \mathbb{P}(Y_k \leqslant t) &= \mathbb{P}(X_k \leqslant t) + \mathbb{P}(X_k > k) \\ &= \mathbb{P}(X_1 \leqslant t) + \mathbb{P}(X_1 > k) \\ &= \mathbb{P}(Z_k \leqslant t). \end{aligned}$$

Assim, as funções de distribuição de Y_k e Z_k coincidem em todo $t \in \mathbb{R}$, o que prova a igualdade em distribuição. Como essas variáveis são limitadas, suas esperanças são finitas e iguais. Logo,

$$\mathbb{E}[Y_k] = \mathbb{E}[Z_k] \longrightarrow \mu.$$

Usaremos agora a seguinte propriedade das sequências numéricas: se $a_m \to a$, então suas médias aritméticas também convergem para a, isto é,

$$\frac{a_1 + \cdots + a_n}{n} \longrightarrow a.$$

Essa propriedade é conhecida como convergência de Cesàro. A ideia é que a contribuição de um número finito de termos iniciais para a média tende a zero após a divisão por n; a partir de certo índice, todos os demais termos estão tão próximos de a quanto quisermos.

Aplicamos essa propriedade à sequência $a_m = \mathbb{E}[Y_m]$, cujo limite é μ. Pela linearidade da esperança, obtemos

$$\frac{\mathbb{E}[T_n]}{n} = \frac{1}{n} \sum_{m=1}^{n} \mathbb{E}[Y_m] \longrightarrow \mu. \tag{18}$$

Essa identificação vale para a sequência inteira; mais adiante, será usada também nos índices da subsequência geométrica.

5.5. A Estimativa de Momentos

Na subseção anterior, mostramos que $\mathbb{E}[T_n]/n \to \mu$. Resta controlar a diferença entre a média truncada T_n/n e sua esperança, isto é, provar que $(T_n - \mathbb{E}[T_n])/n \to 0$ quase certamente. Como as variáveis aleatórias Y_k têm segundo momento finito e são duas-a-duas independentes, a variância de T_n é a soma das variâncias dos seus termos. Portanto, para todo $\varepsilon > 0$, a desigualdade de Chebyshev fornece

$$\mathbb{P}\left(\left|\frac{T_n - \mathbb{E}[T_n]}{n}\right| > \varepsilon\right) \leqslant \frac{1}{\varepsilon^2 n^2} \sum_{k=1}^{n} \operatorname{Var}(Y_k). \tag{19}$$

A expressão acima sugere que a prova do teorema poderia ser finalizada se fosse possível aplicar o critério de Borel-Cantelli. O problema é justamente encontrar uma cota superior somável para a expressão que aparece no lado direito da desigualdade (19). A grande ideia de Etemadi foi perceber dois fatos fundamentais: que a convergência quase certa da sequência

$$\frac{T_{k(n)}}{k(n)}, \tag{20}$$

onde $k(n)$ é uma sequência apropriada de índices, implica na convergência quase certa de toda a sequência T_n/n e; que a convergência quase certa da sequência (20) poderia ser obtida por (19), escolhendo $k(n)$ apropriadamente.

Antes de desenvolver a ideia de Etemadi precisamos de alguns lemas técnicos.

Lema 12. Para todo $y \geqslant 0$,

$$2y \sum_{k>y} \frac{1}{k^2} \leqslant 4.$$

Prova. Começamos pela comparação com a integral. Se $m \geqslant 2$, a função $x \mapsto x^{-2}$ é decrescente em $[m-1, \infty)$. Para $x \in [k-1, k]$ com $k \geqslant m$, temos $x \leqslant k$, logo $x^{-2} \geqslant k^{-2}$ e

$$\frac{1}{k^2} \leqslant \int_{k-1}^{k} \frac{1}{x^2}\, dx.$$

Somando em $k \geqslant m$,

$$\sum_{k \geqslant m} \frac{1}{k^2} \leqslant \int_{m-1}^{\infty} \frac{dx}{x^2} = \frac{1}{m-1}. \tag{21}$$

O restante da prova é divido em dois casos. Suponha inicialmente que $y \geqslant 1$. Neste caso, a soma $\sum_{k>y} k^{-2}$ começa em $k = \lfloor y \rfloor + 1 \geqslant 2$, de modo que (21) se aplica com $m = \lfloor y \rfloor + 1$, fornecendo a seguinte desigualdade

$$2y \sum_{k>y} \frac{1}{k^2} = 2y \sum_{k=\lfloor y \rfloor + 1}^{\infty} \frac{1}{k^2} \leqslant \frac{2y}{\lfloor y \rfloor} \leqslant 4,$$

pois, para $y \geqslant 1$, temos $y < \lfloor y \rfloor + 1 \leqslant 2\lfloor y \rfloor$.

Para finalizar resta obter a desigualdade para o caso $0 \leqslant y < 1$. Neste caso, temos que todo inteiro positivo k satisfaz $k > y$, e portanto segue de uma aplicação diretta de (21) com $m = 2$ que

$$2y \sum_{k>y} \frac{1}{k^2} = 2y \sum_{k \geqslant 1} \frac{1}{k^2} \leqslant 2 \sum_{k \geqslant 1} \frac{1}{k^2} = 2\left(1 + \sum_{k \geqslant 2} \frac{1}{k^2}\right) \leqslant 2(1+1) = 4.$$

Em ambos os casos a constante 4 serve. A cota é grosseira, mas suficiente. ■

Observe que a constante exata não importa: qualquer constante finita serviria, porque ela será multiplicada por $\mathbb{E}[|X_1|]$, que é finito. O que importa é a ordem de grandeza, isto é, o fato de a constante não depender de y.

Lema 13 (Estimativa de momentos)**.** Se Y_k, para cada $k \in \mathbb{N}$, é a variável aleatória definida em (16), então

$$\sum_{k=1}^{\infty} \frac{\mathrm{Var}(Y_k)}{k^2} \leqslant 4\,\mathbb{E}[|X_1|] < \infty. \tag{22}$$

Prova. Como $\mathrm{Var}(Y_k) = \mathbb{E}[Y_k^2] - (\mathbb{E}[Y_k])^2 \leqslant \mathbb{E}[Y_k^2]$, basta cotar a soma dos segundos momentos e para estimar $\mathbb{E}[Y_k^2]$, vamos usar a representação do segundo momento mostrada abaixo. Para qualquer variável W com $\mathbb{E}[W^2] < \infty$,

$$\mathbb{E}[W^2] = \int_0^{\infty} 2y\,\mathbb{P}(|W| > y)\,dy. \tag{23}$$

Como $|Y_k| \leqslant k$, temos $\mathbb{P}(|Y_k| > y) = 0$ para $y \geqslant k$. Para $0 \leqslant y < k$, o evento $\{|Y_k| > y\}$ está contido em $\{|X_k| > y\}$, pois $|Y_k| \leqslant |X_k|$. Logo

$$\mathbb{P}(|Y_k| > y) \leqslant \mathbb{1}_{\{y<k\}}\,\mathbb{P}(|X_k| > y) = \mathbb{1}_{\{y<k\}}\,\mathbb{P}(|X_1| > y),$$

onde na última igualdade usamos que $(X_n)_{n\in\mathbb{N}}$ é uma sequência de v.a.'s identicamente distribuídas. Portanto

$$\mathbb{E}[Y_k^2] \leqslant \int_0^{k} 2y\,\mathbb{P}(|X_1| > y)\,dy. \tag{24}$$

Substituindo (24) na soma (22) e trocando a ordem de soma e integral, o que é permitido pelo teorema de Tonelli porque todos os termos são não negativos, obtemos a seguinte estimativa

$$\begin{aligned}
\sum_{k=1}^{\infty} \frac{\mathbb{E}[Y_k^2]}{k^2} &\leqslant \sum_{k=1}^{\infty} \frac{1}{k^2} \int_0^{\infty} \mathbb{1}_{\{y<k\}}\, 2y\,\mathbb{P}(|X_1| > y)\,dy \\
&= \int_0^{\infty} \Big\{ \sum_{k=1}^{\infty} \frac{\mathbb{1}_{\{y<k\}}}{k^2} \Big\}\, 2y\,\mathbb{P}(|X_1| > y)\,dy \\
&= \int_0^{\infty} \Big\{ \sum_{k>y} \frac{1}{k^2} \Big\}\, 2y\,\mathbb{P}(|X_1| > y)\,dy.
\end{aligned}$$

A soma entre chaves é cotada pelo Lema 12

$$2y \sum_{k>y} \frac{1}{k^2} \leqslant 4 \qquad \text{para todo } y \geqslant 0.$$

Logo concluímos das desigualdades acima que

$$\sum_{k=1}^{\infty} \frac{\mathrm{Var}(Y_k)}{k^2} \leqslant \sum_{k=1}^{\infty} \frac{\mathbb{E}[Y_k^2]}{k^2} \leqslant \int_0^{\infty} 4\,\mathbb{P}(|X_1| > y)\,dy = 4\,\mathbb{E}[|X_1|] < \infty,$$

onde na última igualdade usamos novamente a fórmula da esperança por integração da cauda e a hipótese de finitude da esperança de $|X_1|$ e com isto encerramos a prova do lema. ■

5.6. O Argumento de Subsequência

Suponha, como permite a Seção 5.2, que $X_n \geqslant 0$ para todo n. Então $Y_n \geqslant 0$ e $\mu = \mathbb{E}[X_1] \geqslant 0$.

Seja $\alpha > 1$ fixo, porém arbitrário. Defina

$$k(n) = \lfloor \alpha^n \rfloor, \qquad n \geqslant 1. \tag{25}$$

A sequência $k(n)$ é não decrescente e tende a infinito, podendo repetir alguns termos iniciais.

Queremos aplicar o primeiro lema de Borel-Cantelli às médias truncadas nesses índices. Para isso, precisamos mostrar que as probabilidades de desvio obtidas pela desigualdade de Chebyshev têm soma finita. Mais precisamente, provaremos que, para cada $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}\Big(|T_{k(n)} - \mathbb{E}[T_{k(n)}]| \geqslant \varepsilon k(n)\Big) \leqslant C_{\alpha,\varepsilon} \sum_{m=1}^{\infty} \frac{\mathrm{Var}(Y_m)}{m^2},$$

onde $C_{\alpha,\varepsilon}$ é uma constante finita que depende apenas de α e de ε. O Lema 13 garante que a série do lado direito converge. Essa estimativa fornecerá, portanto, a somabilidade necessária para aplicar Borel-Cantelli.

A escolha dos índices geométricos permite obter essa estimativa porque $k(n) \geqslant \alpha^n/2$ e, consequentemente, $k(n)^{-2} \leqslant 4\alpha^{-2n}$. Os termos α^{-2n} formam uma progressão geométrica de razão $\alpha^{-2} < 1$, cujas somas podemos calcular explicitamente. No argumento abaixo, usaremos essa comparação depois de trocar a ordem das somas, para obter a série $\sum_{m=1}^{\infty} \mathrm{Var}(Y_m)/m^2$ que aparece no lado direito.

Como os Y_m são v.a.'s duas-a-duas independentes, então

$$\mathrm{Var}(T_{k(n)}) = \sum_{m=1}^{k(n)} \mathrm{Var}(Y_m).$$

Pela Desigualdade de Chebyshev sabemos que para cada $\varepsilon > 0$ fixado temos

$$\mathbb{P}\Big(|T_{k(n)} - \mathbb{E}[T_{k(n)}]| \geqslant \varepsilon k(n)\Big) \leqslant \frac{\mathrm{Var}(T_{k(n)})}{\varepsilon^2 k(n)^2} = \frac{1}{\varepsilon^2 k(n)^2} \sum_{m=1}^{k(n)} \mathrm{Var}(Y_m).$$

Próximo passo é somar ambos os lados da desigualdade acima sobre $n \geqslant 1$. O lado direito passa a ser uma soma dupla, onde para cada $n \in \mathbb{N}$, somamos os termos com $1 \leqslant m \leqslant k(n)$.

Para inverter a ordem das somas, fixemos agora $m \geqslant 1$. A parcela $\mathrm{Var}(Y_m)/k(n)^2$ aparece precisamente para os índices n que satisfazem $k(n) \geqslant m$. Assim, podemos somar primeiro sobre esses índices e depois sobre todos os $m \geqslant 1$. Como todos os termos são não negativos, o teorema de Tonelli permite essa troca mesmo antes de sabermos se a soma é finita. Obtemos

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}\Big(|T_{k(n)} - \mathbb{E}[T_{k(n)}]| \geqslant \varepsilon k(n)\Big) &\leqslant \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \sum_{m=1}^{k(n)} \frac{\mathrm{Var}(Y_m)}{k(n)^2} \\
&= \frac{1}{\varepsilon^2} \sum_{m=1}^{\infty} \sum_{\substack{n \geqslant 1 \\ k(n) \geqslant m}} \frac{\mathrm{Var}(Y_m)}{k(n)^2} \\
&= \frac{1}{\varepsilon^2} \sum_{m=1}^{\infty} \mathrm{Var}(Y_m) \sum_{\substack{n \geqslant 1 \\ k(n) \geqslant m}} \frac{1}{k(n)^2}. \qquad (26)
\end{aligned}$$

Na última igualdade, colocamos $\mathrm{Var}(Y_m)$ em evidência na soma em n, pois essa variância não depende de n.

Próximo passo é dominar a soma interna pela cauda de uma série geométrica. Como $k(n) = \lfloor \alpha^n \rfloor \geqslant \alpha^n/2$ para $n \geqslant 1$, temos $k(n)^{-2} \leqslant 4\alpha^{-2n}$. Defina

$$n_0(m) = \min\{n \geqslant 1 : \alpha^n \geqslant m\}.$$

Como m é inteiro, $\lfloor \alpha^n \rfloor \geqslant m$ equivale a $\alpha^n \geqslant m$, logo os índices da soma são exatamente $n \geqslant n_0(m)$. Escrevendo $n_0 = n_0(m)$, temos $\alpha^{-2n_0} \leqslant m^{-2}$ e

$$\sum_{n:\, k(n)\geqslant m} \frac{1}{k(n)^2} \leqslant 4 \sum_{n\geqslant n_0} \alpha^{-2n} = 4\,\alpha^{-2n_0} \sum_{j=0}^{\infty} \alpha^{-2j} \leqslant \frac{4}{1-\alpha^{-2}} \cdot \frac{1}{m^2}. \tag{27}$$

Substituindo (27) em (26) e usando o Lema 13,

$$\sum_{n=1}^{\infty} \mathbb{P}\Big(|T_{k(n)} - \mathbb{E}[T_{k(n)}]| \geqslant \varepsilon k(n)\Big) \leqslant \frac{4}{\varepsilon^2(1-\alpha^{-2})} \sum_{m=1}^{\infty} \frac{\mathrm{Var}(Y_m)}{m^2} \leqslant \frac{16\,\mathbb{E}[|X_1|]}{\varepsilon^2(1-\alpha^{-2})} < \infty. \tag{28}$$

Como $k(n) \geqslant 1$, a estimativa (28) mostra que, para todo $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{T_{k(n)} - \mathbb{E}[T_{k(n)}]}{k(n)}\right| \geqslant \varepsilon\right) < \infty.$$

Está, portanto, verificada a hipótese do critério de Borel-Cantelli para a convergência quase certa, recordado na Seção 2.2. Aplicando esse critério à sequência dos desvios normalizados, obtemos

$$\frac{T_{k(n)} - \mathbb{E}[T_{k(n)}]}{k(n)} \longrightarrow 0 \qquad \text{quase certamente.} \tag{29}$$

Além disso, em (18) já mostramos que a sequência numérica $\mathbb{E}[T_m]/m$ converge para μ. Como $k(n) \to \infty$, essa convergência também vale ao longo dos índices $k(n)$:

$$\frac{\mathbb{E}[T_{k(n)}]}{k(n)} \longrightarrow \mu.$$

Para concluir, decompomos a média truncada em duas parcelas:

$$\frac{T_{k(n)}}{k(n)} = \frac{T_{k(n)} - \mathbb{E}[T_{k(n)}]}{k(n)} + \frac{\mathbb{E}[T_{k(n)}]}{k(n)}.$$

A primeira tende a zero quase certamente, por (29), e a segunda tende a μ. Portanto,

$$\frac{T_{k(n)}}{k(n)} \longrightarrow \mu \qquad \text{quase certamente.} \tag{30}$$

5.7. Os Valores Intermediários

O limite quase certo garantido por (30) fornece a convergência apenas ao longo dos índices $k(n)$. Para completar a prova falta controlar os índices m que ficam entre dois valores consecutivos da sequência $k(n)$. A hipótese $X_n \geqslant 0$ será a chave para fazer o controle da sequência ao longo desses índices, por meio de um argumento de monotonicidade.

Já que $Y_k \geqslant 0$, então podemos afirmar que a sequência $T_m = Y_1 + \cdots + Y_m$ é não decrescente em m. Seja m um inteiro positivo satisfazendo

$$k(n) \leqslant m < k(n+1).$$

Pela monotonicidade, $T_{k(n)} \leqslant T_m \leqslant T_{k(n+1)}$. Usando também a não negatividade das variáveis aleatórias T_n's e as desigualdades acima, obtemos

$$\frac{T_{k(n)}}{k(n+1)} \leqslant \frac{T_m}{k(n+1)} \leqslant \frac{T_m}{m} \leqslant \frac{T_{k(n+1)}}{m} \leqslant \frac{T_{k(n+1)}}{k(n)},$$

isto é,

$$\frac{T_{k(n)}}{k(n+1)} \leqslant \frac{T_m}{m} \leqslant \frac{T_{k(n+1)}}{k(n)}. \tag{31}$$

Para usar (31), precisamos calcular os limites de seus dois extremos. Em cada um deles, multiplicaremos e dividiremos pelo índice da soma que aparece no numerador. Assim, faremos aparecer as médias cuja convergência já conhecemos, acompanhadas de razões entre índices consecutivos.

Comecemos por essas razões. Pela definição da parte inteira, $0 \leqslant \alpha^n - k(n) < 1$. Dividindo por α^n e fazendo $n \to \infty$, obtemos $k(n)/\alpha^n \to 1$. Consequentemente,

$$\frac{k(n+1)}{k(n)} = \frac{k(n+1)}{\alpha^{n+1}} \cdot \alpha \cdot \frac{\alpha^n}{k(n)} \longrightarrow \alpha,$$

e a razão inversa tende a α^{-1}.

No que segue trabalhamos no evento de probabilidade 1 em que vale (30). Nesse evento,

$$\frac{T_{k(n)}}{k(n+1)} = \frac{T_{k(n)}}{k(n)} \cdot \frac{k(n)}{k(n+1)} \longrightarrow \frac{\mu}{\alpha},$$

$$\frac{T_{k(n+1)}}{k(n)} = \frac{T_{k(n+1)}}{k(n+1)} \cdot \frac{k(n+1)}{k(n)} \longrightarrow \alpha\mu.$$

Na desigualdade (31), as duas cotas dependem apenas de n. Assim, elas controlam de uma só vez todas as médias T_m/m com $k(n) \leqslant m < k(n+1)$. Vamos usar esse controle em cada intervalo para obter uma estimativa válida para todos os índices suficientemente grandes. A Figura 3 ilustra como os intervalos se sucedem.

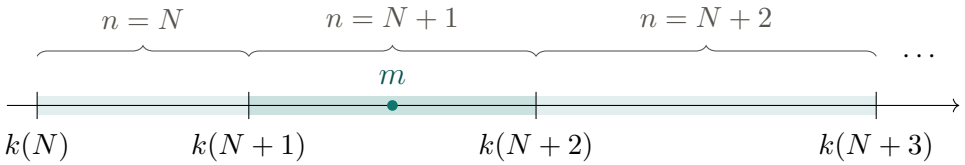


Figura 3: Os intervalos $[k(n), k(n+1))$, com $n \geqslant N$, cobrem todos os inteiros $m \geqslant k(N)$. Em cada intervalo, a desigualdade (31) controla todas as médias T_m/m pelas mesmas duas cotas. O desenho é esquemático.

Fixemos $\delta > 0$. Pelos dois limites que acabamos de obter, existe N tal que, para todo $n \geqslant N$,

$$\frac{\mu}{\alpha} - \delta \leqslant \frac{T_{k(n)}}{k(n+1)} \qquad \text{e} \qquad \frac{T_{k(n+1)}}{k(n)} \leqslant \alpha\mu + \delta.$$

Combinando essas cotas com (31), obtemos

$$\frac{\mu}{\alpha} - \delta \leqslant \frac{T_m}{m} \leqslant \alpha\mu + \delta \qquad \text{se } k(n) \leqslant m < k(n+1), \quad n \geqslant N.$$

Como $k(n) \to \infty$, esses intervalos cobrem todos os inteiros $m \geqslant k(N)$. Portanto, a estimativa acima vale para todo $m \geqslant k(N)$.

Uma vez que $\delta > 0$ é arbitrário, o limite inferior das médias é pelo menos μ/α, e o limite superior é no máximo $\alpha\mu$. Concluímos, no evento de probabilidade 1 em que estamos trabalhando, que

$$\frac{\mu}{\alpha} \leqslant \liminf_{m\to\infty} \frac{T_m}{m} \leqslant \limsup_{m\to\infty} \frac{T_m}{m} \leqslant \alpha\mu. \tag{32}$$

Para concluir que o limite é μ, precisamos aproximar as duas cotas de (32) desse mesmo valor. Isso sugere escolher α cada vez mais próximo de 1, o que podemos fazer porque o argumento vale para qualquer $\alpha > 1$.

Precisamos, porém, que as desigualdades usadas nessa passagem valham no mesmo evento de probabilidade 1. Para cada α fixado, obtivemos um evento com essa propriedade, que pode depender de α. Basta então escolher uma sequência enumerável de parâmetros que tenda a 1:

$$\alpha_j = 1 + \frac{1}{j}, \qquad j \geqslant 1.$$

A interseção dos eventos de probabilidade 1 correspondentes a esses parâmetros também tem probabilidade 1. Nessa interseção, (32) vale simultaneamente com $\alpha = \alpha_j$ para todo $j \geqslant 1$. Podemos, portanto, fazer $j \to \infty$ nas duas cotas. Como $\mu/\alpha_j \to \mu$ e $\alpha_j\mu \to \mu$, obtemos

$$\mu \leqslant \liminf_{m\to\infty} \frac{T_m}{m} \leqslant \limsup_{m\to\infty} \frac{T_m}{m} \leqslant \mu.$$

Os limites inferior e superior são, assim, iguais a μ, o que prova que $T_m/m \to \mu$ quase certamente. Pelo Lema 11, as médias originais S_m/m também convergem quase certamente para μ. Isso conclui a prova do Teorema 10 para variáveis não negativas; a redução da Seção 5.2 fornece o caso geral.

5.8. Observações sobre a Prova

A prova está completa, mas vale destacar três pontos que costumam passar despercebidos na primeira leitura.

O primeiro é o papel de cada hipótese. A hipótese de distribuições iguais é usada no lema de truncamento, para escrever $\mathbb{P}(|X_k| > k) = \mathbb{P}(|X_1| > k)$, na estimativa dos segundos momentos, para substituir $\mathbb{P}(|X_k| > y)$ pela cauda de X_1, e na identificação de $\mathbb{E}[Y_k]$ com a esperança do truncamento correspondente de X_1. Junto com a fórmula de integração da cauda e a cota do Lema 12, isso produz a cota $4\mathbb{E}[|X_1|]$ no Lema 13. A hipótese de independência dois a dois é usada apenas para que a variância da soma seja a soma das variâncias; não é preciso mais do que isso. A hipótese $\mathbb{E}[|X_1|] < \infty$ é usada no truncamento, para garantir a somabilidade de (17), e na estimativa de momentos, para garantir a finitude do lado direito do Lema 13. Ela também garante que as médias truncadas convergem para um centro finito em (18).

O segundo é o papel do parâmetro α. Ele é arbitrário em $(1, \infty)$ e o argumento funciona para qualquer valor fixado. Como o conjunto nulo pode depender de α, na conclusão intersectamos os conjuntos de probabilidade 1 apenas para a sequência enumerável $\alpha_j = 1 + 1/j$.

Fazendo $j \to \infty$, identificamos o limite. Se tivéssemos fixado $\alpha = 2$, por exemplo, obteríamos apenas

$$\frac{\mu}{2} \leqslant \liminf \frac{T_m}{m} \leqslant \limsup \frac{T_m}{m} \leqslant 2\mu,$$

que, quando $\mu > 0$, ainda não identifica o limite.

O terceiro é o papel da não negatividade. Ela garante tanto a monotonicidade de T_m quanto os sinais necessários às comparações dos denominadores em (31). Sem $Y_k \geqslant 0$, essas desigualdades podem falhar e não teríamos o controle dos valores intermediários usado neste argumento.

6. A Recíproca: A Integrabilidade é Necessária

A hipótese $\mathbb{E}[|X_1|] < \infty$ não é uma conveniência técnica. Nesta seção, mostramos que, sem ela, a conclusão da lei forte falha para sequências independentes e identicamente distribuídas. O argumento é o da Seção 2.3 de [2].

Teorema 14 (Recíproca da lei forte)**.** Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas com $\mathbb{E}[|X_1|] = \infty$. Então

$$\mathbb{P}\Big(|X_n| \geqslant n \text{ infinitas vezes}\Big) = 1.$$

Consequentemente, se $S_n = X_1 + \cdots + X_n$,

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{S_n}{n} \text{ existe e é finito}\right) = 0.$$

Prova. Aplicando a fórmula (5) e a comparação (6) a $W = |X_1|$, obtemos

$$\mathbb{E}[|X_1|] = \int_0^\infty \mathbb{P}(|X_1| > t)\, dt \leqslant \sum_{n=0}^{\infty} \mathbb{P}(|X_1| > n). \tag{33}$$

Como a esperança é infinita, a série do lado direito diverge. Sendo as variáveis identicamente distribuídas,

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n| \geqslant n) \geqslant \sum_{n=1}^{\infty} \mathbb{P}(|X_n| > n) = \sum_{n=1}^{\infty} \mathbb{P}(|X_1| > n) = \infty.$$

Os eventos $\{|X_n| \geqslant n\}$ são independentes, logo o segundo lema de Borel-Cantelli fornece

$$\mathbb{P}\Big(|X_n| \geqslant n \text{ infinitas vezes}\Big) = 1. \tag{34}$$

Para concluir, mostraremos que a convergência de S_n/n a um limite finito obrigaria $X_n/n \to 0$, em contradição com (34). Suponha que exista ω tal que $S_n(\omega)/n$ convirja para um número real ℓ. Da identidade

$$\frac{S_n}{n} - \frac{S_{n+1}}{n+1} = \frac{S_n}{n(n+1)} - \frac{X_{n+1}}{n+1},$$

obtemos

$$\frac{X_{n+1}}{n+1} = \frac{S_n}{n(n+1)} - \left(\frac{S_n}{n} - \frac{S_{n+1}}{n+1}\right).$$

O primeiro termo do lado direito vale $\frac{1}{n+1} \cdot \frac{S_n}{n}$, que tende a zero porque $S_n/n \to \ell$ e ℓ é finito. O segundo tende a $\ell - \ell = 0$. Logo

$$\frac{|X_{n+1}(\omega)|}{n+1} \longrightarrow 0. \tag{35}$$

Por outro lado, (35) é equivalente, após reindexação, a $|X_n(\omega)|/n \to 0$. Mas o evento $\{|X_n| \geqslant n \text{ infinitas vezes}\}$ significa exatamente que $|X_n(\omega)|/n \geqslant 1$ para infinitos índices n, o que contradiz a convergência para zero. Portanto o conjunto em que S_n/n converge para um limite finito é disjunto de $\{|X_n| \geqslant n \text{ infinitas vezes}\}$. Por (34), esse último conjunto tem probabilidade 1, e a conclusão segue. ■

A ideia do argumento é identificar uma condição necessária para a convergência das médias: se S_n/n converge para um limite finito, então $X_n/n \to 0$. Sem integrabilidade, porém, quase certamente $|X_n| \geqslant n$ para infinitos índices n. Assim, $|X_n|/n$ é pelo menos 1 infinitas vezes, o que impede essa razão de tender a zero. Portanto, as médias também não podem convergir para um limite finito.

Exemplo 15 (A distribuição de Cauchy)**.** Sejam $X_1, X_2, \ldots$ independentes com a distribuição de Cauchy padrão, cuja densidade é

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

Então

$$\mathbb{E}[|X_1|] = \int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)}\, dx = \frac{2}{\pi}\int_0^{\infty} \frac{x}{1+x^2}\, dx = \infty,$$

de modo que o Teorema 14 se aplica: S_n/n não converge quase certamente.

Por outro lado, a média de variáveis de Cauchy independentes também é de Cauchy. Este fato será provado mais a frente quando introduzirmos a função característica. De qualquer forma usando este resultado podemos afirmar que S_n/n tem exatamente a mesma distribuição para todo n e, em particular, $S_n/n \Rightarrow X_1$. O exemplo mostra que a convergência em distribuição pode valer trivialmente enquanto a convergência quase certa falha, e que a hipótese de integrabilidade é indispensável para a lei forte.

No exemplo da Cauchy, a média também não converge em probabilidade para nenhuma constante $c \in \mathbb{R}$ já que para todo $\varepsilon > 0$,

$$\mathbb{P}(|S_n/n - c| \geqslant \varepsilon) = \mathbb{P}(|X_1 - c| \geqslant \varepsilon) > 0,$$

independentemente de n. Isso não contradiz a lei fraca, porque a lei fraca na forma demonstrada exige variância finita, e a Cauchy não tem nem esperança. No caso independente e identicamente distribuído, a integrabilidade é suficiente, mas não necessária, para a lei fraca. Já o Teorema 14 mostra que ela é necessária para a convergência quase certa das médias a um limite finito.

7. Aplicações

7.1. Números Normais

O Teorema 9 mostra que, para quase todo número real, cada dígito decimal tem frequência assintótica 1/10. Podemos perguntar também pelas frequências de pares, trincas e blocos mais longos. Por exemplo, será que os blocos 00 e 37 têm, cada um, frequência 1/100?

O equilíbrio dos dígitos isolados não basta para responder a essa pergunta. Considere o número

$$x = 0{,}012345678\,90123456789\,0123456789\ldots.$$

Cada dígito aparece uma vez a cada dez posições, mas o bloco 00 nunca aparece. Esse número é simplesmente normal na base 10, mas não é normal. A normalidade, definida na Definição 4, exige que *toda* palavra finita u de comprimento ℓ tenha frequência $10^{-\ell}$. Vamos mostrar que quase todo número real satisfaz essa propriedade.

7.1.1. Contar as ocorrências de uma palavra

Retomemos o espaço $[0,1)$ com a probabilidade de Lebesgue λ e os dígitos $d_j(x)$ da expansão decimal de x, com a convenção da Seção 4.3. Pelo Lema 8, essas variáveis são independentes e uniformes em $\{0,\ldots,9\}$.

Fixemos uma palavra $u = u_1 \ldots u_\ell$ de comprimento $\ell \geqslant 1$. Para registrar se ela aparece a partir da posição j, definimos

$$Z_j^{(u)}(x) = \mathbb{1}_{\{d_j = u_1, \ldots, d_{j+\ell-1} = u_\ell\}}(x).$$

Assim, $Z_j^{(u)}(x) = 1$ quando o bloco que começa em j é u, e vale 0 caso contrário. A frequência de ocorrências entre as n primeiras posições de início é, portanto,

$$F_n^{(u)}(x) = \frac{1}{n} \sum_{j=1}^{n} Z_j^{(u)}(x).$$

Contamos todas as posições de início, inclusive as que produzem ocorrências sobrepostas. Por exemplo, o bloco 00 aparece duas vezes em 000, com inícios na primeira e na segunda posições.

Pela independência dos dígitos, a probabilidade de encontrar u em qualquer posição fixada é

$$\lambda\big(Z_j^{(u)} = 1\big) = \prod_{i=1}^{\ell} \lambda\big(d_{j+i-1} = u_i\big) = \left(\frac{1}{10}\right)^{\ell} = 10^{-\ell}.$$

Queremos mostrar que a frequência observada $F_n^{(u)}$ converge quase certamente para essa probabilidade.

7.1.2. Separar os blocos para obter independência

Uma tentativa seria aplicar a lei forte diretamente à sequência $(Z_j^{(u)})$. A dificuldade é que blocos com posições iniciais próximas usam alguns dos mesmos dígitos. Por exemplo, se $u = 372$, a condição $Z_1^{(u)} = 1$ exige $d_2 = 7$, enquanto $Z_2^{(u)} = 1$ exige $d_2 = 3$. Esses dois eventos têm probabilidade positiva, mas não podem ocorrer simultaneamente. Logo, as indicadoras não são independentes.

A sobreposição sugere como proceder: separar as posições iniciais de modo que os blocos examinados usem posições disjuntas da expansão decimal. Para palavras de comprimento 3, os blocos que começam em $1, 4, 7, \ldots$ examinam, respectivamente,

$$(d_1, d_2, d_3), \qquad (d_4, d_5, d_6), \qquad (d_7, d_8, d_9), \qquad \ldots.$$

Como os dígitos são independentes e nenhum deles é usado em dois desses blocos, as indicadoras correspondentes são independentes. Podemos fazer o mesmo com os inícios $2, 5, 8, \ldots$

e com os inícios $3, 6, 9, \ldots$. Esses três grupos, reunidos, incluem todas as posições de início, como mostra a Figura 4.

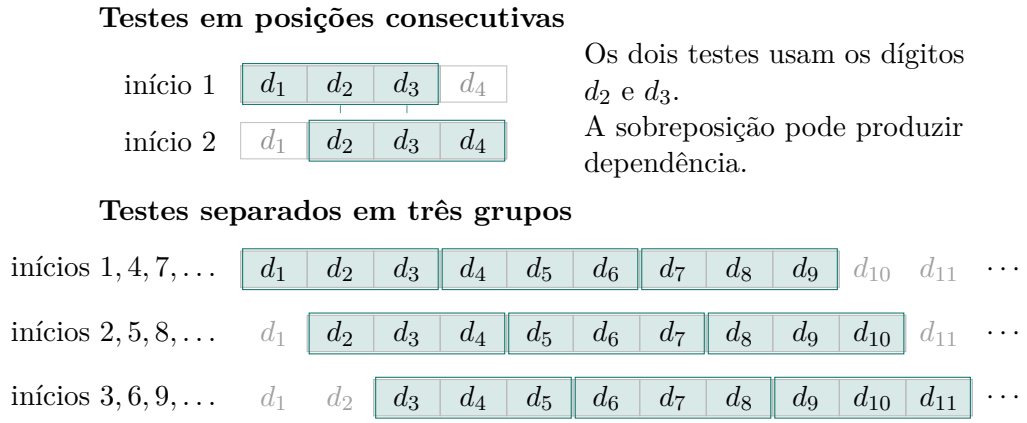


Figura 4: Cada retângulo destacado representa um bloco de três dígitos examinado na procura de uma palavra fixada. Acima, os testes compartilham dígitos. Abaixo, os blocos de uma mesma linha usam posições disjuntas, o que garante a independência das indicadoras nessa linha. As três linhas incluem, juntas, todas as posições de início.

Para uma palavra de comprimento ℓ, fazemos a mesma separação em ℓ grupos. Fixado $r \in \{1, \ldots, \ell\}$, consideramos as posições

$$r, \quad r+\ell, \quad r+2\ell, \quad \ldots.$$

As indicadoras $Z_r^{(u)}, Z_{r+\ell}^{(u)}, Z_{r+2\ell}^{(u)}, \ldots$ são independentes, têm a mesma distribuição de Bernoulli com parâmetro $10^{-\ell}$ e são limitadas por 1. Pelo Corolário 7,

$$\frac{1}{q}\sum_{k=0}^{q-1} Z_{r+k\ell}^{(u)} \longrightarrow 10^{-\ell} \qquad \text{quase certamente.}$$

Aplicamos esse argumento a cada um dos ℓ grupos. A interseção dos ℓ eventos de probabilidade 1 também tem probabilidade 1; nela, todas essas frequências convergem simultaneamente.

7.1.3. Reunir as frequências dos grupos

Vamos primeiro contar as ocorrências com início nas primeiras $q\ell$ posições. Entre elas, há exatamente q posições de cada grupo. Portanto, a frequência total é a média aritmética das ℓ frequências que acabamos de estudar:

$$F_{q\ell}^{(u)} = \frac{1}{\ell}\sum_{r=1}^{\ell}\left(\frac{1}{q}\sum_{k=0}^{q-1} Z_{r+k\ell}^{(u)}\right) \longrightarrow \frac{1}{\ell}\sum_{r=1}^{\ell} 10^{-\ell} = 10^{-\ell}.$$

Essa identidade explica como a separação em grupos preserva a contagem de todas as ocorrências, inclusive as sobrepostas.

Para tratar todos os $n \geqslant \ell$, escrevemos $n = q\ell + s$, com $0 \leqslant s < \ell$. Além das primeiras $q\ell$ posições, restam apenas s posições a examinar. Assim,

$$F_n^{(u)} = \frac{q\ell}{n}\, F_{q\ell}^{(u)} + \frac{1}{n} \sum_{j=q\ell+1}^{n} Z_j^{(u)}.$$

A última parcela está entre 0 e $(\ell - 1)/n$, pois cada indicadora vale no máximo 1. Como $n - q\ell = s$ permanece limitado, temos $q \to \infty$ e $q\ell/n \to 1$ quando $n \to \infty$. Obtemos, portanto,

$$F_n^{(u)} \longrightarrow 10^{-\ell} \qquad \text{quase certamente.} \tag{36}$$

Se exigirmos que cada ocorrência esteja inteiramente contida nos primeiros n dígitos, como no Exercício 3, basta excluir os inícios posteriores à posição $n - \ell + 1$. São no máximo $\ell - 1$ ocorrências a menos, de modo que a diferença entre as duas frequências é no máximo $(\ell - 1)/n$ e tende a zero. As duas maneiras de contar fornecem o mesmo limite.

7.1.4. Todas as palavras e todas as bases

Até aqui, fixamos uma palavra u. Para cada comprimento ℓ, existem apenas 10^ℓ palavras, de modo que a família de todas as palavras finitas é enumerável. Para cada uma delas, (36) falha apenas em um conjunto de medida zero. A união desses conjuntos excepcionais também tem medida zero. Fora dela, todas as palavras têm, simultaneamente, a frequência exigida pela normalidade.

Concluímos que quase todo $x \in [0, 1)$ é normal na base 10. Como na Seção 4.3, o resultado se estende a $\mathbb{R}$ por translações inteiras: os dígitos considerados são os da parte fracionária, e as translações do conjunto excepcional ainda formam uma união enumerável de conjuntos de medida zero.

Por fim, a escolha da base 10 não foi essencial. Em uma base inteira $b \geqslant 2$, dividimos o intervalo em b partes iguais a cada etapa. Os dígitos são então independentes e uniformes em $\{0, \ldots, b - 1\}$, e a probabilidade de uma palavra de comprimento ℓ é $b^{-\ell}$. O mesmo argumento mostra que quase todo número real é normal nessa base.

Um número é chamado *completamente normal* quando é normal em toda base inteira $b \geqslant 2$. Para cada base, acabamos de obter um conjunto excepcional de medida zero. Como há apenas uma quantidade enumerável de bases inteiras, a união desses conjuntos também tem medida zero. Fora dela, a normalidade vale em todas as bases simultaneamente. Portanto, quase todo número real é completamente normal.

7.2. Integração de Monte Carlo

O método de Monte Carlo para calcular uma integral se apoia na lei forte. Seja $f : [0, 1] \to \mathbb{R}$ uma função Borel-mensurável com $\int_0^1 |f(x)|\, dx < \infty$, e sejam $U_1, U_2, \ldots$ independentes e uniformes em $[0, 1]$. Então as variáveis $f(U_i)$ são independentes e identicamente distribuídas, com

$$\mathbb{E}\big[|f(U_1)|\big] = \int_0^1 |f(x)|\, dx < \infty.$$

Pelo Teorema 10,

$$\frac{1}{n} \sum_{i=1}^{n} f(U_i) \longrightarrow \int_0^1 f(x)\, dx \qquad \text{quase certamente.} \tag{37}$$

A integral é o valor esperado de $f(U_1)$, e a média empírica é o estimador obtido por amostragem. A lei forte garante que o estimador converge ao longo de quase toda realização da sequência, e não apenas em probabilidade. Se, adicionalmente, $\int_0^1 f(x)^2\,dx < \infty$, a desigualdade de Chebyshev fornece uma cota para a probabilidade de erro em cada n finito. Sob as hipóteses usuais do teorema central do limite, aqui satisfeitas quando $f(U_1)$ tem variância finita e positiva, obtemos também uma aproximação normal assintótica para o erro; essa aproximação não é uma cota para n finito.

Tomando $f = \mathbb{1}_{[k/10,(k+1)/10)}$, a indicadora aplicada a cada U_i registra se o primeiro dígito decimal da i-ésima amostra é k. A média empírica é, portanto, a frequência desse dígito entre as amostras, e a integral vale $1/10$. Esses indicadores têm a mesma distribuição conjunta que os indicadores dos dígitos sucessivos de um único número uniforme em $[0, 1]$: ambas as sequências são formadas por variáveis independentes com distribuição de Bernoulli com parâmetro $1/10$.

7.3. A Função de Distribuição Empírica

Sejam $X_1, X_2, \ldots$ independentes e identicamente distribuídas com função de distribuição F. Para cada $x \in \mathbb{R}$, a variável $\mathbb{1}_{\{X_i \leqslant x\}}$ é de Bernoulli com parâmetro $F(x)$, e as variáveis são independentes. Pelo Teorema 10,

$$F_n(x) \equiv \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leqslant x\}} \longrightarrow F(x) \qquad \text{quase certamente,} \tag{38}$$

para cada x fixado. A função F_n é a *função de distribuição empírica* e $F_n(x)$ é a frequência observada de valores menores ou iguais a x.

A afirmação acima é pontual: para cada x fixado, vale quase certamente, mas isso, por si só, não fornece um único evento de probabilidade 1 que sirva para todos os x. Primeiro suponha que F seja contínua. Como $\mathbb{Q}$ é enumerável, há um evento de probabilidade 1 no qual $F_n(q) \to F(q)$ para todo racional q. Fixado $\varepsilon > 0$, escolha uma malha finita racional $q_1 < \cdots < q_m$ que cubra a parte central da distribuição, com as massas entre pontos consecutivos menores que ε e com as duas caudas menores que ε. Para construir essa malha, escolhemos primeiro extremos racionais com caudas pequenas, usando os limites de F em $-\infty$ e $+\infty$. A continuidade uniforme de F no intervalo compacto entre esses extremos permite então subdividi-lo em uma malha racional com as variações desejadas. Para $q_j \leqslant x \leqslant q_{j+1}$, a monotonicidade dá, em ambas as direções,

$$|F_n(x) - F(x)| \leqslant \max_{1 \leqslant j \leqslant m} |F_n(q_j) - F(q_j)| + \varepsilon,$$

e os mesmos controles valem nas caudas. Como a malha é finita, o máximo dos erros nos seus pontos tende a zero. Aplicando o argumento a uma sequência enumerável de ε que tende a zero, obtemos

$$\sup_{x \in \mathbb{R}} \big|F_n(x) - F(x)\big| \longrightarrow 0 \qquad \text{quase certamente.}$$

Esse é o teorema de Glivenko-Cantelli no caso contínuo. Se F tem átomos, a mesma intercalação por racionais garante a convergência nos pontos de continuidade, mas não garante a convergência em todos os saltos, que podem ocorrer em pontos irracionais. Nos saltos racionais, a convergência já está incluída no evento escolhido. Os átomos formam um conjunto enumerável, e a convergência em cada posição atômica também vale quase certamente pela

lei forte, assim como a convergência dos limites à esquerda $F_n(a-)$ para $F(a-)$ em cada átomo a. Para obter a extensão uniforme, separam-se os finitos átomos de massa maior que uma tolerância fixada e refinam-se os intervalos entre eles até que as variações restantes de F sejam pequenas; a última parte do Exercício 8 detalha essa ideia.

7.4. Distribuições de Cauda Pesada: A Família de Pareto

Um dos aspectos mais marcantes do Teorema 10 é que ele exige apenas a finitude do primeiro momento ($\mathbb{E}[|X_1|] < \infty$), dispensando completamente a existência de variância ou de momentos de ordem superior. Para compreender o alcance dessa generalidade, é instrutivo examinar situações em que a média teórica existe, mas as flutuações ao redor dela são dominadas por caudas pesadas.

Um modelo clássico com esse comportamento é a *distribuição de Pareto*. Sejam $\alpha > 0$ e $x_0 > 0$. Uma variável aleatória contínua X tem distribuição de Pareto com parâmetro de forma α e escala x_0 se sua densidade é dada por

$$f(x) = \begin{cases} \dfrac{\alpha x_0^{\alpha}}{x^{\alpha+1}}, & \text{para } x \geqslant x_0, \\ 0, & \text{para } x < x_0. \end{cases} \tag{39}$$

Para fixar ideias e simplificar a álgebra, fixemos a escala $x_0 = 1$. A probabilidade de cauda decai como uma lei de potência:

$$\mathbb{P}(X > t) = \int_t^{\infty} \frac{\alpha}{x^{\alpha+1}}\, dx = \frac{1}{t^{\alpha}}, \qquad t \geqslant 1.$$

Calculando os momentos de ordem $p > 0$, temos

$$\mathbb{E}[X^p] = \int_1^{\infty} x^p \frac{\alpha}{x^{\alpha+1}}\, dx = \alpha \int_1^{\infty} x^{p-\alpha-1}\, dx,$$

integral que converge se, e somente se, $p < \alpha$.

Consideremos agora o regime $1 < \alpha \leqslant 2$, tomando por exemplo $\alpha = 3/2$. Nesse caso:

(i) O primeiro momento é finito:

$$\mathbb{E}[X_1] = \frac{\alpha}{\alpha - 1} = \frac{3/2}{1/2} = 3 < \infty;$$

(ii) O segundo momento diverge:

$$\mathbb{E}[X_1^2] = \frac{3}{2} \int_1^{\infty} x^{2-5/2}\, dx = \frac{3}{2} \int_1^{\infty} \frac{1}{\sqrt{x}}\, dx = \infty,$$

e, consequentemente, $\operatorname{Var}(X_1) = \infty$ e $\mathbb{E}[X_1^4] = \infty$.

Essa classe de distribuições desempenha um papel fundamental em finanças, ciência de redes e teoria do risco, modelando fenômenos sujeitos a eventos extremos (como grandes sinistros em companhias de seguros, distribuições de renda ou volumes de tráfego de dados na internet).

Suponha agora que $X_1, X_2, \ldots$ sejam variáveis aleatórias independentes com essa distribuição e denote por $S_n = X_1 + \cdots + X_n$ a soma acumulada. O que podemos afirmar sobre o comportamento assintótico da média empírica S_n/n?

- **A abordagem pelo quarto momento (Teorema 5) não se aplica:** como $\mathbb{E}[X_1^4] = \infty$, a estimativa via $(S_n - n\mu)^4$ sequer está bem definida;
- **A abordagem de Chebyshev e a Lei Fraca elementar falham:** como $\mathrm{Var}(X_1) = \infty$, a desigualdade de Chebyshev não fornece nenhum controle útil para $\mathbb{P}(|S_n/n - 3| \geqslant \varepsilon)$;
- **A Lei Forte de Etemadi (Teorema 10) aplica-se plenamente:** como as variáveis são independentes e $\mathbb{E}[|X_1|] = 3 < \infty$, o teorema garante que

$$\frac{S_n}{n} \longrightarrow 3 \qquad \text{quase certamente.} \tag{40}$$

Esse exemplo ilustra de forma nítida o poder da demonstração de Etemadi: a convergência quase certa prevalece mesmo quando as observações individuais produzem flutuações e saltos de variância infinita.

8. Exercícios Propostos

1. Seja X uma variável aleatória. Dê uma demonstração alternativa da comparação (6), aplicada a $W = |X|$:

$$\sum_{k=1}^{\infty} \mathbb{P}(|X| > k) \leqslant \mathbb{E}[|X|] \leqslant \sum_{k=0}^{\infty} \mathbb{P}(|X| > k),$$

e conclua que a série $\sum_k \mathbb{P}(|X| > k)$ converge se, e somente se, $\mathbb{E}[|X|] < \infty$. *Sugestão:* compare pontualmente $|X|$ com as somas $\sum_{k=1}^{\infty} \mathbb{1}_{\{|X|>k\}}$ e $\sum_{k=0}^{\infty} \mathbb{1}_{\{|X|>k\}}$ e tome esperanças.

2. Sejam $X_1, X_2, \ldots$ independentes com $\mathbb{P}(X_i = 1) = p$ e $\mathbb{P}(X_i = 0) = 1 - p$, e $S_n = X_1 + \cdots + X_n$.

(a) Calcule $\mathbb{E}[(S_n - np)^4]$ em função de n e de p e mostre que ela é $O(n^2)$. Mais precisamente, para $0 < p < 1$, mostre que é da ordem de n^2; trate também os casos degenerados $p = 0$ e $p = 1$.

(b) Aplique a desigualdade de Markov a $|S_n - np|^4$ para mostrar que $\mathbb{P}(|S_n/n - p| \geqslant \varepsilon) \leqslant C/(n^2\varepsilon^4)$ e conclua a convergência quase certa.

3. No espaço $\Omega = \{0, 1, \ldots, 9\}^{\mathbb{N}}$ com a medida produto uniforme, seja $u = u_1 \ldots u_\ell$ uma palavra de comprimento $\ell \geqslant 1$. Defina

$$N_n(u) = \#\{1 \leqslant j \leqslant n - \ell + 1 : X_j = u_1, \ldots, X_{j+\ell-1} = u_\ell\}.$$

Mostre que $N_n(u)/n \to 10^{-\ell}$ quase certamente. *Sugestão:* considere primeiro as posições $j = 1, 1 + \ell, 1 + 2\ell, \ldots$ e depois intercale as ℓ progressões aritméticas de razão ℓ.

4. Seja $f : [0, 1] \to \mathbb{R}$ Borel-mensurável, com $\int_0^1 |f(x)|\, dx < \infty$, e sejam $U_1, U_2, \ldots$ independentes e uniformes em $[0, 1]$.

(a) Mostre que $\frac{1}{n}\sum_{i=1}^{n} f(U_i) \to \int_0^1 f(x)\, dx$ quase certamente.

(b) Suponha que $|f(x)| \leqslant M < \infty$ para todo $x \in [0, 1]$. Para $\varepsilon > 0$, use a desigualdade de Chebyshev para estimar $\mathbb{P}(|\frac{1}{n}\sum_{i=1}^{n} f(U_i) - \int_0^1 f| \geqslant \varepsilon)$ em função de n, ε e M.

5. Sejam $X_1, X_2, \ldots$ independentes, identicamente distribuídas e não negativas, com $\mathbb{E}[X_1] = \infty$, e defina $S_n = \sum_{i=1}^n X_i$. Mostre que

$$\frac{S_n}{n} \longrightarrow \infty \qquad \text{quase certamente.}$$

Sugestão: para $M > 0$, aplique a lei forte às variáveis truncadas $X_i^M = \min\{X_i, M\}$ e use a convergência monótona para mostrar que $\mathbb{E}[X_1^M] \to \infty$.

6. Sejam $X_1, X_2, \ldots$ independentes e identicamente distribuídas com densidade $f(x) = x^{-2}$ para $x \geqslant 1$ e $f(x) = 0$ para $x < 1$. Defina $S_n = \sum_{i=1}^n X_i$.

(a) Mostre que $\mathbb{E}[|X_1|] = \infty$ e que $\mathbb{P}(X_1 > t) = 1/t$ para $t \geqslant 1$.

(b) Conclua, usando o exercício anterior, que S_n/n não converge quase certamente para um limite finito.

(c) Mostre que $\mathbb{P}(|X_n| \geqslant n \text{ infinitas vezes}) = 1$, usando diretamente o segundo lema de Borel-Cantelli.

7. Mostre que existe uma variável aleatória X com $\mathbb{E}[|X|] < \infty$ e $\mathbb{E}[X^4] = \infty$. Conclua que a hipótese do Teorema 5 é estritamente mais forte que a do Teorema 10.

∗ **8.** Sejam $X_1, X_2, \ldots$ independentes e identicamente distribuídas com função de distribuição F contínua, e seja F_n a função de distribuição empírica.

(a) Mostre que, para cada x, $F_n(x) \to F(x)$ quase certamente.

(b) Mostre que, quase certamente, $F_n(q) \to F(q)$ simultaneamente para todo racional q.

(c) Para cada $\varepsilon > 0$, escolha um número finito de pontos racionais $x_1 < \cdots < x_k$ que cubra a parte central da distribuição, de modo que as variações de F entre pontos consecutivos e nas duas caudas sejam menores que ε. Justifique a existência da malha escolhendo primeiro extremos racionais com caudas pequenas e usando a continuidade uniforme de F no compacto entre eles. Considere também os extremos $x_0 = -\infty$ e $x_{k+1} = +\infty$, convencionando $F_n(-\infty) = F(-\infty) = 0$ e $F_n(+\infty) = F(+\infty) = 1$. Usando a monotonicidade de F_n e de F, mostre que

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leqslant \max_{0 \leqslant j \leqslant k+1} |F_n(x_j) - F(x_j)| + \varepsilon$$

e conclua o teorema de Glivenko-Cantelli no caso contínuo.

(d) Agora suponha que F possa ter átomos. Explique por que a malha de (c), com variações menores que ε, pode não existir. Mostre que a convergência vale simultaneamente em cada ponto atômico (há no máximo enumeravelmente muitos) e também para os limites à esquerda nesses pontos, observando que

$$F_n(a-) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i < a\}}.$$

Descreva como isolar cada um dos finitos átomos a de massa maior que ε, usando pontos de continuidade $c < a < d$ suficientemente próximos para que $F(a-) - F(c) < \varepsilon$ e $F(d) - F(a) < \varepsilon$; nos intervalos restantes, refine a partição para que cada incremento de F seja, por exemplo, no máximo 2ε. Use a monotonicidade e o controle nos pontos da partição e nos limites à esquerda dos átomos isolados para obter a convergência uniforme.

∗ **9.** Reveja a prova do Teorema 10 e identifique exatamente em que ponto a hipótese de independência dois a dois é usada. Mostre, por um exemplo, que a prova do Teorema 5 não funciona com independência apenas dois a dois, indicando qual passagem falha.

Referências

[1] Patrick Billingsley. *Probability and measure*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, anniversary edition, 2012. With a foreword by Steve Lalley and a brief biography of Billingsley by Steve Koppes.

[2] Rick Durrett. *Probability—theory and examples*, volume 49 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, fifth edition, 2019.

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