UnB | Departamento de Matemática

Modos de Convergência de Sequências de Variáveis Aleatórias

L. Cioletti

Outubro de 2026

Resumo

Nestas notas, estudamos as convergências quase certa, em probabilidade, em média, em média quadrática e em distribuição para sequências de variáveis aleatórias. Demonstramos as principais implicações entre essas noções e apresentamos contra-exemplos para as recíprocas falsas. Discutimos o critério de Borel-Cantelli para a convergência quase certa e o papel da dominação na obtenção das convergências em média e em média quadrática. Como aplicação, deduzimos a Lei Fraca dos Grandes Números sob a hipótese de variância finita. A exposição inclui exemplos e exercícios e segue de perto as referências [1, 2, 3].

1. Introdução

Nas notas anteriores, estudamos uma variável aleatória por vez. Fixado o espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, descrevemos a distribuição de X, calculamos sua esperança, sua variância e seus momentos e usamos essas grandezas para estimar a probabilidade de eventos associados a X. A desigualdade de Chebyshev, por exemplo, transforma informação sobre a variância em uma cota para $\mathbb{P}(|X - \mathbb{E}[X]| \geqslant t)$, com $t > 0$, isto é, para a probabilidade de um desvio de tamanho pelo menos t. Em todas essas situações, o objeto de interesse era um único experimento aleatório.

Nestas notas, o objeto de interesse passa a ser uma sequência $X_1, X_2, X_3, \ldots$ de variáveis aleatórias definidas no mesmo espaço de probabilidade. Sequências aparecem naturalmente em vários problemas. Ao lançar uma moeda repetidamente, podemos acompanhar a frequência relativa de caras nos n primeiros lançamentos e perguntar se ela se estabiliza. Ao repetir um experimento para estimar uma média desconhecida, como na integração de Monte Carlo, calculamos médias de um número crescente de observações e queremos saber se elas se aproximam do valor procurado. Ao considerar o máximo $M_n = \max\{U_1, \ldots, U_n\}$ de n variáveis uniformes em $(0, 1)$, vemos que M_n aumenta com n e parece se aproximar de 1. Em cada um desses exemplos, temos uma sequência de números aleatórios e uma pergunta do tipo: esta sequência se aproxima de algum limite? Em que sentido? Com que nível de incerteza?

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

A expressão "em que sentido" é o ponto central destas notas, porque há várias respostas razoáveis e elas não são equivalentes. A resposta mais direta é exigir que, para cada ponto ω do espaço amostral, a sequência de números reais $X_n(\omega)$ convirja para $X(\omega)$ no sentido usual de sequência de números reais. Essa forma de convergência, como veremos à frente, é fortemente relacionada à convergência quase certa. Esta noção de convergência é mais forte que a convergência em probabilidade e que a convergência em distribuição, mas não é comparável com as convergências em média e em média quadrática, como veremos na Seção 5. Há, porém, duas dificuldades. A primeira é que a convergência pode falhar em um conjunto de pontos, e esse conjunto precisa ser pequeno do ponto de vista da medida de probabilidade, não do ponto de vista da cardinalidade. A segunda é que, em muitos problemas concretos, não há como controlar o comportamento ponto a ponto de toda a sequência: o que sabemos estimar é a probabilidade de que X_n se afaste de X em uma etapa fixada n. Isso sugere uma noção mais fraca, na qual se pede que a probabilidade de um desvio de tamanho ε tenda a zero quando n cresce. Essa é a convergência em probabilidade.

Além dessas duas noções, há as convergências definidas por médias dos erros. Se o valor absoluto do erro, $|X_n - X|$, tem média pequena, então, por Markov, a probabilidade de que esse valor seja grande também é pequena. Assim, exigir que $\mathbb{E}|X_n - X|$ tenda a zero, a convergência em média, ou que $\mathbb{E}[(X_n - X)^2]$ tenda a zero, a convergência em média quadrática, produz noções mais fortes que a convergência em probabilidade. A vantagem dessas noções é que elas se reduzem a cálculos de esperança, com os quais já estamos habituados. A desvantagem é que a média pode ser estragada por poucos pontos onde o erro é enorme, mesmo que a sequência convirja em quase todos os pontos.

Há ainda uma quarta possibilidade, que abandona a comparação entre os valores de X_n e os de X e compara apenas as distribuições. Dizemos que X_n converge em distribuição para X quando as funções de distribuição $F_{X_n}(x) = \mathbb{P}(X_n \leqslant x)$ convergem para $F_X(x)$ em todo ponto x em que F_X é contínua. Essa noção é a mais fraca de todas e é a que aparece no Teorema do Limite Central, quando afirmamos que as somas normalizadas, sob certas condições, têm distribuição aproximadamente normal, sem que isso signifique que elas se aproximem de uma variável normal específica. A convergência em distribuição também é a única das noções que faz sentido quando as variáveis X_n vivem em espaços de probabilidade diferentes, ou quando a variável limite não está definida no mesmo espaço que as demais.

O objetivo destas notas é apresentar as cinco noções, compará-las e fornecer ferramentas para verificar convergência em problemas concretos. A Seção 2 recorda os resultados sobre limites de eventos e o lema de Borel-Cantelli, tratados em notas de aula anteriores. Em seguida, na Seção 3, tratamos da convergência quase certa e discutimos com cuidado a definição por ε e n_0, que costuma ser a primeira dificuldade real do assunto. A Seção 4 introduz a convergência em probabilidade e mostra que ela é consequência da convergência quase certa, mas não é equivalente a ela. A Seção 5 discute as convergências em média e em média quadrática e as comparações entre elas. A Seção 6 introduz a convergência em distribuição e prova que a convergência em probabilidade implica a convergência em distribuição. A Seção 7 reúne todas as implicações em um diagrama que mostra como a convergência quase certa, acompanhada de hipóteses adequadas de dominação, recupera todas as outras noções.

A técnica mais empregada neste texto para obter a convergência quase certa de uma sequência de variáveis aleatórias é a de se estimar séries da forma

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

analisar sua convergência e aplicar o primeiro lema de Borel-Cantelli, quando a série acima for convergente. Fixado $\varepsilon > 0$, cada termo dessa série é a probabilidade de que X_n se afaste de X por pelo menos ε. Para obter a convergência quase certa, precisamos controlar a repetição desses desvios ao longo da sequência. É nesse ponto que a somabilidade é útil: pelo primeiro lema de Borel-Cantelli, se a série converge, então, para quase todo ω, tais desvios ocorrem apenas finitas vezes. Assim, $|X_n(\omega) - X(\omega)| < \varepsilon$ a partir de algum índice. Se a série converge para todo $\varepsilon > 0$, aplicamos esse argumento às tolerâncias $\varepsilon = 1/m$, com $m \geqslant 1$. Como são enumeráveis, podemos excluir um único conjunto de probabilidade zero fora do qual $X_n(\omega) \to X(\omega)$. Esse é o conteúdo da Proposição 8, que será usada em notas de aula futuras sobre a Lei Forte dos Grandes Números.

Por outro lado, se exigirmos apenas que $\mathbb{P}(|X_n - X| \geqslant \varepsilon) \to 0$ para todo $\varepsilon > 0$, obtemos a convergência em probabilidade. Para cada tolerância fixada, essa condição garante que a probabilidade de um desvio seja pequena em todo índice suficientemente grande. Mas os conjuntos de pontos ω em que os desvios ocorrem podem variar com n. Mesmo que suas probabilidades tendam a zero, um mesmo ponto pode pertencer a infinitos desses conjuntos, impedindo a convergência de $X_n(\omega)$ para $X(\omega)$. A somabilidade permite excluir essa possibilidade para quase todo ω, pelo primeiro lema de Borel-Cantelli. Ela é, portanto, uma condição suficiente para a convergência quase certa, mas não necessária, como veremos na Seção 3.6.

A exposição segue de perto as Seções 1 e 2 do Capítulo 1 e a Seção 25 do Capítulo 5 de [1], as Seções 8.1, 8.2 e 8.5 do Capítulo 8 de [3] e a Seção 2.3 de [2]. Os exemplos foram escolhidos para que cada implicação e cada recíproca falsa tenham pelo menos um exemplo concreto associado.

2. Preliminares: o Lema de Borel-Cantelli

Nesta seção, recordamos os resultados sobre sequências de eventos que serão usados em quase todas as demonstrações. As provas estão nas notas de aula *Desigualdades de Markov e Chebyshev e o lema de Borel-Cantelli*, onde também se encontra a discussão completa dos limites superior e inferior. Aqui repetimos os enunciados para conveniência do leitor e acrescentamos alguns comentários necessários ao uso que faremos destes resultados.

Definição 1 (Limite superior de uma sequência de eventos)**.** Seja $(A_n)_{n\geqslant 1}$ uma sequência de eventos. O *limite superior* de (A_n) é o evento

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Escrevemos também $\{A_n \text{ infinitas vezes}\}$ ou $\{A_n \text{ i.v.}\}$ para denotar este evento.

Um ponto ω pertence a $\bigcup_{k\geqslant n} A_k$ quando pertence a pelo menos um A_k com $k \geqslant n$. Dizer que ω está na interseção sobre todos os n significa, portanto, que para cada n existe um índice $k \geqslant n$ tal que $\omega \in A_k$, isto é, que os índices k para os quais $\omega \in A_k$ não são limitados. Em palavras,

$$\limsup_{n\to\infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ para infinitos índices } n\}.$$

O limite inferior tem definição dual e corresponde a "a partir de algum momento":

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \{\omega \in \Omega : \omega \in A_n \text{ para todo } n \text{ suficientemente grande}\}.$$

Temos $\liminf_n A_n \subseteq \limsup_n A_n$, e o complementar troca um limite pelo outro:

$$\left(\limsup_{n\to\infty} A_n\right)^c = \liminf_{n\to\infty} A_n^c.$$

Teorema 2 (Primeiro lema de Borel-Cantelli)**.** Seja $(A_n)_{n\geqslant 1}$ uma sequência de eventos tal que $\sum_{n\geqslant 1}\mathbb{P}(A_n) < \infty$. Então

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0.$$

Em outras palavras, com probabilidade 1 ocorrem apenas finitos dos eventos A_n.

Enfatizamos que para aplicar o lema acima não precisamos da hipótese de independência dos eventos. Este primeiro lema de Borel-Cantelli vale para qualquer sequência de eventos. Por outro lado, para aplicar o segundo lema é importante verificar a hipótese de independência.

Teorema 3 (Segundo lema de Borel-Cantelli)**.** Seja $(A_n)_{n\geqslant 1}$ uma sequência de eventos independentes com $\sum_{n\geqslant 1}\mathbb{P}(A_n) = \infty$. Então

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

A prova, que está nas notas citadas, usa a desigualdade $1 - x \leqslant e^{-x}$ para estimar $\mathbb{P}\left(\bigcap_{n=M}^N A_n^c\right)$ e depois a continuidade da medida. Não repetiremos o argumento porque ele não será usado diretamente aqui; o segundo lema aparecerá apenas em exemplos e na discussão da recíproca do critério de Borel-Cantelli.

Encerramos a seção com uma observação sobre o uso desses resultados. O primeiro lema transforma uma informação sobre a *soma* das probabilidades em uma informação sobre o *conjunto* dos pontos onde os eventos ocorrem infinitas vezes. Essa passagem é uma maneira elementar particularmente útil de obter convergência quase certa a partir de estimativas de probabilidade, e é por isso que a série $\sum_n \mathbb{P}(A_n)$ aparecerá tantas vezes no que segue.

3. Convergência Quase Certa

3.1. A Definição

A primeira noção de convergência que estudaremos é a mais próxima da convergência de sequências de números reais. Fixado ω, a sequência $X_1(\omega), X_2(\omega), \ldots$ é uma sequência de números reais, e podemos perguntar se ela converge no sentido usual. A convergência quase certa pede que isso ocorra para quase todo ω.

Definição 4 (Convergência quase certa)**.** Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias definidas em um mesmo espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$. Dizemos que (X_n) *converge quase certamente* para X se

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

Nesse caso, escrevemos $X_n \xrightarrow{\text{q.c.}} X$ ou, simplesmente, $X_n \to X$ quase certamente.

O conjunto que aparece na definição é o conjunto dos pontos onde a sequência numérica converge. Denotaremos esse conjunto por

$$E \equiv \Big\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\Big\}.$$

A definição pede que E tenha probabilidade 1, ou seja, que seu complementar tenha probabilidade zero. Os pontos de E^c são os pontos onde a sequência não converge para $X(\omega)$, e a exigência é que eles formem um conjunto pequeno do ponto de vista da medida $\mathbb{P}$. Note que E^c pode ser não vazio e pode até ser não enumerável; o que a definição controla é a sua probabilidade.

Antes de usar a definição, precisamos saber que E é um evento, isto é, que $E \in \mathcal{F}$. A próxima proposição resolve essa questão e, ao mesmo tempo, traduz a definição em uma forma que será usada em todas as demonstrações.

3.2. O Conjunto de Convergência em Termos de Eventos

Proposição 5 (Caracterização da convergência quase certa)**.** Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias. Então

$$E = \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \Big\{|X_n - X| < \frac{1}{m}\Big\}, \tag{1}$$

e, equivalentemente,

$$E^c = \bigcup_{m=1}^{\infty} \limsup_{n\to\infty} \Big\{|X_n - X| \geqslant \frac{1}{m}\Big\}. \tag{2}$$

Em particular, $E \in \mathcal{F}$ e

$$X_n \xrightarrow{\text{q.c.}} X \qquad \Longleftrightarrow \qquad \mathbb{P}\Big(\limsup_{n\to\infty} \big\{|X_n - X| \geqslant \tfrac{1}{m}\big\}\Big) = 0 \quad \text{para todo } m \geqslant 1.$$

Prova. Vamos demonstrar a identidade (1) verificando as duas inclusões. Suponha primeiro que $\omega \in E$ e fixe $m \geqslant 1$. Pela definição de E, temos $X_n(\omega) \to X(\omega)$. Aplicando a definição de convergência à tolerância $1/m$, obtemos um índice $N \in \mathbb{N}$ tal que $|X_n(\omega) - X(\omega)| < 1/m$ para todo $n \geqslant N$. A validade dessa desigualdade para todos os índices a partir de N significa que

$$\omega \in \bigcap_{n=N}^{\infty} \Big\{|X_n - X| < \frac{1}{m}\Big\}.$$

O índice N pode depender de m e de ω. A existência de um tal índice é expressa pela união sobre N:

$$\omega \in \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \Big\{|X_n - X| < \frac{1}{m}\Big\}.$$

Como $m \geqslant 1$ foi arbitrário, ω pertence a essa união para todo $m \geqslant 1$. Portanto, ω pertence à interseção dessas uniões:

$$\omega \in \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \Big\{|X_n - X| < \frac{1}{m}\Big\}.$$

Como ω foi escolhido arbitrariamente em E, concluímos que

$$E \subseteq \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \Big\{|X_n - X| < \frac{1}{m}\Big\}.$$

Isso prova a primeira inclusão necessária para (1).

Para provar a inclusão contrária, suponha que ω pertence ao conjunto do lado direito da identidade (1). Vamos mostrar que $X_n(\omega) \to X(\omega)$ quando $n \to \infty$.

Fixe $\varepsilon > 0$. Como $1/m \to 0$, existe $m \geqslant 1$ tal que $1/m < \varepsilon$. Por hipótese, ω pertence a todos os conjuntos da interseção em m. Em particular, para o m que acabamos de escolher, temos

$$\omega \in \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{ |X_n - X| < \frac{1}{m} \right\}.$$

Pertencer a essa união significa pertencer a pelo menos um dos conjuntos que a formam. Assim, existe $N \in \mathbb{N}$ tal que

$$\omega \in \bigcap_{n=N}^{\infty} \left\{ |X_n - X| < \frac{1}{m} \right\}.$$

Isso significa que a desigualdade vale para todos os índices a partir de N. Portanto,

$$|X_n(\omega) - X(\omega)| < \frac{1}{m} < \varepsilon \qquad \text{para todo } n \geqslant N.$$

Como $\varepsilon > 0$ foi arbitrário, concluímos que $X_n(\omega) \to X(\omega)$ e, portanto, $\omega \in E$. Isto mostra que

$$\bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{ |X_n - X| < \frac{1}{m} \right\} \subseteq E.$$

As duas inclusões demonstradas acima nos permitem concluir que vale a identidade (1).

A identidade (1) também mostra que E é um evento. De fato, como X_n e X são variáveis aleatórias, cada conjunto $\{|X_n - X| < 1/m\}$ pertence a $\mathcal{F}$. Pela identidade, E é obtido a partir desses conjuntos por uniões e interseções enumeráveis. Como $\mathcal{F}$ é uma σ-álgebra, essas operações produzem um conjunto que ainda pertence a $\mathcal{F}$. Logo $E \in \mathcal{F}$.

Vamos agora calcular E^c. Para isso, tomamos o complementar dos dois lados de (1) e usamos as leis de De Morgan: o complementar de uma interseção é a união dos complementares, e o complementar de uma união é a interseção dos complementares. Além disso, o complementar de $\{|X_n - X| < 1/m\}$ é $\{|X_n - X| \geqslant 1/m\}$. Assim,

$$\begin{aligned} E^c &= \left(\bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{ |X_n - X| < \frac{1}{m} \right\} \right)^c \\ &= \bigcup_{m=1}^{\infty} \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \left\{ |X_n - X| \geqslant \frac{1}{m} \right\} \\ &= \bigcup_{m=1}^{\infty} \limsup_{n \to \infty} \left\{ |X_n - X| \geqslant \frac{1}{m} \right\}. \end{aligned}$$

Na última igualdade usamos a definição de limite superior de eventos. Isso prova (2). Cada limite superior nessa fórmula é um evento, pois é uma interseção enumerável de uniões enumeráveis de eventos.

Resta demonstrar a equivalência que aparece no final do enunciado. Suponha primeiro que $X_n \to X$ quase certamente. Então $\mathbb{P}(E) = 1$ e $\mathbb{P}(E^c) = 0$. Por (2), cada evento da união está contido em E^c. Logo, para todo $m \geqslant 1$,

$$0 \leqslant \mathbb{P}\Big(\limsup_{n\to\infty}\Big\{|X_n - X| \geqslant \tfrac{1}{m}\Big\}\Big) \leqslant \mathbb{P}(E^c) = 0.$$

Reciprocamente, suponha que cada uma dessas probabilidades seja zero. Usando (2) e a desigualdade de Boole, obtemos

$$\mathbb{P}(E^c) \leqslant \sum_{m=1}^{\infty} \mathbb{P}\Big(\limsup_{n\to\infty}\Big\{|X_n - X| \geqslant \tfrac{1}{m}\Big\}\Big) = 0.$$

Portanto, $\mathbb{P}(E) = 1$, isto é, $X_n \to X$ quase certamente. ■

A proposição anterior mostra como provar a convergência quase certa usando apenas os números $1, 1/2, 1/3, \ldots$. Para cada $m \geqslant 1$, consideramos os pontos ω em que a desigualdade $|X_n(\omega) - X(\omega)| \geqslant 1/m$ vale para infinitos valores de n. Precisamos mostrar que esse conjunto tem probabilidade zero.

Se cada um desses conjuntos tem probabilidade zero, a desigualdade de Boole garante que sua união também tem probabilidade zero, pois a união é enumerável. Fora dessa união, para cada m, o erro é menor que $1/m$ a partir de algum índice. Isso implica $X_n(\omega) \to X(\omega)$, já que dado $\varepsilon > 0$, basta escolher m com $1/m < \varepsilon$.

Se considerássemos diretamente todos os $\varepsilon > 0$, teríamos uma família não enumerável de conjuntos excepcionais e não poderíamos aplicar o mesmo argumento. Uma união não enumerável de conjuntos de probabilidade zero pode ter probabilidade positiva. Por exemplo, na medida uniforme em $(0, 1)$, cada conjunto de um único ponto tem probabilidade zero, mas a união de todos eles é o intervalo inteiro, de probabilidade 1. A sequência $1/m$ permite evitar essa dificuldade: ela tende a zero e tem apenas uma quantidade enumerável de termos.

3.3. A Definição de Convergência em Detalhes

Nesta subseção, discutimos a definição com o cuidado que ela merece, porque é aqui que aparece a primeira dificuldade real do assunto. A dificuldade não está na matemática, mas na ordem dos quantificadores.

Em análise, dizemos que $x_n \to x$ quando

$$\text{para todo } \varepsilon > 0 \text{ existe } N \in \mathbb{N} \text{ tal que } |x_n - x| < \varepsilon \text{ para todo } n \geqslant N. \tag{3}$$

A negação dessa afirmação é

$$\text{existe } \varepsilon > 0 \text{ tal que, para todo } N \in \mathbb{N}, \text{ existe } n \geqslant N \text{ com } |x_n - x| \geqslant \varepsilon. \tag{4}$$

O ponto importante é que, em (3), o número N pode depender de ε, e em (4) o índice n pode depender de N. Ao passar para variáveis aleatórias, acrescentamos uma dependência adicional: N pode depender de ω. A definição de convergência quase certa não pede que exista um único N que sirva para todos os pontos; ela pede que, para cada ponto de um conjunto de probabilidade 1, exista um N próprio daquele ponto.

Para ver como essa dependência se traduz em eventos, fixe $\varepsilon > 0$ e considere o conjunto dos pontos para os quais a desigualdade $|X_n - X| < \varepsilon$ vale a partir de algum índice:

$$B_\varepsilon \equiv \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \big\{|X_n - X| < \varepsilon\big\}.$$

Um ponto ω pertence a B_ε se, e somente se, existe N tal que $|X_n(\omega) - X(\omega)| < \varepsilon$ para todo $n \geqslant N$. Isso é exatamente a afirmação "a tolerância ε é respeitada a partir de um certo momento", com o momento dependendo de ω. O evento B_ε é, portanto, a versão probabilística da condição que aparece em (3) para uma tolerância fixada.

A convergência em ω exige que essa condição valha para todas as tolerâncias ao mesmo tempo. Assim,

$$E = \bigcap_{\varepsilon > 0} B_\varepsilon.$$

Essa interseção não é enumerável, mas a monotonicidade de B_ε em ε permite substituí-la por uma interseção enumerável: se $B_{1/m}$ ocorre para todo m, então B_ε ocorre para todo $\varepsilon > 0$, porque basta escolher m com $1/m < \varepsilon$ e usar $B_{1/m} \subseteq B_\varepsilon$. Logo

$$E = \bigcap_{m=1}^{\infty} B_{1/m},$$

que é a identidade (1).

Resta entender por que a interseção enumerável preserva a probabilidade 1. Suponha que $\mathbb{P}(B_{1/m}) = 1$ para todo m. Então $\mathbb{P}(B_{1/m}^c) = 0$ para todo m e, pela subaditividade enumerável,

$$\mathbb{P}\Big(\bigcup_{m=1}^{\infty} B_{1/m}^c\Big) \leqslant \sum_{m=1}^{\infty} \mathbb{P}(B_{1/m}^c) = 0.$$

Tomando complementares, $\mathbb{P}(E) = 1$. Na preservação da probabilidade 1, a enumerabilidade é usada neste passo, por meio da subaditividade enumerável. É também aqui que muitos argumentos informais erram. Não é verdade que uma interseção arbitrária de eventos de probabilidade 1 tenha probabilidade 1; é verdade para interseções enumeráveis.

A negação da convergência em um ponto ω, que será usada para construir contraexemplos, tem a seguinte forma. O ponto ω não pertence a E se, e somente se, existe $m \geqslant 1$ tal que, para todo N, existe $n \geqslant N$ com $|X_n(\omega) - X(\omega)| \geqslant 1/m$. Em símbolos,

$$\omega \in E^c \qquad \Longleftrightarrow \qquad \exists\, m \geqslant 1 \;\; \forall N \geqslant 1 \;\; \exists\, n \geqslant N : \quad |X_n(\omega) - X(\omega)| \geqslant \frac{1}{m}.$$

Essa é a versão de (4) com a dependência em ω explicitada. A diferença entre

$$\forall m \; \exists N \; \forall n \geqslant N \qquad \text{e} \qquad \exists m \; \forall N \; \exists n \geqslant N$$

é toda a diferença entre convergência e não convergência, e a segunda forma é a que devemos verificar quando queremos mostrar que uma sequência *não* converge em um ponto. Na segunda forma, a tolerância que testemunha a falha é uma só, escolhida uma vez; ela não precisa funcionar para todos os m.

Vale registrar uma consequência prática da discussão. Para provar que $X_n \to X$ quase certamente, é suficiente exibir, para cada m, um evento B_m de probabilidade 1 tal que, em B_m, a desigualdade $|X_n - X| < 1/m$ valha para todo n suficientemente grande. Em outras palavras, podemos trabalhar com uma tolerância de cada vez, desde que o façamos para todas as tolerâncias da sequência $1/m$ e que os eventos excepcionais sejam enumeráveis.

Essa leitura também mostra o que a convergência em probabilidade não fornece. Ela dá, para cada tolerância fixada, a informação de que a probabilidade do evento ruim é pequena em cada etapa, mas não diz nada sobre a organização dos índices ruins ao longo do tempo. O exemplo da sequência que varre o intervalo, na Seção 4.4, mostrará que os conjuntos ruins podem se deslocar indefinidamente, de modo que cada ponto seja atingido infinitas vezes ainda que a probabilidade de cada conjunto tenda a zero.

3.4. Primeiros Exemplos

Exemplo 6. Seja X uma variável aleatória qualquer e defina $X_n = X + 1/n$. Então $X_n \to X$ quase certamente. De fato, para todo ω,

$$|X_n(\omega) - X(\omega)| = \frac{1}{n} \longrightarrow 0,$$

de modo que $E = \Omega$ e $\mathbb{P}(E) = 1$. O exemplo é trivial, mas serve para fixar a notação: a convergência quase certa é uma afirmação sobre a sequência numérica obtida em cada ponto, e não sobre a distribuição das variáveis.

Exemplo 7 (O máximo de variáveis uniformes)**.** Sejam $U_1, U_2, \ldots$ variáveis aleatórias independentes com distribuição uniforme em $(0, 1)$ e seja $M_n = \max\{U_1, \ldots, U_n\}$. Então $M_n \to 1$ quase certamente.

Para calcular a probabilidade de um desvio, observamos que o máximo de dois números é menor ou igual a z exatamente quando ambos são menores ou iguais a z. Para variáveis aleatórias X e Y, isso significa que, para cada $\omega \in \Omega$ e $z \in \mathbb{R}$,

$$\max\{X(\omega), Y(\omega)\} \leqslant z \quad \Longleftrightarrow \quad X(\omega) \leqslant z \text{ e } Y(\omega) \leqslant z.$$

Logo, os dois eventos abaixo contêm exatamente os mesmos pontos:

$$\Big\{\max\{X, Y\} \leqslant z\Big\} = \{X \leqslant z\} \cap \{Y \leqslant z\}.$$

Com n variáveis, precisamos que a desigualdade valha para todas elas. Portanto, para $\varepsilon \in (0, 1)$,

$$\{M_n \leqslant 1 - \varepsilon\} = \bigcap_{k=1}^{n} \{U_k \leqslant 1 - \varepsilon\}.$$

Como as variáveis U_k são independentes, os eventos dessa interseção também são independentes. Além disso, como a distribuição de cada U_k é uniforme em $(0, 1)$, cada evento tem probabilidade $1 - \varepsilon$. Usando ainda que $M_n < 1$ quase certamente, obtemos

$$\begin{aligned}
\mathbb{P}\Big(|M_n - 1| \geqslant \varepsilon\Big) &= \mathbb{P}\Big(M_n \leqslant 1 - \varepsilon\Big) \\
&= \mathbb{P}\Big(\bigcap_{k=1}^{n} \{U_k \leqslant 1 - \varepsilon\}\Big) \\
&= \prod_{k=1}^{n} \mathbb{P}\Big(U_k \leqslant 1 - \varepsilon\Big) \\
&= (1 - \varepsilon)^n.
\end{aligned}$$

Para cada $m \geqslant 2$, tome $\varepsilon = 1/m$. A série geométrica $\sum_n (1 - 1/m)^n$ converge, logo o primeiro lema de Borel-Cantelli garante que

$$\mathbb{P}\Big(|M_n - 1| \geqslant \tfrac{1}{m} \text{ infinitas vezes}\Big) = 0.$$

Para $m = 1$, o evento correspondente também tem probabilidade zero, pois $0 < M_n < 1$ quase certamente. Como isso vale para cada $m \geqslant 1$ e a união sobre m é enumerável, o conjunto em que algum desses eventos ocorre tem probabilidade zero. Pela Proposição 5, $M_n \to 1$ quase certamente.

O exemplo anterior antecipa o critério da próxima subseção: a convergência quase certa foi obtida a partir da somabilidade das probabilidades dos desvios. O cálculo de $\mathbb{P}(M_n \leqslant 1-\varepsilon)$ é elementar porque as variáveis são independentes, e a somabilidade decorre de a probabilidade decair geometricamente. Nos problemas de médias, a probabilidade de um desvio decai apenas como $1/n$, e a série diverge; será preciso um argumento mais fino, que é o conteúdo das notas de aula sobre a lei forte.

3.5. O Critério de Borel-Cantelli

O resultado a seguir é a ferramenta padrão para provar convergência quase certa. Ele transforma uma estimativa sobre cada n em uma conclusão sobre o comportamento de toda a sequência.

Proposição 8 (Critério de Borel-Cantelli)**.** Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias. Suponha que, para todo $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}\Big(|X_n - X| \geqslant \varepsilon\Big) < \infty. \tag{5}$$

Então $X_n \to X$ quase certamente.

Prova. Fixe $m \geqslant 1$ e defina $A_{n,m} = \{|X_n - X| \geqslant 1/m\}$. A hipótese, aplicada a $\varepsilon = 1/m$, diz que $\sum_n \mathbb{P}(A_{n,m}) < \infty$. Pelo Teorema 2,

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

Como isso vale para cada m, a união enumerável $\bigcup_m \limsup_n A_{n,m}$ também tem probabilidade zero. Pela Proposição 5, esse conjunto é exatamente E^c. Logo $\mathbb{P}(E) = 1$, isto é, $X_n \to X$ quase certamente. ∎

A leitura da prova esclarece o papel de cada hipótese. O primeiro lema de Borel-Cantelli é aplicado a uma tolerância fixa; a passagem para todas as tolerâncias é feita pela união enumerável em m; e a identificação do conjunto excepcional com E^c é feita pela Proposição 5. Nenhuma hipótese de independência é usada, o que é coerente com o fato de o primeiro lema não exigir independência.

Vale comparar (5) com a definição de convergência em probabilidade, que estudaremos na próxima seção. Aquela definição pede apenas que cada termo $\mathbb{P}(|X_n - X| \geqslant \varepsilon)$ tenda a zero. A hipótese (5) pede que a soma desses termos seja finita, o que é estritamente mais forte. A diferença é a mesma que existe entre dizer que uma sequência de números não negativos tende a zero e dizer que ela é somável: a segunda afirmação controla o conjunto dos índices, e não apenas o termo genérico.

Na prática, raramente verificamos (5) para todos os ε. O corolário seguinte mostra que basta uma sequência de tolerâncias que tende a zero.

Corolário 9 (Critério com tolerâncias variáveis)**.** Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias e seja (ε_n) uma sequência de números positivos com $\varepsilon_n \to 0$. Se

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

então $X_n \to X$ quase certamente.

Prova. Fixe $\varepsilon > 0$. Como $\varepsilon_n \to 0$, existe n_0 tal que $\varepsilon_n < \varepsilon$ para todo $n \geqslant n_0$. Para esses índices,

$$\Big\{|X_n - X| \geqslant \varepsilon\Big\} \subseteq \Big\{|X_n - X| \geqslant \varepsilon_n\Big\},$$

logo

$$\sum_{n=n_0}^{\infty} \mathbb{P}\Big(|X_n - X| \geqslant \varepsilon\Big) \leqslant \sum_{n=n_0}^{\infty} \mathbb{P}\Big(|X_n - X| \geqslant \varepsilon_n\Big) < \infty.$$

Os primeiros $n_0 - 1$ termos formam uma soma finita e cada um deles é no máximo 1, logo a série completa $\sum_{n=1}^{\infty}$ também converge. Agora, para finalizar a prova, basta aplicar a Proposição 8. ■

O corolário fornece também um controle do erro. Pelo primeiro lema de Borel-Cantelli, a hipótese de somabilidade garante que, para quase todo ω,

$$|X_n(\omega) - X(\omega)| < \varepsilon_n$$

para todo n suficientemente grande. Como $\varepsilon_n \to 0$, essa desigualdade implica $X_n(\omega) \to X(\omega)$. Portanto, além de provar a convergência quase certa, obtemos uma sequência que limita o erro a partir de algum índice.

3.6. O Critério Não É Necessário

A hipótese (5) é suficiente, mas não é necessária. O exemplo a seguir mostra uma sequência que converge quase certamente e para a qual a soma das probabilidades dos desvios diverge.

Exemplo 10 (Convergência quase certa com série divergente)**.** Seja $\Omega = (0, 1]$ com a medida de Lebesgue e defina $A_n = (0, 1/n)$ e $X_n = \mathbb{1}_{A_n}$. Para cada $\omega \in (0, 1]$, existe n_0 tal que $1/n < \omega$ para todo $n \geqslant n_0$; para esses índices, $\omega \notin A_n$ e $X_n(\omega) = 0$. Logo $X_n(\omega) \to 0$ para todo ω, e a convergência é quase certa (na verdade, vale em todo ponto). Por outro lado, para $0 < \varepsilon \leqslant 1$,

$$\mathbb{P}\Big(|X_n - 0| \geqslant \varepsilon\Big) = \mathbb{P}(A_n) = \frac{1}{n},$$

de modo que $\sum_n \mathbb{P}(|X_n| \geqslant \varepsilon) = \infty$. A hipótese do critério falha, mas a conclusão vale.

O que distingue o exemplo anterior de uma situação em que o critério é necessário é a estrutura dos eventos. No exemplo, os eventos A_n são encaixados: $A_{n+1} \subseteq A_n$. Cada ponto sai de todos os A_n a partir de um certo índice, e a probabilidade de A_n tende a zero sem que a soma convirja. O primeiro lema de Borel-Cantelli não se aplica porque ele é um critério baseado na soma, e a soma é grande justamente porque os eventos se repetem nos mesmos pontos. Compare com o exemplo a seguir, no qual as probabilidades são as mesmas, mas os eventos são independentes.

Exemplo 11 (Desvios independentes com série divergente)**.** Seja (A_n) uma sequência de eventos independentes com $\mathbb{P}(A_n) = 1/n$ e $X_n = \mathbb{1}_{A_n}$. Pelo segundo lema de Borel-Cantelli, $\mathbb{P}(\limsup_n A_n) = 1$. Os eventos A_n^c também são independentes e $\sum_n \mathbb{P}(A_n^c) = \infty$, já que $\mathbb{P}(A_n^c) = 1 - 1/n$ para $n \geqslant 2$; aplicando de novo o segundo lema, $\mathbb{P}(\limsup_n A_n^c) = 1$. Seja ω na interseção dos dois limites superiores, que tem probabilidade 1. Então $X_n(\omega) = 1$ para infinitos índices e $X_n(\omega) = 0$ para infinitos índices, de modo que a sequência não converge em ω. Logo a probabilidade de convergência é zero.

Os dois exemplos têm a mesma distribuição marginal para cada X_n e conduzem a conclusões opostas. Isso mostra que a convergência quase certa não é uma propriedade da coleção de distribuições individuais: ela depende da estrutura de dependência entre as variáveis, isto é, da distribuição conjunta da sequência. A convergência em distribuição, que estudaremos na Seção 6, é justamente a noção que ignora essa estrutura.

4. Convergência em Probabilidade

4.1. A Definição

A convergência em probabilidade é a noção que corresponde à pergunta "qual é a chance de que X_n se afaste de X em uma etapa fixada?". Ela não tenta controlar o comportamento simultâneo de todos os índices, e é justamente por isso que é mais fraca que a convergência quase certa.

Definição 12 (Convergência em probabilidade)**.** Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias em um mesmo espaço de probabilidade. Dizemos que $(X_n)_{n\in\mathbb{N}}$ *converge em probabilidade* para X se, para todo $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\Big(|X_n - X| \geqslant \varepsilon\Big) = 0.$$

Nesse caso, escrevemos $X_n \xrightarrow{\mathbb{P}} X$.

A definição é equivalente a dizer que, para todos $\varepsilon > 0$ e $\delta > 0$, existe n_0 tal que

$$\mathbb{P}\Big(|X_n - X| < \varepsilon\Big) \geqslant 1 - \delta \qquad \text{para todo } n \geqslant n_0.$$

Essa forma é útil quando queremos exibir n_0 explicitamente. A forma com o limite é mais compacta e será a que usaremos nas demonstrações.

Duas observações sobre a definição. A primeira é que o evento cuja probabilidade se controla depende de ε, mas não de ω: a exigência é que, em cada etapa, a probabilidade do desvio seja pequena, e não que o desvio seja pequeno em cada ponto. A segunda é que a escolha de $\geqslant$ ou $>$ no evento não altera a definição, porque $\{|X_n - X| \geqslant \varepsilon\} \subseteq \{|X_n - X| > \varepsilon'\}$ para todo $\varepsilon' < \varepsilon$. Usaremos $\geqslant$ para manter a coerência com as desigualdades de Markov e Chebyshev, que são enunciadas nessa forma.

4.2. Exemplos

Exemplo 13 (Convergência em probabilidade de uma translação)**.** Seja X uma variável aleatória e $X_n = X + 1/n$. Para todo $\varepsilon > 0$ e todo $n > 1/\varepsilon$,

$$\mathbb{P}\Big(|X_n - X| \geqslant \varepsilon\Big) = \mathbb{P}\Big(\frac{1}{n} \geqslant \varepsilon\Big) = 0,$$

logo $X_n \to X$ em probabilidade. O exemplo é o mesmo do Exemplo 6, e nele a convergência vale em todas as noções que estudaremos.

Exemplo 14 (A frequência de um resultado no dado)**.** Lancemos um dado honesto n vezes, de forma independente, e seja N_n o número de vezes em que o resultado é 5 ou 6. Então $N_n \sim \mathrm{Bin}(n, 1/3)$ e, consequentemente,

$$\mathbb{E}[N_n] = \frac{n}{3} \qquad \text{e} \qquad \mathrm{Var}(N_n) = \frac{2n}{9}.$$

Pela desigualdade de Chebyshev aplicada à variável aleatória N_n/n, concluímos que

$$\mathbb{P}\Big(\Big|\frac{N_n}{n} - \frac{1}{3}\Big| \geqslant \varepsilon\Big) \leqslant \frac{1}{\varepsilon^2}\operatorname{Var}\Big(\frac{N_n}{n}\Big) = \frac{1}{\varepsilon^2}\cdot\frac{2n/9}{n^2} = \frac{2}{9n\varepsilon^2} \longrightarrow 0.$$

Logo $N_n/n \to 1/3$ em probabilidade. O número $1/3$ é a probabilidade de o resultado ser 5 ou 6 em um lançamento, e o exemplo é, como veremos mais à frente, uma versão mais simples da chamada Lei dos Grandes Números.

4.3. Convergências Quase Certa e em Probabilidade

A convergência quase certa é mais forte que a convergência em probabilidade. Como veremos abaixo, a prova é curta e baseada na caracterização dada pela Proposição 5 e na continuidade da medida.

Proposição 15. Se $X_n \to X$ quase certamente, então $X_n \to X$ em probabilidade.

Prova. Fixe $\varepsilon > 0$. Escolha m com $1/m < \varepsilon$ e defina os eventos

$$A_n = \{|X_n - X| \geqslant \varepsilon\} \qquad \text{e} \qquad A_{n,m} = \{|X_n - X| \geqslant 1/m\}.$$

Como $A_n \subseteq A_{n,m}$, temos $\limsup_n A_n \subseteq \limsup_n A_{n,m}$. Pela Proposição 5,

$$\mathbb{P}(\limsup_n A_{n,m}) = 0 \qquad \Longrightarrow \qquad \mathbb{P}(\limsup_n A_n) = 0.$$

Considere a sequência decrescente

$$B_n = \bigcup_{k=n}^{\infty} A_k.$$

Temos $A_n \subseteq B_n$ e $\bigcap_n B_n = \limsup_n A_n$. Pela continuidade da medida

$$\lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}\Big(\bigcap_{n=1}^{\infty} B_n\Big) = \mathbb{P}\Big(\limsup_{n\to\infty} A_n\Big) = 0.$$

Como $0 \leqslant \mathbb{P}(A_n) \leqslant \mathbb{P}(B_n)$, segue que $\mathbb{P}(A_n) \to 0$, isto é, $X_n \to X$ em probabilidade. ■

Vale sublinhar a estrutura da prova, porque ela se repete. Para passar de "quase certamente" para "em probabilidade", não comparamos X_n com X diretamente. Em vez disso, usamos que o evento $\limsup_n A_n$, no qual ocorrem infinitos desvios de tamanho pelo menos ε, tem probabilidade zero, e que a probabilidade de A_n é majorada pela probabilidade da cauda B_n, que converge para a probabilidade do limite superior. A continuidade da medida faz a ponte entre a informação sobre o limite superior e a informação sobre cada termo.

4.4. As Duas Noções Não São Equivalentes

A convergência em probabilidade não implica convergência quase certa. O contraexemplo padrão é uma sequência de indicadoras de intervalos que percorrem o intervalo $(0, 1]$ em blocos de comprimento decrescente. Em cada etapa, a probabilidade de o indicador valer 1 é pequena, mas cada ponto é visitado infinitas vezes.

Exemplo 16 (A sequência que varre o intervalo)**.** Seja $\Omega = (0, 1]$ com a medida de Lebesgue. Para cada $k \geqslant 1$, particione $(0, 1]$ nos intervalos

$$I_{k,j} = \left(\frac{j-1}{k}, \frac{j}{k}\right], \qquad j = 1, \ldots, k.$$

A Figura 1 ilustra o efeito de variar j com k fixo: a posição do intervalo muda, mas seu comprimento permanece igual a $1/k$. Por outro lado, com j fixo, aumentar k diminui esse comprimento e aproxima as extremidades de 0.

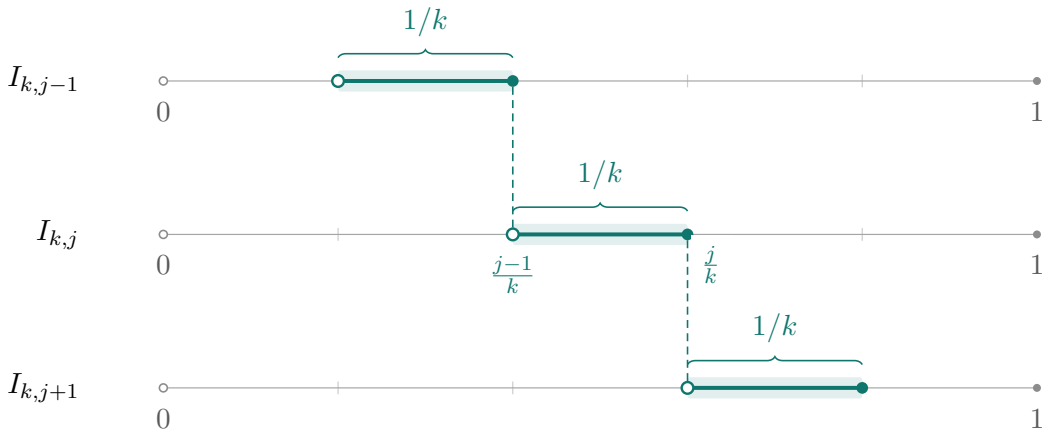


Figura 1: Os intervalos $I_{k,j}$.

Para construir uma sequência a partir desses intervalos, vamos colocá-los em uma única lista. Começamos pelo intervalo do bloco $k = 1$, depois listamos os dois intervalos do bloco $k = 2$ e assim por diante. Dentro de cada bloco, seguimos a ordem crescente de j:

$$\underbrace{I_{1,1}}_{\text{bloco 1}}, \quad \underbrace{I_{2,1},\ I_{2,2}}_{\text{bloco 2}}, \quad \underbrace{I_{3,1},\ I_{3,2},\ I_{3,3}}_{\text{bloco 3}}, \quad \ldots$$

Assim, k indica o bloco e j indica a posição do intervalo dentro dele. Antes de iniciar o bloco k, já listamos

$$1 + 2 + \cdots + (k-1) = \frac{k(k-1)}{2}$$

intervalos. Como $I_{k,j}$ é o j-ésimo intervalo desse bloco, sua posição na lista é

$$n(k, j) = \frac{k(k-1)}{2} + j.$$

Por exemplo, $I_{3,2}$ ocupa a posição $n(3, 2) = 3 + 2 = 5$. Definimos, então,

$$X_n = \mathbb{1}_{I_{k,j}} \qquad \text{quando } n = n(k, j).$$

Então $X_n \to 0$ em probabilidade, mas (X_n) não converge em nenhum ponto de Ω.

Prova. Para $n = n(k, j)$ temos $\mathbb{P}(X_n = 1) = |I_{k,j}| = 1/k$. Dado $\varepsilon \in (0, 1]$,

$$\mathbb{P}\Big(|X_n - 0| \geqslant \varepsilon\Big) = \mathbb{P}(X_n = 1) = \frac{1}{k} \longrightarrow 0 \qquad \text{quando } n \to \infty,$$

porque $k \to \infty$ ao longo dos blocos. Logo $X_n \to 0$ em probabilidade.

Fixe agora $\omega \in (0,1]$. Em cada bloco k, o ponto ω pertence a exatamente um dos intervalos $I_{k,j}$, digamos $I_{k,j(k)}$, e não pertence aos demais. No bloco k, portanto, $X_n(\omega) = 1$ para um índice e $X_n(\omega) = 0$ para os outros $k-1$ índices. Como há infinitos blocos, a sequência $X_n(\omega)$ vale 1 em infinitos índices e 0 em infinitos índices. Uma sequência de números reais que assume os valores 0 e 1 infinitas vezes não converge. Logo $(X_n(\omega))$ não converge para nenhum número real, e a convergência quase certa falha em todos os pontos. ■

Vale entender por que o exemplo funciona, e não apenas verificar as contas. A convergência em probabilidade olha para a etapa n e pergunta qual é a massa do conjunto onde X_n é grande. No exemplo, essa massa é $1/k$, que tende a zero. Mas os conjuntos onde X_n é grande não param de se mover: cada ponto é coberto uma vez em cada bloco. A convergência quase certa olha para cada ponto e pergunta se o conjunto dos índices em que ele é coberto é finito. No exemplo, esse conjunto é infinito para todos os pontos. As duas perguntas são diferentes, e o exemplo as separa.

Observação 17 (Subsequências)**.** Embora a convergência em probabilidade não implique convergência quase certa da sequência inteira, ela implica a existência de uma subsequência que converge quase certamente. Mais precisamente, $X_n \to X$ em probabilidade se, e somente se, toda subsequência de (X_n) possui uma subsubsequência que converge quase certamente para X. A implicação da esquerda para a direita se prova escolhendo adequadamente os índices n_k com $\mathbb{P}(|X_{n_k} - X| \geqslant 1/k) \leqslant 2^{-k}$ e aplicando a Corolário 9. A recíproca se prova por contraposição: se $X_n \not\to X$ em probabilidade, existem $\varepsilon > 0$, $\delta > 0$ e uma subsequência com $\mathbb{P}(|X_{n_k} - X| \geqslant \varepsilon) \geqslant \delta$ para todo k, e nenhuma subsubsequência dessa subsequência pode convergir quase certamente, já que a convergência quase certa implica a convergência em probabilidade pela Proposição 15. O exemplo da Exemplo 16 mostra que a subsequência não pode ser substituída pela sequência inteira.

5. Convergência em Média e em Média Quadrática

5.1. As Definições

As noções desta seção substituem a pergunta "o erro é pequeno?" pela pergunta "o erro médio é pequeno?". Elas têm a vantagem de se reduzirem a cálculos de esperança, que já sabemos fazer, e a desvantagem de serem sensíveis a poucos pontos onde o erro é grande, como veremos nos exemplos.

Definição 18 (Convergência em média)**.** Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias integráveis, isto é, com $\mathbb{E}\big[|X_n|\big] < \infty$ para todo $n \in \mathbb{N}$, e $\mathbb{E}\big[|X|\big] < \infty$. Dizemos que (X_n) *converge em média* para X se

$$\lim_{n\to\infty} \mathbb{E}\big[|X_n - X|\big] = 0.$$

Nesse caso, escrevemos $X_n \xrightarrow{L^1} X$.

Definição 19 (Convergência em média quadrática)**.** Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias com $\mathbb{E}[X_n^2] < \infty$ e $\mathbb{E}[X^2] < \infty$. Dizemos que (X_n) *converge em média quadrática* para X se

$$\lim_{n\to\infty} \mathbb{E}\big[(X_n - X)^2\big] = 0.$$

Nesse caso, escrevemos $X_n \xrightarrow{L^2} X$.

As definições pedem que os momentos envolvidos sejam finitos, pois do contrário as esperanças não estariam bem definidas. No caso da média quadrática, a desigualdade de Cauchy-Schwarz garante que $\mathbb{E}[|X_n - X|]$ também é finito, de modo que a convergência em média quadrática só faz sentido quando a convergência em média também faz.

Exemplo 20 (Convergência em média quadrática)**.** Seja Z_n uma variável aleatória com

$$\mathbb{P}(Z_n = 1) = \frac{1}{n}, \qquad \mathbb{P}(Z_n = 2) = 1 - \frac{1}{n}.$$

Então $Z_n \to 2$ em média quadrática, pois

$$\mathbb{E}\Big[(Z_n - 2)^2\Big] = (1-2)^2 \cdot \frac{1}{n} + (2-2)^2 \cdot \left(1 - \frac{1}{n}\right) = \frac{1}{n} \longrightarrow 0.$$

O exemplo é o mesmo que aparece em Grimmett e Welsh e mostra que a convergência em média quadrática pode ser verificada por um cálculo direto, sem invocar nenhum teorema.

Exemplo 21 (A fronteira da média quadrática)**.** Seja $\alpha > 0$ e defina Z_n por

$$\mathbb{P}(Z_n = n^\alpha) = \frac{1}{n}, \qquad \mathbb{P}(Z_n = 0) = 1 - \frac{1}{n}.$$

Então

$$\mathbb{E}\Big[Z_n^2\Big] = n^{2\alpha} \cdot \frac{1}{n} = n^{2\alpha - 1} \qquad \text{e} \qquad \mathbb{E}\Big[|Z_n|\Big] = n^\alpha \cdot \frac{1}{n} = n^{\alpha - 1}.$$

Portanto, $Z_n \to 0$ em média quadrática se, e somente se, $\alpha < 1/2$, e em média se, e somente se, $\alpha < 1$. Por outro lado, para todo $\alpha > 0$ e todo $\varepsilon > 0$, quando $n^\alpha \geqslant \varepsilon$ temos

$$\mathbb{P}\Big(|Z_n| \geqslant \varepsilon\Big) = \mathbb{P}(Z_n = n^\alpha) = \frac{1}{n} \longrightarrow 0,$$

logo $Z_n \to 0$ em probabilidade para todo $\alpha > 0$. Assim, para $1/2 \leqslant \alpha < 1$, a sequência converge em probabilidade e em média, mas não em média quadrática. O exemplo será retomado na Seção 5.3.

5.2. Uma Hierarquia Parcial entre os Modos de Convergência

A proposição seguinte reúne as comparações entre as convergências em média, em média quadrática e em probabilidade. A parte relativa à distribuição será completada na Seção 6, depois de introduzirmos essa noção.

Proposição 22. Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias.

(a) Se $X_n \to X$ em média quadrática, então $X_n \to X$ em média.

(b) Se $X_n \to X$ em média quadrática, então $X_n \to X$ em probabilidade.

(c) Se $X_n \to X$ em média, então $X_n \to X$ em probabilidade.

Prova. Para (a), aplicamos a desigualdade de Cauchy-Schwarz à variável $Z = X_n - X$:

$$\mathbb{E}|Z| \leqslant \Big(\mathbb{E}[Z^2]\Big)^{1/2}.$$

Se $\mathbb{E}[(X_n - X)^2] \to 0$, então $\mathbb{E}[|X_n - X|] \to 0$.

Para (b), o resultado decorre da combinação de (a) e (c), mas também pode ser obtido diretamente aplicando a desigualdade de Chebyshev à mesma variável, o que fornece uma cota em termos de ε^{-2}. Para $\varepsilon > 0$,

$$\mathbb{P}\Big(|X_n - X| \geqslant \varepsilon\Big) \leqslant \frac{1}{\varepsilon^2}\mathbb{E}\Big[(X_n - X)^2\Big].$$

Se o lado direito tende a zero, o lado esquerdo também tende a zero, e a convergência em probabilidade segue.

Para (c), aplicamos a desigualdade de Markov a $|X_n - X|$:

$$\mathbb{P}\Big(|X_n - X| \geqslant \varepsilon\Big) \leqslant \frac{1}{\varepsilon}\mathbb{E}\Big[|X_n - X|\Big],$$

e o argumento é o mesmo de (b). ■

A proposição mostra a cadeia

$$L^2 \implies L^1 \implies \mathbb{P}.$$

A primeira implicação é estrita, como mostra o Exemplo 21 com $1/2 \leqslant \alpha < 1$; a segunda também é estrita, como mostra o exemplo a seguir.

Observação 23 (Convergência em L^p)**.** Poderíamos definir a convergência em média de ordem $p \geqslant 1$ exigindo $\mathbb{E}|X_n - X|^p \to 0$. A desigualdade de Lyapunov, demonstrada nas notas de aula *Desigualdades de Markov e Chebyshev e o lema de Borel-Cantelli*, diz que

$$\Big(\mathbb{E}|Z|^q\Big)^{1/q} \leqslant \Big(\mathbb{E}|Z|^p\Big)^{1/p} \qquad \text{para } 1 \leqslant q \leqslant p,$$

de modo que a convergência em L^p implica a convergência em L^q para todo $q \leqslant p$. A implicação (a) da Proposição 22 é o caso $p = 2$, $q = 1$. Nestas notas, tratamos apenas $p = 1$ e $p = 2$, que são os casos usados no curso; a mesma proposição vale, com a mesma prova, para qualquer $p \geqslant 1$.

Exemplo 24 (Convergência em probabilidade sem convergência em média)**.** Seja Z_n com $\mathbb{P}(Z_n = n) = 1/n$ e $\mathbb{P}(Z_n = 0) = 1 - 1/n$. Então $Z_n \to 0$ em probabilidade. De fato, dado $\varepsilon \in (0, 1]$ temos, para todo $n \in \mathbb{N}$, que $\mathbb{P}(|Z_n| \geqslant \varepsilon) = 1/n \to 0$. Além disso,

$$\mathbb{E}\Big[|Z_n|\Big] = \frac{1}{n} \cdot n = 1 \longrightarrow 1,$$

logo Z_n não converge para 0 em média. Tampouco converge em média quadrática, pois $\mathbb{E}[Z_n^2] = n \to \infty$. O exemplo é o caso $\alpha = 1$ do Exemplo 21: a massa $1/n$ tende a zero, mas seu produto pelo valor n mantém a esperança igual a 1.

5.3. O que a Convergência em Média Ignora

A convergência quase certa e a convergência em média, em geral, não são comparáveis. Já sabemos que a convergência em média implica convergência em probabilidade e, portanto, nada impede que ela coexista com a convergência quase certa. Mas a recíproca falha, e o exemplo a seguir é o mais simples possível.

Exemplo 25 (Convergência quase certa sem convergência em média)**.** Seja $\Omega = (0,1]$ com a medida de Lebesgue e defina

$$X_n = n\,\mathbb{1}_{(0,1/n)}.$$

Para cada $\omega \in (0,1]$, existe n_0 tal que $1/n < \omega$ para todo $n \geqslant n_0$; para esses índices, $\omega \notin (0,1/n)$ e $X_n(\omega) = 0$. Logo $X_n(\omega) \to 0$ para todo ω, e a convergência é quase certa. Por outro lado,

$$\mathbb{E}|X_n| = n \cdot \frac{1}{n} = 1 \qquad \text{para todo } n,$$

de modo que X_n não converge para 0 em média. Também não converge em média quadrática, pois $\mathbb{E}[X_n^2] = n \to \infty$. O exemplo mostra que a convergência quase certa não controla a magnitude do erro nos pontos onde ele é grande: em cada etapa, o erro vale n em um conjunto de medida $1/n$.

O exemplo anterior é a razão pela qual o teorema da convergência dominada, que enunciaremos na Seção 7, precisa de uma hipótese adicional. A convergência quase certa garante que o erro tende a zero em cada ponto, mas não garante que o erro médio tende a zero, porque a massa de pontos com erro grande pode se concentrar em conjuntos cada vez menores sem que a integral diminua.

Exemplo 26 (Convergência em média sem convergência quase certa)**.** A sequência da Exemplo 16 converge em probabilidade e também em média, pois

$$\mathbb{E}|X_n| = \mathbb{P}(X_n = 1) = \frac{1}{k} \longrightarrow 0$$

ao longo dos blocos. Como X_n só assume os valores 0 e 1, também $\mathbb{E}[X_n^2] = 1/k \longrightarrow 0$. No entanto, ela não converge em nenhum ponto. Logo nem a convergência em média nem a convergência em média quadrática implica convergência quase certa.

Reunindo os exemplos, obtemos a Tabela 1.

Sequência	Converge em	Não converge em
Z_n do Exemplo 21, $\alpha = 0{,}7$	$\mathbb{P}$, L^1	L^2
$n\mathbb{1}_{(0,1/n)}$	q.c., $\mathbb{P}$	L^1, L^2
X_n do Exemplo 16	$\mathbb{P}$, L^1, L^2	q.c.

Tabela 1: Comparação entre os modos de convergência.

A tabela não esgota as possibilidades, mas cobre as separações que aparecem com mais frequência. Note que nenhuma das linhas exibe convergência em média quadrática sem convergência em média ou em probabilidade, o que é coerente com a Proposição 22.

5.4. Um Cálculo que Será Reutilizado

Terminamos a seção com o cálculo que dá origem à lei fraca dos grandes números. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes com média μ e variância $\sigma^2 < \infty$, e seja $S_n = X_1 + \cdots + X_n$. Como a variância de uma soma de variáveis independentes é a soma das variâncias,

$$\operatorname{Var}(S_n) = n\sigma^2.$$

Logo

$$\mathbb{E}\left[\left(\frac{S_n}{n}-\mu\right)^2\right] = \operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\operatorname{Var}(S_n) = \frac{\sigma^2}{n} \longrightarrow 0.$$

Segue da definição e da identidade acima que $S_n/n \to \mu$ em média quadrática. Além do mais, sabemos da Proposição 22 que a convergência ocorre também em média e em probabilidade. A última afirmação é exatamente o que afirma a Lei Fraca dos Grandes Números, sob a hipótese de variância finita.

A Lei Forte dos Grandes Números afirma que $S_n/n \to \mu$ também quase certamente. Para obter essa conclusão pelo critério de Borel-Cantelli, precisaríamos mostrar que, para cada $\varepsilon > 0$, a soma das probabilidades de desvio é finita. Entretanto, uma das nossas melhores ferramentas, a desigualdade de Chebyshev, fornece apenas a seguinte estimativa

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

Quando $\sigma^2 > 0$, a soma das cotas à direita diverge, pois é um múltiplo da série harmônica. Portanto, essa estimativa não permite aplicar diretamente a Proposição 8; a demonstração da lei forte exige um argumento adicional.

A Lei Forte vai além do caso de variância finita. Para variáveis independentes e identicamente distribuídas, basta supor

$$\mathbb{E}[|X_1|] < \infty$$

para concluir que $S_n/n \to \mathbb{E}[X_1]$ quase certamente. Essa conclusão permanece válida mesmo quando $\mathbb{E}[X_1^2] = \infty$, situação em que o cálculo de variâncias acima não pode ser usado. Assim, a Lei Forte garante a convergência quase certa sob uma hipótese de momentos mais fraca do que a empregada nesta demonstração da Lei Fraca.

A Lei Fraca também vale sob essa hipótese de primeiro momento absoluto finito, para variáveis independentes e identicamente distribuídas. A variância finita foi usada aqui para dar uma prova curta e obter também a convergência em média quadrática. A demonstração da Lei Forte sob a hipótese de primeiro momento finito, que estudaremos em notas de aula futuras, exige outras ferramentas, entre elas um truncamento apropriado.

6. Convergência em Distribuição

6.1. A Definição e os Pontos de Continuidade

As noções anteriores comparam os valores de X_n com os de X e, por isso, exigem que as variáveis vivam em um mesmo espaço de probabilidade. A convergência em distribuição abandona essa comparação e olha apenas para as funções de distribuição. Ela é a noção que aparece no Teorema do Limite Central, quando se diz que uma soma adequadamente centralizada e normalizada tem distribuição aproximadamente normal sem que se especifique uma variável limite no mesmo espaço.

Definição 27 (Convergência em distribuição)**.** Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias com funções de distribuição $F_n(x) = \mathbb{P}(X_n \leqslant x)$ e $F(x) = \mathbb{P}(X \leqslant x)$. Dizemos que (X_n) *converge em distribuição* para X, e escrevemos $X_n \Rightarrow X$, se

$$\lim_{n\to\infty} F_n(x) = F(x)$$

para todo $x \in \mathbb{R}$ em que F é contínua.

A restrição aos pontos de continuidade de F não é um detalhe técnico dispensável. Ela é o que torna a definição estável e o que permite enunciar os teoremas de convergência sem hipóteses artificiais sobre as variáveis. O exemplo seguinte mostra o que aconteceria se exigíssemos convergência em todos os pontos.

Exemplo 28 (Por que os pontos de continuidade aparecem)**.** Seja $X \equiv 0$ e $X_n = 1/n$. A sequência converge para X em todos os sentidos estudados até aqui: quase certamente, em probabilidade, em média, em média quadrática. As funções de distribuição são $F(x) = \mathbb{1}_{[0,\infty)}(x)$ e $F_n(x) = \mathbb{1}_{[1/n,\infty)}(x)$. Para $x < 0$ e para $x > 0$ com n grande, $F_n(x) = F(x)$; mas em $x = 0$ temos $F_n(0) = 0$ para todo n, enquanto $F(0) = 1$. A convergência pontual das funções de distribuição produz, no limite, a função $\mathbb{1}_{(0,\infty)}$, que não é contínua à direita em 0 e não é a função de distribuição de X. A exigência de convergência apenas nos pontos de continuidade de F elimina o problema, porque nesses pontos F não tem salto e o valor pontual é determinado pelos valores vizinhos. Note que 0 é justamente o ponto de descontinuidade de F.

O exemplo anterior também explica por que a convergência em distribuição é uma propriedade da *distribuição* e não das variáveis. Se $X_n \Rightarrow X$ e Y_n tem a mesma distribuição que X_n para todo n, então $Y_n \Rightarrow X$. A relação entre X_n e X não desempenha nenhum papel, apenas as funções de distribuição. Em particular, não é necessário que as variáveis estejam definidas no mesmo espaço de probabilidade.

6.2. Exemplos

Exemplo 29 (Discretização da uniforme)**.** Para cada $n \geqslant 1$, seja X_n uniformemente distribuída no conjunto

$$\left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}\right\}.$$

Isso significa que cada um desses n pontos tem probabilidade $1/n$. Seja U uniforme em $(0, 1)$. Vamos mostrar que $X_n \Rightarrow U$ calculando as funções de distribuição.

Para $x \in [0, 1)$, calcular $F_n(x) = \mathbb{P}(X_n \leqslant x)$ consiste em contar quantos pontos do conjunto acima são menores ou iguais a x e somar suas probabilidades. A condição $k/n \leqslant x$ equivale a $k \leqslant nx$, de modo que os índices possíveis são $0, 1, \ldots, \lfloor nx \rfloor$, onde $\lfloor nx \rfloor$ denota o maior inteiro menor ou igual a nx. Como a contagem começa em 0, há $\lfloor nx \rfloor + 1$ pontos. Portanto,

$$F_n(x) = \frac{\lfloor nx \rfloor + 1}{n}.$$

Para comparar $F_n(x)$ com x, escrevemos

$$\begin{aligned} F_n(x) - x &= \frac{\lfloor nx \rfloor + 1}{n} - x \\ &= \frac{\lfloor nx \rfloor + 1 - nx}{n}. \end{aligned}$$

Pela definição de parte inteira, temos $\lfloor nx \rfloor \leqslant nx < \lfloor nx \rfloor + 1$. A desigualdade $nx < \lfloor nx \rfloor + 1$ mostra que o numerador é positivo. Por outro lado, como $\lfloor nx \rfloor \leqslant nx$, esse numerador é no máximo 1. Portanto,

$$0 < \lfloor nx \rfloor + 1 - nx \leqslant 1.$$

Dividindo por $n > 0$ e usando a expressão de $F_n(x) - x$ acima, obtemos

$$0 < F_n(x) - x \leqslant \frac{1}{n}.$$

Como $1/n \to 0$, o teorema do confronto garante que $F_n(x) - x \to 0$, isto é, $F_n(x) \to x = F_U(x)$ para todo $x \in [0, 1)$.

Para $x < 0$, temos $F_n(x) = F_U(x) = 0$. Por último, para $x \geqslant 1$, temos $F_n(x) = F_U(x) = 1$, em ambos os casos para todo $n \in \mathbb{N}$. Logo $F_n(x) \to F_U(x)$ para todo $x \in \mathbb{R}$. Como F_U é contínua em toda a reta, concluímos que $X_n \Rightarrow U$.

Assim, embora cada X_n assuma apenas um número finito de valores, suas distribuições convergem para uma distribuição contínua.

Exemplo 30 (Alternância de sinais)**.** Vamos construir uma sequência que converge em distribuição para U, mas não converge em probabilidade para essa mesma variável. A ideia é escolher U de modo que trocar seu sinal preserve a distribuição e, em seguida, alternar entre U e $-U$.

Seja U uma variável aleatória que assume os valores -1 e 1, cada um com probabilidade $1/2$. Defina

$$Z_n = \begin{cases} U, & n \text{ ímpar}, \\ -U, & n \text{ par}. \end{cases}$$

Quando n é ímpar, $Z_n = U$ e as distribuições são iguais. Quando n é par, $Z_n = -U$: o valor -1 de U passa a ser 1, e o valor 1 passa a ser -1. Como os dois valores têm a mesma probabilidade, essa troca preserva a distribuição. Portanto, para todo n,

$$\mathbb{P}(Z_n = -1) = \mathbb{P}(Z_n = 1) = \frac{1}{2}.$$

Assim, Z_n e U têm a mesma função de distribuição: $F_{Z_n}(x) = F_U(x)$ para todo $x \in \mathbb{R}$ e todo n. Em particular, $Z_n \Rightarrow U$.

Para verificar a convergência em probabilidade, precisamos comparar os valores de Z_n e U no mesmo ponto ω. Nos índices ímpares, a diferença é zero. Nos índices pares, temos $Z_n - U = -2U$ e, como $|U| = 1$ quase certamente, o valor absoluto dessa diferença é 2. Logo, quase certamente,

$$|Z_n - U| = \begin{cases} 0, & n \text{ ímpar}, \\ 2, & n \text{ par}. \end{cases}$$

Tomando $\varepsilon = 1$ na definição de convergência em probabilidade, obtemos

$$\mathbb{P}\big(|Z_n - U| \geqslant 1\big) = \begin{cases} 0, & n \text{ ímpar}, \\ 1, & n \text{ par}. \end{cases}$$

Essas probabilidades alternam entre 0 e 1 e, portanto, não tendem a zero. Assim, Z_n não converge para U em probabilidade.

Neste exemplo, a distribuição é determinada pelas probabilidades dos valores -1 e 1. Essas probabilidades são as mesmas para Z_n e U, embora seus valores sejam opostos em todos os índices pares. A convergência em probabilidade exige também o controle da diferença entre as variáveis no mesmo espaço de probabilidade.

6.3. Convergência em Probabilidade Implica a Distribuição

O resultado seguinte é o análogo, para a convergência em distribuição, da Proposição 15. A prova usa um argumento de sanduíche que será reutilizado em outras situações.

Proposição 31. Se $X_n \to X$ em probabilidade, então $X_n \Rightarrow X$.

Prova. Denote por F_n e F as funções de distribuição de X_n e X, respectivamente. Fixe um ponto x em que F é contínua. Pela definição de convergência em distribuição, precisamos mostrar que

$$F_n(x) = \mathbb{P}(X_n \leqslant x) \xrightarrow{\quad n\to\infty \quad} \mathbb{P}(X \leqslant x) = F(x).$$

Observe que a hipótese fornece informações diretas sobre a diferença $|X_n - X|$. Vamos usar esse controle para comparar as probabilidades acima. Se $X(\omega)$ estiver muito próximo de x, uma pequena diferença entre $X_n(\omega)$ e $X(\omega)$ pode colocá-los em lados opostos de x. Por isso, vamos comparar $F_n(x)$ com $F(x-\varepsilon)$ e $F(x+\varepsilon)$. A tolerância ε permitirá obter as inclusões de eventos necessárias; ao final, usaremos a continuidade de F para aproximar esses dois valores de $F(x)$.

Fixe $\varepsilon > 0$ e denote por B_n o evento em que o erro é pelo menos ε:

$$B_n \equiv \{|X_n - X| \geqslant \varepsilon\}.$$

Fora de B_n, temos $|X_n - X| < \varepsilon$. Além disso, a hipótese de convergência em probabilidade garante que $\mathbb{P}(B_n) \to 0$ para esse ε fixado.

1. Uma cota inferior para $F_n(x)$. Suponha que $X(\omega) \leqslant x - \varepsilon$ e que $\omega \notin B_n$. Usando que um número real é menor ou igual ao seu valor absoluto, obtemos

$$\begin{aligned} X_n(\omega) = X(\omega) + \big(X_n(\omega) - X(\omega)\big) &\leqslant X(\omega) + |X_n(\omega) - X(\omega)| \\ &< X(\omega) + \varepsilon \\ &\leqslant (x-\varepsilon) + \varepsilon = x. \end{aligned}$$

Portanto, $\{X \leqslant x-\varepsilon\} \cap B_n^c \subseteq \{X_n \leqslant x\}$. Desta forma podemos concluir que

$$\begin{aligned} \{X \leqslant x-\varepsilon\} &= \big(\{X \leqslant x-\varepsilon\} \cap B_n^c\big) \cup \big(\{X \leqslant x-\varepsilon\} \cap B_n\big) \\ &\subseteq \{X_n \leqslant x\} \cup \big(\{X \leqslant x-\varepsilon\} \cap B_n\big) \subseteq \{X_n \leqslant x\} \cup B_n. \end{aligned}$$

A Figura 2 ilustra essa inclusão.

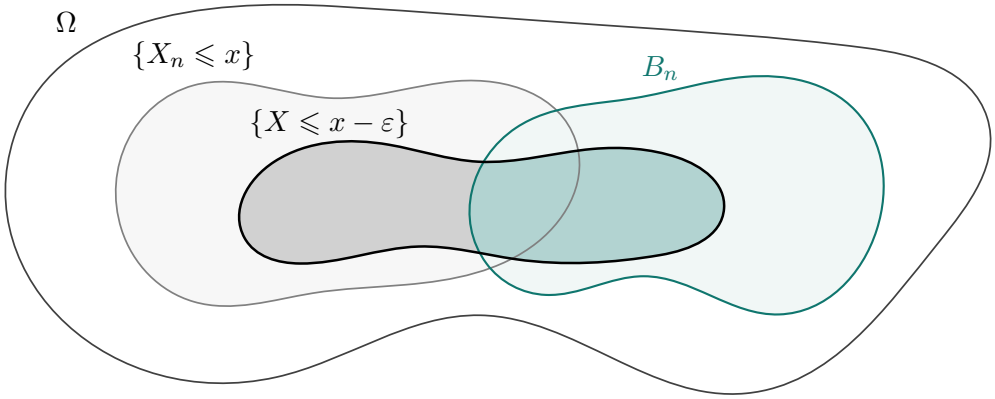


Figura 2: Os pontos de $\{X \leqslant x-\varepsilon\}$ fora de B_n (parte cinza) pertencem a $\{X_n \leqslant x\}$. A parte verde pertence a B_n; juntas, as duas partes ficam contidas em $\{X_n \leqslant x\} \cup B_n$.

Pela monotonicidade e pela subaditividade da probabilidade,

$$F(x-\varepsilon) \leqslant \mathbb{P}\Big(\{X_n \leqslant x\} \cup B_n\Big) \leqslant F_n(x) + \mathbb{P}(B_n).$$

Subtraindo $\mathbb{P}(B_n)$, obtemos a cota inferior

$$F(x-\varepsilon) - \mathbb{P}(B_n) \leqslant F_n(x).$$

2. Uma cota superior para $F_n(x)$. Agora suponha que $X_n(\omega) \leqslant x$ e que $\omega \notin B_n$. Queremos limitar o valor de $X(\omega)$. Para isso, escrevemos

$$\begin{aligned} X(\omega) &= X_n(\omega) + \Big(X(\omega) - X_n(\omega)\Big) \\ &\leqslant X_n(\omega) + |X_n(\omega) - X(\omega)| \\ &< X_n(\omega) + \varepsilon \\ &\leqslant x + \varepsilon. \end{aligned}$$

Assim, todo ponto de $\{X_n \leqslant x\}$ que não pertence a B_n pertence a $\{X \leqslant x+\varepsilon\}$. Logo,

$$\{X_n \leqslant x\} \subseteq \{X \leqslant x+\varepsilon\} \cup B_n.$$

Usando novamente a monotonicidade e a subaditividade, obtemos

$$\begin{aligned} F_n(x) &\leqslant \mathbb{P}\Big(\{X \leqslant x+\varepsilon\} \cup B_n\Big) \\ &\leqslant F(x+\varepsilon) + \mathbb{P}(B_n). \end{aligned}$$

3. Conclusão pela continuidade de F. Reunindo as duas cotas, temos

$$F(x-\varepsilon) - \mathbb{P}(B_n) \leqslant F_n(x) \leqslant F(x+\varepsilon) + \mathbb{P}(B_n).$$

Essas cotas valem para qualquer $\varepsilon > 0$. Precisamos agora que $F(x-\varepsilon)$ e $F(x+\varepsilon)$ fiquem próximos de $F(x)$ e que $\mathbb{P}(B_n)$ seja pequena. Vamos escolher primeiro ε e depois um índice a partir do qual essa probabilidade seja pequena.

Seja $\eta > 0$ o erro que queremos permitir na aproximação de $F_n(x)$ por $F(x)$. Em cada cota, o erro vem da variação de F e da probabilidade de desvio. Vamos fazer cada uma dessas contribuições menor que $\eta/2$. Como F é contínua em x, podemos escolher $\varepsilon > 0$ tão pequeno que

$$|F(x-\varepsilon) - F(x)| < \frac{\eta}{2} \qquad \text{e} \qquad |F(x+\varepsilon) - F(x)| < \frac{\eta}{2}.$$

Fixado esse valor de ε, a convergência em probabilidade fornece $N \in \mathbb{N}$ tal que

$$\mathbb{P}(B_n) < \frac{\eta}{2} \qquad \text{para todo } n \geqslant N.$$

Substituindo essas estimativas nas cotas anteriores, para $n \geqslant N$ obtemos

$$\begin{aligned} F_n(x) &\geqslant F(x-\varepsilon) - \mathbb{P}(B_n) \\ &> F(x) - \frac{\eta}{2} - \frac{\eta}{2} = F(x) - \eta, \end{aligned}$$

$$\begin{aligned} F_n(x) &\leqslant F(x+\varepsilon) + \mathbb{P}(B_n) \\ &< F(x) + \frac{\eta}{2} + \frac{\eta}{2} = F(x) + \eta. \end{aligned}$$

Portanto, $|F_n(x) - F(x)| < \eta$ para todo $n \geqslant N$. Como $\eta > 0$ foi arbitrário, isso prova que $F_n(x) \to F(x)$. O ponto x foi escolhido arbitrariamente entre os pontos de continuidade de F, de modo que $X_n \Rightarrow X$. ■

Corolário 32 (A convergência em média quadrática implica as demais)**.** Se $X_n \to X$ em média quadrática, em média, em probabilidade ou quase certamente, então $X_n \Rightarrow X$.

Prova. A convergência em média quadrática implica a convergência em média e a convergência em probabilidade pela Proposição 22; a convergência em média implica a convergência em probabilidade pela mesma proposição; a convergência quase certa implica a convergência em probabilidade pela Proposição 15. Em todos os casos, a Proposição 31 fornece a convergência em distribuição. ■

A proposição completa a cadeia de implicações

$$L^2 \implies L^1 \implies \mathbb{P} \implies \text{distribuição},$$

que liga quatro das cinco noções estudadas. A Proposição 15 acrescenta a convergência quase certa no início da cadeia, mas apenas até a convergência em probabilidade:

$$\text{q.c.} \implies \mathbb{P} \implies \text{distribuição}.$$

6.4. Quando a Recíproca Vale: o Limite Constante

A recíproca da Proposição 31 é falsa, como mostra o Exemplo 30. Há, no entanto, um caso importante em que ela vale: se o limite é uma constante, então convergência em distribuição e convergência em probabilidade são equivalentes.

Proposição 33. Seja $c \in \mathbb{R}$. Então $X_n \Rightarrow c$ se, e somente se, $X_n \to c$ em probabilidade.

Prova. Se $X_n \to c$ em probabilidade, a Proposição 31 já garante que $X_n \Rightarrow c$.

Resta provar a recíproca. Suponha que $X_n \Rightarrow c$ e denote por F_n a função de distribuição de X_n. Fixe $\varepsilon > 0$. Queremos mostrar que

$$\mathbb{P}\big(|X_n - c| \geqslant \varepsilon\big) \longrightarrow 0.$$

A distribuição limite concentra toda a probabilidade em c. Vamos usar a convergência em distribuição para mostrar que a probabilidade de X_n estar em um pequeno intervalo em torno de c tende a 1. Escolheremos esse intervalo com extremidades $c - \varepsilon/2$ e $c + \varepsilon/2$, pois até esses pontos estão a uma distância menor que ε de c.

1. A função de distribuição do limite. A variável limite vale sempre c. Portanto, sua função de distribuição é

$$F(x) = \mathbb{P}(c \leqslant x) = \begin{cases} 0, & x < c, \\ 1, & x \geqslant c. \end{cases}$$

Essa função é contínua em todos os pontos diferentes de c. Em particular, $c - \varepsilon/2$ e $c + \varepsilon/2$ são pontos de continuidade de F. Pela hipótese $X_n \Rightarrow c$, obtemos

$$F_n\Big(c - \frac{\varepsilon}{2}\Big) \longrightarrow F\Big(c - \frac{\varepsilon}{2}\Big) = 0,$$

$$F_n\Big(c + \frac{\varepsilon}{2}\Big) \longrightarrow F\Big(c + \frac{\varepsilon}{2}\Big) = 1.$$

2. A probabilidade de X_n estar perto de c. Considere o evento

$$A_n \equiv \Big\{c - \frac{\varepsilon}{2} < X_n \leqslant c + \frac{\varepsilon}{2}\Big\}.$$

Vamos mostrar que $\mathbb{P}(A_n) \to 1$. Para calcular essa probabilidade por meio de F_n, separamos o evento $\{X_n \leqslant c + \varepsilon/2\}$ em duas partes. A primeira reúne os pontos em que $X_n \leqslant c - \varepsilon/2$. A segunda reúne os pontos em que $c - \varepsilon/2 < X_n \leqslant c + \varepsilon/2$, isto é, os pontos de A_n. Essas duas condições não podem ocorrer simultaneamente. Assim,

$$\left\{X_n \leqslant c + \frac{\varepsilon}{2}\right\} = \left\{X_n \leqslant c - \frac{\varepsilon}{2}\right\} \cup A_n,$$

onde a união é disjunta. A Figura 3 ilustra essa decomposição.

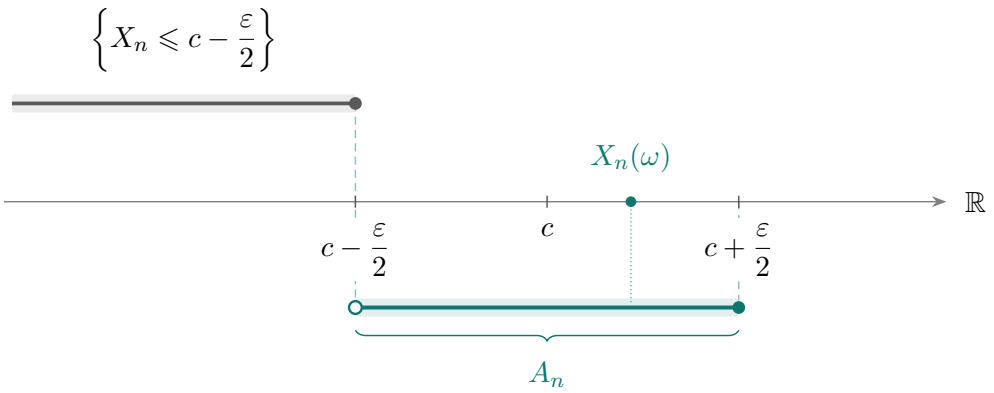


Figura 3: O extremo $c - \varepsilon/2$ é incluído na semirreta cinza e excluído do intervalo verde. Assim, as duas condições sobre X_n são disjuntas. O ponto marcado representa $X_n(\omega)$ para um $\omega \in A_n$.

Pela aditividade da probabilidade,

$$F_n\left(c + \frac{\varepsilon}{2}\right) = F_n\left(c - \frac{\varepsilon}{2}\right) + \mathbb{P}(A_n).$$

Subtraindo $F_n(c - \varepsilon/2)$ dos dois lados, obtemos

$$\mathbb{P}(A_n) = F_n\left(c + \frac{\varepsilon}{2}\right) - F_n\left(c - \frac{\varepsilon}{2}\right).$$

Pela primeira etapa, o primeiro termo dessa diferença tende a 1 e o segundo tende a 0. Portanto, $\mathbb{P}(A_n) \to 1$.

3. O controle da probabilidade de desvio. Se $\omega \in A_n$, então, subtraindo c nas desigualdades que definem esse evento, temos

$$-\frac{\varepsilon}{2} < X_n(\omega) - c \leqslant \frac{\varepsilon}{2}.$$

Portanto, $|X_n(\omega) - c| \leqslant \varepsilon/2 < \varepsilon$. Assim, todo ponto em que $|X_n - c| \geqslant \varepsilon$ está fora de A_n, isto é,

$$\{|X_n - c| \geqslant \varepsilon\} \subseteq A_n^c.$$

Pela monotonicidade da probabilidade e pela fórmula da probabilidade do complementar,

$$0 \leqslant \mathbb{P}\Big(|X_n - c| \geqslant \varepsilon\Big) \leqslant \mathbb{P}(A_n^c) = 1 - \mathbb{P}(A_n).$$

Como $\mathbb{P}(A_n) \to 1$, o lado direito tende a zero. Pelo teorema do confronto, concluímos que $\mathbb{P}(|X_n - c| \geqslant \varepsilon) \to 0$. Como $\varepsilon > 0$ foi arbitrário, isso prova que $X_n \to c$ em probabilidade. ■

Quando o limite é a constante c, toda a probabilidade da distribuição limite está concentrada nesse ponto. A proposição mostra que, se $X_n \Rightarrow c$, então, para cada $\varepsilon > 0$, a probabilidade de X_n ficar fora do intervalo $(c-\varepsilon, c+\varepsilon)$ tende a zero.

Em aplicações, X_n pode ser uma estimativa calculada a partir de uma amostra, e c a média de uma população ou a probabilidade de um evento. O valor de c pode ser desconhecido, mas é fixo. Se demonstrarmos $X_n \Rightarrow c$, a proposição garante que, para cada $\varepsilon > 0$, a probabilidade de o erro $|X_n - c|$ ser pelo menos ε tende a zero.

6.5. Limitações da Convergência em Distribuição

Vale resumir o que a convergência em distribuição controla e o que ela ignora. Ela controla a massa que cada variável coloca em intervalos da forma $(-\infty, x]$, ou seja, a forma das distribuições. Ela não controla a dependência entre X_n e X, como mostra o Exemplo 30; não controla os momentos, no sentido de que $X_n \Rightarrow X$ não implica $\mathbb{E}[X_n] \to \mathbb{E}[X]$; e não controla a probabilidade de eventos que dependem conjuntamente de X_n e X.

Exemplo 34 (Convergência em distribuição sem convergência de médias)**.** Seja $X \equiv 0$ e $X_n = n\mathbb{1}_{(0,1/n)}$, com $\Omega = (0,1]$ e medida de Lebesgue. Para todo $x < 0$, $F_n(x) = 0 = F(x)$. Para $x \geqslant 0$ e $n > x$, temos $F_n(x) = \mathbb{P}(X_n = 0) = 1 - 1/n$, que tende a $1 = F(x)$; em $x = 0$ o mesmo cálculo se aplica. Logo $X_n \Rightarrow 0$. No entanto, $\mathbb{E}[X_n] = 1$ para todo n, enquanto $\mathbb{E}[X] = 0$. A convergência em distribuição não vê a massa que escapa para valores grandes em conjuntos pequenos.

O exemplo anterior é o mesmo do Exemplo 25, lido sob outra ótica. A sequência converge em todos os sentidos que não envolvem a magnitude do erro e falha em todos os que a envolvem. A lição é que cada noção de convergência responde a uma pergunta diferente, e a escolha da noção deve ser feita de acordo com a pergunta.

7. O Quadro Geral das Implicações

7.1. O Diagrama

Reunimos na Figura 4 as implicações demonstradas até aqui. As setas contínuas valem sem hipóteses adicionais. A seta tracejada representa o teorema da convergência dominada, enunciado na Seção 7.3, e exige que os módulos $|X_n|$ sejam dominados por uma mesma variável integrável.

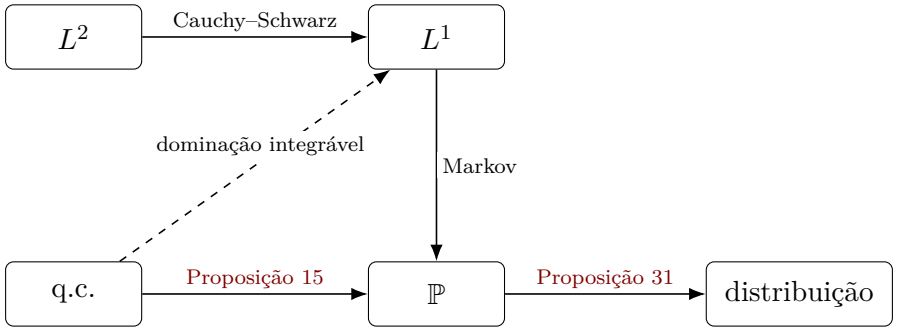


Figura 4: Implicações entre os modos de convergência.

O diagrama resume relações lógicas, e não um roteiro único de demonstração. A ausência de setas contínuas entre a convergência quase certa e as convergências em média registra que essas noções não são comparáveis em geral. As implicações para a convergência em distribuição também podem ser obtidas pela composição das setas que passam pela convergência em probabilidade.

7.2. A Tabela

A Tabela 2 reúne as definições, as implicações e, em cada linha, um exemplo que refuta uma das recíprocas.

Modo	Definição	Implica	Contraexemplo
q.c.	$\mathbb{P}(X_n \to X) = 1$	$\mathbb{P}$, distribuição	Exemplo 16: converge em $\mathbb{P}$, não em q.c.
em probabilidade	$\mathbb{P}(\lvert X_n - X\rvert \geqslant \varepsilon) \to 0$ para todo $\varepsilon > 0$	distribuição	Exemplo 30: converge em distribuição, não em $\mathbb{P}$
em média	$\mathbb{E}\lvert X_n - X\rvert \to 0$	$\mathbb{P}$, distribuição	Exemplo 24: converge em $\mathbb{P}$, não em média
em média quadrática	$\mathbb{E}[(X_n - X)^2] \to 0$	média, $\mathbb{P}$, distribuição	Exemplo 21 com $\alpha = 0{,}7$
em distribuição	$F_n(x) \to F(x)$ nos pontos de continuidade	nenhuma, em geral	Exemplo 30: converge em distribuição, não em $\mathbb{P}$

Tabela 2: Definições, implicações e recíprocas falsas.

A coluna "Implica" inclui consequências obtidas por composição. Por exemplo, a convergência quase certa implica a convergência em distribuição porque implica a convergência em probabilidade. A última coluna apresenta apenas algumas separações. Em particular, a convergência em distribuição não implica as demais em geral, mas é equivalente à convergência em probabilidade quando o limite é constante, pela Proposição 33.

7.3. Convergência Dominada

Falta tratar de uma situação que aparece com frequência nas aplicações. Muitas vezes sabemos que $X_n \to X$ quase certamente e queremos concluir que as médias também convergem. O Exemplo 25 mostra que isso não vale em geral: a convergência quase certa não controla a magnitude do erro. O teorema da convergência dominada dá a hipótese adicional que corrige o problema.

Teorema 35 (Convergência dominada)**.** Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias tais que $X_n \to X$ quase certamente. Suponha que exista uma variável aleatória Y com $\mathbb{E}[Y] < \infty$ e

$$\lvert X_n\rvert \leqslant Y \qquad \text{quase certamente, para todo } \; n \in \mathbb{N}.$$

Então X é integrável, $\mathbb{E}\big[\lvert X_n - X\rvert\big] \to 0$ e, em particular, $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

Prova. Como $X_n \to X$ quase certamente, $\lvert X_n - X\rvert \to 0$ quase certamente. Além disso,

$$\lvert X_n - X\rvert \leqslant \lvert X_n\rvert + \lvert X\rvert \leqslant 2Y \qquad \text{quase certamente,}$$

sendo a última desigualdade consequência de $|X| \leqslant Y$ quase certamente (que se obtém passando ao limite em $|X_n| \leqslant Y$). A variável aleatória $2Y$ é integrável e domina a sequência $|X_n - X|$. Pelo teorema da convergência dominada de Lebesgue (na forma para variáveis aleatórias, idêntica à enunciada nas notas de aula sobre funções de distribuição),

$$\lim_{n\to\infty} \mathbb{E}\Big[|X_n - X|\Big] = \mathbb{E}\left[\lim_{n\to\infty} |X_n - X|\right] = 0.$$

A conclusão sobre as esperanças segue de $\Big|\mathbb{E}[X_n] - \mathbb{E}[X]\Big| \leqslant \mathbb{E}\Big[|X_n - X|\Big]$. ■

Corolário 36. Sob as hipóteses do Teorema 35, a sequência $(X_n)_{n\in\mathbb{N}}$ converge em média, em probabilidade e em distribuição. Se, além disso, $\mathbb{E}[Y^2] < \infty$, então a convergência também vale em média quadrática.

Prova. Pelo Teorema 35, temos $X_n \to X$ em média. A Proposição 22(c) garante então que $X_n \to X$ em probabilidade. Por sua vez, a Proposição 31 permite concluir que $X_n \Rightarrow X$.

Suponha agora que $\mathbb{E}[Y^2] < \infty$. Para obter a convergência em média quadrática, precisamos mostrar que $\mathbb{E}\Big[(X_n - X)^2\Big] \to 0$. Vamos aplicar o teorema da convergência dominada à sequência dos quadrados dos erros. Como $X_n \to X$ quase certamente, também $(X_n - X)^2 \to 0$ quase certamente. Precisamos ainda de uma variável aleatória integrável que domine esses quadrados e não dependa de n.

A hipótese $|X_n| \leqslant Y$ fornece essa dominação. De fato, passando ao limite, obtemos também $|X| \leqslant Y$ quase certamente. Assim, para todo n, temos quase certamente

$$0 \leqslant (X_n - X)^2 = |X_n - X|^2 \leqslant \Big(|X_n| + |X|\Big)^2 \leqslant (2Y)^2 = 4Y^2.$$

Como $\mathbb{E}[4Y^2] = 4\mathbb{E}[Y^2] < \infty$, a variável aleatória $4Y^2$ é integrável. Podemos, portanto, aplicar o teorema da convergência dominada e concluir que

$$\lim_{n\to\infty} \mathbb{E}\Big[(X_n - X)^2\Big] = \mathbb{E}\left[\lim_{n\to\infty} (X_n - X)^2\right] = 0.$$

Isso prova que $X_n \to X$ em média quadrática. ■

Sob dominação integrável, a convergência quase certa implica a convergência em média e, portanto, as convergências em probabilidade e em distribuição. Se a dominante também tem quadrado integrável, obtemos ainda a convergência em média quadrática. A hipótese de dominação é essencial: sem ela, o Exemplo 25 mostra que a convergência quase certa pode coexistir com a divergência das médias.

Exemplo 37 (O máximo de uniformes converge em todos os modos)**.** Sejam $U_1, U_2, \ldots$ independentes e uniformes em $(0,1)$ e $M_n = \max\{U_1, \ldots, U_n\}$. No Exemplo 7, mostramos que $M_n \to 1$ quase certamente. Como $0 \leqslant M_n \leqslant 1$ e 1 é integrável, o Teorema 35 se aplica com $Y \equiv 1$. Logo $M_n \to 1$ em média, em probabilidade e em distribuição. Como $Y^2 = 1$ também é integrável, o Corolário 36 garante a convergência em média quadrática.

Podemos verificar o resultado por cálculo direto. Para $t \in [0,1]$,

$$\mathbb{P}(M_n \leqslant t) = t^n,$$

logo a densidade de M_n é nt^{n-1} em $(0,1)$ e

$$\mathbb{E}[M_n] = \int_0^1 t \cdot nt^{n-1}\, dt = \frac{n}{n+1} \longrightarrow 1.$$

Do mesmo modo,

$$\mathbb{E}[M_n^2] = \int_0^1 t^2 \cdot n t^{n-1}\, dt = \frac{n}{n+2} \longrightarrow 1,$$

e portanto

$$\mathbb{E}\Big[(1-M_n)^2\Big] = 1 - 2\,\mathbb{E}[M_n] + \mathbb{E}[M_n^2] = \frac{2}{(n+1)(n+2)} \longrightarrow 0.$$

O exemplo é instrutivo porque a convergência em todos os modos pode ser verificada de duas maneiras independentes: pelo teorema da convergência dominada, que não usa a forma da distribuição, e pelo cálculo explícito dos momentos, que não usa o teorema.

7.4. Considerações Finais

As cinco noções não formam uma hierarquia linear, mas formam uma rede com duas entradas fortes, a convergência quase certa e a convergência em média quadrática, e uma saída fraca, a convergência em distribuição. A convergência em probabilidade é o ponto de encontro: ela é consequência de ambas as entradas e é a hipótese que basta para a convergência em distribuição. A convergência em distribuição é a única que não exige um espaço comum e a única que não controla o erro.

Na prática, a escolha da noção é ditada pelas estimativas disponíveis. Se conseguimos mostrar que $\mathbb{P}(|X_n - X| \geqslant \varepsilon) \to 0$ para todo $\varepsilon > 0$, concluímos que há convergência em probabilidade. Se mostramos que $\mathbb{E}[(X_n - X)^2] \to 0$, obtemos a convergência em média quadrática. Quando as probabilidades dos desvios são somáveis em n para cada tolerância, a Proposição 8 fornece a convergência quase certa. Se o interesse está apenas na forma assintótica da distribuição, como no Teorema do Limite Central, a convergência em distribuição é a noção apropriada.

8. Exercícios Propostos

1. Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias.

(a) Mostre que $X_n \to X$ quase certamente se, e somente se, $X_n - X \to 0$ quase certamente.

(b) Mostre que, se $X_n \to X$ quase certamente, então $|X_n| \to |X|$ quase certamente.

(c) Mostre que, se $X_n \to X$ em probabilidade, então $|X_n| \to |X|$ em probabilidade.

2. Suponha que $X_n \to X$ quase certamente e $X_n \to Y$ quase certamente. Mostre que $X = Y$ quase certamente. A conclusão continua válida se a convergência for apenas em probabilidade? E se for apenas em distribuição?

3. Sejam $X_n \to X$ e $Y_n \to Y$ em probabilidade. Mostre que $X_n + Y_n \to X + Y$ em probabilidade. *Sugestão:* use a desigualdade $|X_n + Y_n - X - Y| \leqslant |X_n - X| + |Y_n - Y|$ e a subaditividade da probabilidade.

4. Seja X com distribuição normal padrão e defina $X_n = (-1)^n X$. Mostre que $X_n \Rightarrow X$, mas (X_n) não converge em probabilidade para X. Compare com o Exemplo 30.

5. Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias com $X_n \Rightarrow X$ e seja (c_n) uma sequência de números reais com $c_n \to 0$. Mostre que $X_n + c_n \Rightarrow X$.

6. Suponha que $X_n \to X$ em média quadrática. Mostre que $\mathbb{E}[X_n] \to \mathbb{E}[X]$ e $\mathrm{Var}(X_n) \to \mathrm{Var}(X)$. Dê um exemplo em que $\mathbb{E}[X_n] \to \mathbb{E}[X]$ e, no entanto, X_n não converge para X em média quadrática.

7. Seja (A_n) uma sequência de eventos com $\mathbb{P}(A_n) = 2^{-n}$. Mostre que

$$\sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega) < \infty \qquad \text{para quase todo } \omega.$$

Interprete o resultado em termos do número de eventos A_n que ocorrem.

8. Seja $\Omega = \{0, 1, \ldots, 9\}^{\mathbb{N}}$, com a medida produto determinada por $\mathbb{P}(X_n = k) = 1/10$, onde $X_n(\omega) = \omega_n$. Para k fixo, defina

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

(a) Calcule $\mathbb{E}[F_{n,k}]$ e $\mathrm{Var}(F_{n,k})$.

(b) Conclua que $F_{n,k} \to 1/10$ em probabilidade.

(c) Mostre que a cota obtida em (b) não é somável em n e explique por que o critério de Borel-Cantelli não se aplica diretamente. A convergência quase certa será demonstrada nas notas de aula sobre a lei forte dos grandes números.

∗ **9.** No conjunto das classes de equivalência de variáveis aleatórias, identificando X e Y quando $X = Y$ quase certamente, defina

$$d(X, Y) = \mathbb{E}\Big[\min\{1, |X - Y|\}\Big].$$

(a) Mostre que d está bem definida, que $d(X, Y) = d(Y, X)$ e que $d(X, Y) = 0$ se, e somente se, $X = Y$ quase certamente.

(b) Mostre a desigualdade triangular.

(c) Mostre que $d(X_n, X) \to 0$ se, e somente se, $X_n \to X$ em probabilidade.

∗ **10.** Suponha que $X_n \to X$ em probabilidade. Construa, indutivamente, uma subsequência (X_{n_k}) tal que

$$\mathbb{P}\Big(|X_{n_k} - X| \geqslant \tfrac{1}{k}\Big) \leqslant 2^{-k} \qquad \text{para todo } k \geqslant 1,$$

e conclua que $X_{n_k} \to X$ quase certamente. Use o resultado para provar a implicação da esquerda para a direita no Observação 17.

Referências

[1] P. 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.