Introdução aos Processos Estocásticos - Desigualdades e Convergência

Eduardo M. A. M. Mendes

DELT - UFMG
Programa de Pós-Graduação em Engenharia Elétrica
Universidade Federal de Minas Gerais
emmendes@cpdee.ufmg.br

Introdução

- A seguir serão descritas as várias desigualdades usadas no contexto de probabilidade e processos estocásticos.
- Serão mencionadas também as definições de convergência.

Desigualdade de Hölder

- Se p e q são número reais maiores do que 1 com $\frac{1}{p} + \frac{1}{q} = 1$ e se as variáveis aleatórias X, Y e $|X|^p$, $|Y|^q$ são integráveis, então

$$E(|XY|) \leq (E(|X|^p))^{\frac{1}{p}} (E(|Y|^q))^{\frac{1}{q}}$$

- Prova: Seja x um número positivo e considere a função $\phi(x) = \frac{x^p}{p} + \frac{x^{-q}}{q}$. Suponha que essa função tenha um mínimo em $x = 1$ com $\phi(1) = 1$.
Substituindo $x = \frac{b^{\frac{1}{q}}}{a^{\frac{1}{p}}} = b^{\frac{1}{q}} a^{-\frac{1}{p}}$ com $a, b > 0$, então

$$\begin{aligned} \phi(x) &= (ap)^{-1} b^{\frac{p}{q}} + (bq)^{-1} a^{\frac{q}{p}} \geq 1 \\ &= p^{-1} b^{\frac{(p+q)}{q}} + q^{-1} a^{\frac{(p+q)}{p}} \geq ab \end{aligned}$$

Desigualdade de Hölder (cont.)

Como $\frac{1}{p} + \frac{1}{q} = 1$ ou $\frac{(p+q)}{pq} = 1$, temos $\frac{(p+q)}{q} = p$ e $\frac{(p+q)}{p} = q$ e

$$ab \leq \frac{b^p}{p} + \frac{a^q}{q}$$

Substituindo

$$b = \frac{|X|}{(E(|X|^p))^{\frac{1}{p}}} \text{ e } a = \frac{|Y|}{(E(|Y|^q))^{\frac{1}{q}}}$$

na expressão logo acima, temos:

$$\begin{aligned} \frac{|X||Y|}{(E(|X|^p))^{\frac{1}{p}}(E(|Y|^q))^{\frac{1}{q}}} &\leq \frac{\left(\frac{|X|}{(E(|X|^p))^{\frac{1}{p}}}\right)^p}{p} + \frac{\left(\frac{|Y|}{(E(|Y|^q))^{\frac{1}{q}}}\right)^q}{q} \\ &\leq \frac{\frac{|X|^p}{E(|X|^p)}}{p} + \frac{\frac{|Y|^q}{E(|Y|^q)}}{q} \end{aligned}$$

Desigualdade de Hölder (cont.)

Tomando o valor esperado obtemos

$$\frac{E(XY)}{(E(|X|^p))^{\frac{1}{p}}(E(|Y|^q))^{\frac{1}{q}}} \leq \frac{1}{p} + \frac{1}{q} = 1$$

a desigualdade segue direto da relação acima.

Desigualdade de Schwartz

Fazendo $p = q = 2$ na desigualdade de Hölder, obtemos

$$E(|XY|) \leq \sqrt{E(|X|^2)E(|Y|^2)}$$

Desigualdade de Minkowski

- Se p é um número real com $p \geq 1$ e se as variáveis aleatórias X, Y e $|X|^p$, $|Y|^p$ são integráveis, então

$$(E(|X+Y|^p))^{\frac{1}{p}} \leq (E(|X|^p))^{\frac{1}{p}} + (E(|Y|^p))^{\frac{1}{p}}$$

- Prova: Se $p = 1$ o resultado segue da desigualdade referente ao triângulo.
- Para $p > 1$

$$\begin{aligned} E(|X+Y|^p) &= E(|X+Y||X+Y|^{p-1}) \\ &\leq E(|X||X+Y|^{p-1}) + E(|Y||X+Y|^{p-1}) \end{aligned}$$

Usando a desigualdade de Hölder e $\frac{1}{q} = 1 - \frac{1}{p}$, temos

$$E(|X+Y|^p) \leq (E(|X|^p)^{\frac{1}{p}} (E(|X+Y|)^p)^{1-\frac{1}{p}} + (E(|Y|^p)^{\frac{1}{p}} (E(|X+Y|)^p)^{1-\frac{1}{p}}$$

dividindo por $(E(|X+Y|)^p)^{1-\frac{1}{p}}$ resulta na desigualdade.

Desigualdade de Jensen

- Se ϕ é uma função convexa e contínua e se ambos X e $\phi(X)$ são integráveis, então
$$\phi(E(X)) \leq E(\phi(X))$$
- Prova: Se ϕ é uma função convexa, podemos construir uma corda $a_i + b_i x$ tal que
$$a_i + b_i x \leq \phi(x)$$
e
$$\sup_x (a_i + b_i x) = \phi(x)$$
logo
$$a_i + b_i X \leq \phi(X)$$
e
$$a_i + b_i E(X) \leq E(\phi(X))$$

Desigualdade de Jensen (cont.)

ou

$$sup(a_i + b_i E(X)) = \phi(E(X)) \leq E(\phi(X))$$

Objetivo: Achar um limite para a probabilidade a partir da variância

$$\boxed{P(|X - E(X)| > \gamma)}$$

O que pode ser dito se $p_X(x)$ não pode ser integrada ou não é conhecida?

- Podemos estimar $E(X)$ e $Var(X)$, então

$$P(|X - E(X)| > \gamma) \leq B$$

onde

$$B = \frac{Var(X)}{\gamma^2}$$

Desigualdade de Chebyshev (cont.)

Exemplo: $X \sim N(0,1), \quad E(X) = 0, \quad Var(X) = 1$. Supondo $\gamma = 3$, temos

$$P(|X - 0| > 3) \leq \frac{1}{3^2} = \frac{1}{9} \approx 0,11$$

Calculando

$$P(|X| > 3) = 2P(X > 3) \approx 0,0027 < 0,11$$

limite muito conservador!

Exemplo: PDF Laplaciano com $\sigma^2 = 1 = Var(X)$

$$P(|X - 0| > 3) \leq \frac{1}{9} \approx 0,11$$

Conclusão: Mesmo limite para todas as PDFs com $Var(X) = 1$.

Desigualdade de Chebyshev (cont.)

- Se $Var(X) \to 0$, então

$$P(|X - E(X)| > \gamma) \to 0$$

para qualquer $\gamma > 0$.

Prova:

$$\begin{aligned}
Var(X) &= \int_{-\infty}^{\infty} (x - E(X))^2 p_X(x)dx \\
&= \int_{\{x:|x-E(X)|>\gamma\}} (x - E(X))^2 p_X(x)dx \\
&\quad + \int_{\{x:|x-E(X)|\leq\gamma\}} (x - E(X))^2 p_X(x)dx \\
&\geq \int_{\{x:|x-E(X)|>\gamma\}} (x - E(X))^2 p_X(x)dx \\
&\geq \int_{\{x:|x-E(X)|>\gamma\}} \gamma^2 p_X(x)dx = \gamma^2 P(|x - E(X)| > \gamma)
\end{aligned}$$

Desigualdade de Chebyshev (cont.)

o que conclui a nossa prova.

Teorema de Markov e Desigualdade de Chebyshev

Teorema: Se X é uma variável aleatória e $g(X)$ é uma transformação de X tal que $g(X) > 0$, então, para qualquer $K > 0$

$$P(g(X) \geq K) \leq \frac{E(g(X))}{K}$$

Prova: Usando a função indicadora, ou seja

$$I(g(X)) = \begin{cases} 1 & \text{se } g(X) \geq K, \\ 0 & \text{caso contrário} \end{cases}$$

Se $g(X) \geq 0$ e $I(g(X)) \leq 1$, podemos escrever para a primeira condição

$$I(g(X)) \leq \frac{g(X)}{K}$$

Tomando o valor esperado

Teorema de Markov e Desigualdade de Chebyshev (cont.)

$$
\begin{aligned}
E(I(g(X))) &\leq \frac{E(g(X))}{K} \\
\sum_x I(g(X)) \times p(x) &\leq \\
[1 \times P(I(g(X)) = 1)] + [0 \times P(I(g(X)) = 0)] &\leq \\
[1 \times P(g(X) \geq K)] + [0 \times P(g(X) < K)] &\leq \\
P(g(X) \geq K) &\leq \frac{E(g(X))}{K}
\end{aligned}
$$

Teorema de Markov e Desigualdade de Chebyshev (cont.)

- Repare que se fizermos $g(X) = (X - \mu)^2$ e $K = k^2\sigma^2$, temos

$$P\left((X-\mu)^2 \geq k^2\sigma^2\right) \leq \frac{E\left((X-\mu)^2\right)}{k^2\sigma^2} = \frac{\sigma^2}{k^2\sigma^2}$$

Podemos usar $(X - \mu)^2 \geq k^2\sigma^2$ da seguinte maneira

$$\begin{aligned}
(X-\mu)^2 \geq k^2\sigma^2 \quad & \rightarrow \quad \left(X - \mu \geq \sqrt{k^2\sigma^2}\right) \text{ ou } \left(X - \mu \leq -\sqrt{k^2\sigma^2}\right) \\
& \rightarrow \quad \left(|X-\mu| \geq \sqrt{k^2\sigma^2}\right) \\
& \rightarrow \quad (|X-\mu| \geq k\sigma)
\end{aligned}$$

Logo

$$P\left(|X-\mu| \geq k\sigma\right) \leq \frac{1}{k^2}$$

que é a desigualdade de Chebyshev para $k = \frac{\gamma}{\sigma}$. Lembrando que σ^2 nada mais é do que $Var(\bullet)$.

Limites de Chernoff

- Fornece um limite menos conservador comparado com Markov e Chebyshev.
- É aplicado quando a variável aleatória é a soma de várias variáveis aleatórias mutuamente independentes.
- É o limite na probabilidade de se desviar da média.
- Não se aplica a todas as distribuições. Só aplicável a soma de VAs definidas no intervalo $[0, 1]$.
- Não depende diretamente do número de variáveis somadas.

Limites de Chernoff (cont.)

Teorema: Seja $X_1, X_2, \ldots, X_N$ variáveis aleatórias mutuamente independentes definidas no intervalo $[0,1]$. Seja $X = \sum_{j=1}^{N} X_j$, então para qualquer $c > 1$

$$P\left(X \geq cE(X)\right) \leq e^{-\alpha E(X)}$$

onde $\alpha = c ln(c) + 1 - c > 0$.

Prova: Definindo $R_j = X_j - E(X_j)$ para $j = 1, 2, \ldots, N$, temos

$$-E(X_j) \leq R_j \leq 1 - E(X_j)$$

e $E(R_j) = 0$. Logo

Limites de Chernoff (cont.)

$$\begin{aligned} P\left(X \geq cE(X)\right) &= P\left(R \geq (c-1)E(X)\right) \\ &= P\left(c^R \geq c^{(c-1)E(X)}\right) \\ &\leq \frac{E(c^R)}{c^{(c-1)E(X)}} \quad \text{Usando o Teorema de Markov} \end{aligned}$$

Sabemos

$$\begin{aligned} E(c^R) &= E\left(c^{R_1+R_2+\cdots+R_N}\right) \\ &= E\left(\prod_{j=1}^{N} c^{R_j}\right) \\ &= \prod_{j=1}^{N} E\left(c^{R_j}\right) \quad \text{independência} \end{aligned}$$

Limites de Chernoff (cont.)

Precisamos

1) Se $-m \leq z \leq 1 - m$, então

$$c^z \leq c^{-m}(1 + m(c-1)) + z(c^{1-m} - c^{-m})$$

Repare que o lado direito descreve uma equação de uma reta que corta a curva c^z em dois pontos: $-m$ e $1 - m$. Como c^z é convexa, neste intervalo, a curva se encontra abaixo da reta.

2) $1 + m(c-1) \leq e^{m(c-1)}$ que é consequência direta da expansão em séries de Taylor $(1 + x \leq e^x)$.

Limites de Chernoff (cont.)

Podemos continuar com a prova fazendo $m = E(X_j)$ e $z = R_j$

$$\begin{aligned} E\left(c^{R_j}\right) &\leq E\left(c^{-m}e^{m(c-1)} + (c^{1-m} - c^{-m})R_j\right) \\ &\leq c^{-m}e^{m(c-1)} \quad \text{pois } E(R_j) = 0 \\ &\leq e^{m(c-1-ln(c))} \end{aligned}$$

Assim

$$\begin{aligned} \prod E\left(c^{R_j}\right) &\leq \prod e^{E(X_j)(c-1-ln(c))} \\ &\leq e^{(c-1-ln(c))\sum E(X_j)} \\ &\leq e^{(c-1-ln(c))E(X)} \end{aligned}$$

Finalmente

Limites de Chernoff (cont.)

$$
\begin{aligned}
P\left(X \geq cE(X)\right) &\leq \frac{e^{(c-1-ln(c))E(X)}}{c^{(c-1)E(X)}} \\
&\leq e^{E(X)(c-1-ln(c)-cln(c)+ln(c))} \\
&\leq e^{E(X)(-cln(c)-1+c)} \\
&\leq e^{-\alpha E(X)}
\end{aligned}
$$

onde $\alpha = cln(c) + 1 - c$.

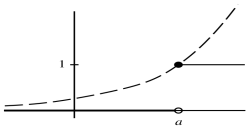

Figura 1: Mostra que $I_{[a,\infty)}(x)$ é um limite para $e^{s(x-a)}$.

Logo se tomarmos a esperança matemática de

$$I_{[a,\infty)}(X) \geq e^{s(X-a)}$$

temos

$$\begin{aligned} E\left(I_{[a,\infty)}(X)\right) &\geq E\left(e^{s(X-a)}\right) \\ &= e^{-sa}E\left(e^{sX}\right) \\ &= e^{-sa}\phi(s) \end{aligned}$$

onde $\phi(s)$ é função característica.
Usando o fato que toda probabilidade pode ser escrita como uma esperança matemática

$$P(X \geq A) = E\left(I_{[a,\infty)}X\right)$$

podemos escrever que

$$P(X \geq A) \geq e^{-sa}\phi(s)$$

que é válida para todo $s > 0$.
O limite de Chernoff é dado por

$$\boxed{P(X \geq A) \geq \min_{s \geq 0}\left(e^{-sa}\phi(s)\right)}$$

Exemplo

Seja X uma variável aleatória contínua com distribuição exponencial e $\lambda = 1$. Determine $P(X \geq 7)$ e os limites de Markov, Chebyshev and Chernoff.

Solução: Sabemos que $E(X) = \frac{1}{\lambda}$, $E(X^2) = \frac{2}{\lambda^2}$, $Var(X) = \frac{1}{\lambda^2}$ e $\phi(s) = \frac{\lambda}{(\lambda - s)}$. Logo:

Exemplo (cont.)

$$
\begin{aligned}
P(X \geq 7) &= \int_7^{\infty} e^{-x} dx = e^{-7} = 0,00091 \\
&\leq \frac{E(X)}{7} = \frac{1}{7} = 0,144 \quad \textcolor{red}{\text{Markov}} \\
&\leq \frac{Var(X)}{7^2} = \frac{1}{7^2} = 0,0204 \quad \textcolor{red}{\text{Chebyshev}} \\
&\leq \min_{s \geq 0} \frac{e^{-7s}}{1-s} = \left. \frac{e^{-7s}}{1-s} \right|_{s=6/7} = 7e^{-6} = 0.017 \quad \textcolor{red}{\text{Chernoff}}
\end{aligned}
$$

Estimação da Média e da Variância

Nas desigualdades propostas, como por exemplo a de Chebyshev, usa-se a variância. Logo seria interessante achar uma estimativa para ela quanto para o valor esperado.

- Média

$$\hat{E}(X) = \frac{1}{M}\sum_{i=1}^{M} x_i \quad \text{Média Amostral}$$

Em geral

$$\hat{E}(g(x)) = \frac{1}{M}\sum_{i=1}^{M} g(x_i)$$

- Para a variância, temos:

$$\hat{Var}(X) = \frac{1}{M}\sum_{i=1}^{M} x_i^2 - \left(\frac{1}{M}\sum_{i=1}^{M} x_i\right)^2$$

Convergência Pointwise

- Definição: Uma sequência $\{X_n\}$ converge para um limite X se e somente se, para qualquer $\epsilon > 0$ por menor que seja, podemos encontrar um inteiro n_o tal que

$$|X_n - X| < \epsilon$$

para todo $n > n_o$.

- Considere agora uma sequência de variáveis aleatórias

$$\{X_1, X_2, \ldots, X_n, \ldots\}$$

e a definição de convergência pointwise para outra variável X, então é necessário ter para todo ω-ponto em Ω a seguinte sequência

$$X_1(\omega), X_2(\omega), \ldots, X_n(\omega)$$

convergindo para $X(\omega)$. Este tipo de convergência é chamada everywhere convergence - altamente restritiva. Devemos olhar para sub-conjuntos de Ω.

Convergência Almost Sure

- Definição: Um sequência de variáveis aleatórias $\{X_n\}$ converge *almost sure* (a.s.) para X se para cada ω-ponto não pertencente ao evento nulo A,

$$\lim_{n\to\infty} |X_n(\omega) - X(\omega)| = 0$$

- Este tipo de convergência é também conhecida como convergência com probabilidade 1 e possui a seguinte notação

$$X_n(\omega) \xrightarrow[n\to\infty]{a.s.} X(\omega) \text{ ou } X(\omega) = \lim_{n\to\infty} X_n(\omega) \ (a.s.)$$

- Se o limite de X não é conhecido *a priori*, podemos definir a convergência *almost sure* mutual. A sequência X_n converge mutualmente *almost sure* se

$$\sup_{m\geq n} |X_m - X_n| \xrightarrow[n\to\infty]{a.s.} 0$$

Convergência Almost Sure (cont.)

- Uma maneira alternativa de se testar a convergência quase certamente é olhar para

$$P\left(\{\omega : X_n(\omega) \to X(\omega)\}\right) = 1 \quad \text{quando} \quad n \to \infty$$

ou

$$\lim_{n\to\infty} P\left(sup_{m\geq n}|X_m - X| > \epsilon\right) = 0 \quad \text{para todo} \quad \epsilon > 0$$

ou

$$\lim_{n\to\infty} P\left(sup_{m\geq n}|X_m - X| \leq \epsilon\right) = 1 \quad \text{para todo} \quad \epsilon > 0$$

- A convergência quase certamente põe restrições no comportamento conjunto de todos elementos aleatórios na sequência $|X_n - X|$, $|X_{n+1} - X|, \ldots$

Convergência em Probabilidade

- Definição: A sequência de variáveis aleatórias $\{X_n\}$ converge *em probabilidade* para X se para $\epsilon > 0$ pequeno

$$\lim_{n\to\infty} P\left(|X_n - X| \geq \epsilon\right) = 0$$

ou

$$\lim_{n\to\infty} P\left(|X_n - X| < \epsilon\right) = 1$$

- A notação usada é a seguinte

$$X_n(\omega) \xrightarrow[n\to\infty]{l.i.p} X(\omega) \text{ ou } X(\omega) = \underset{n\to\infty}{l.i.p}\, X_n(\omega)$$

- A convergência em probabilidade mutual é definida como

$$\lim_{n\to\infty} \sup_{m\geq n} P\left(|X_m - X_n| \geq \epsilon\right) \to 0$$

Convergência em Probabilidade (cont.)

- A convergência em probabilidae olha para a distribuição marginal de $|X_n - X|$ quando $n \to \infty$.
- O conceito de convergência em probabilidade será uma importante ferramenta quando estudarmos a convergência de estimadores e a lei Fraca dos Grandes Números.
- Algums resultados importantes:
  - Se a sequência de variáveis aleatórias $\{X_n\}$ converge *almost sure* para X, então converge em probabilidade para o mesmo limite. O contrário não é verdadeiro.
  - Se $\{X_n\}$ converge em probabilidade para X, então existe uma sub-sequência $\{X_{nk}\}$ de $\{X_n\}$ que converge *almost sure* para o mesmo limite.
  - $\{X_n\}$ converge em probabilidade se e somente se converge mutualmente em probabilidade.

Diferença entre a convergência quase certamente e em probabilidade

Primeira maneira de olhar

- **Convergência em probabilidade** - A chance de ter algum ponto fora dos limites vai para zero quando o $n \to \infty$.
- **Convergência quase certamente** - É mais forte! O número de pontos fora do limite é finito (countável?). Ou seja, à medida que n aumenta, o número de pontos fora dos limites se esgota $\to$ tudo funciona perfeitamente. Infelizmente não se sabe quando isso acontece.

Na prática não há diferença entre os dois modos de convergência.

Segunda maneira de olhar

Diferença entre a convergência quase certamente e em probabilidade (cont.)

- **Convergência em probabilidade** - A probabilidade da sequência de variáveis aleatórias alcançar o alvo decresce exponencialmente e aproxima de zero mas <u>não</u> é zero.
- **Convergência quase certamente** - Algo irá acontecer (o alvo será alacançado) mas não sabemos quando.

- Definição: Uma sequência de variáveis aleatórias $\{X_n\}$ converge na p-ésima média ($p > 0$) para X se

$$\lim_{n\to\infty} E(|X_n - X|^p) \to 0$$

que também por ser escrito como

$$X_n(\omega) \xrightarrow[n\to\infty]{l.i.p.m} X(\omega) \text{ ou } X(\omega) = \underset{n\to\infty}{l.i.p.m}\, X_n(\omega)$$

- Podemos também definir a convergência na p-ésima média mutual

$$\lim_{n\to\infty} \sup_{m\geq n} E(|X_m - X_n)^p) \to 0$$

Convergência na Média (cont.)

- Um conceito importante é a convergência na média quadrada no qual $p = 2$

$$\lim_{n\to\infty} E(|X_n - X|)^2 \to 0$$

ou de maneira equivalente, no caso da convergência em média quadrada mutual

$$\lim_{n\to\infty} \sup_{m\geq n} E(|X_m - X_n)^2) \to 0$$

- A notação usada é

$$X_n(\omega) \xrightarrow[n\to\infty]{l.i.q.m} X(\omega) \text{ ou } X(\omega) = \underset{n\to\infty}{l.i.q.m}\, X_n(\omega)$$

- Mesmo a convergência sendo mais restritiva (forte) do que a convergência em probabilidade, não implica ou não é implicada pela convergência *almost sure*.

Convergência na Média (cont.)

- Alguns resultados importantes são:
  - Se a sequência de variáveis aleatórias $\{X_n\}$ converge na p-ésima média, então converge em probabilidade para o mesmo limite.
  - $\{X_n|\}$ converge na p-ésima média se e somente se converge mutualmente na p-ésima média.

Em probabilidade mas não quase certamente

Exemplo: Seja $X \sim U[0,1]$ e os intervalos binários $I_1 = [0,1]$, $I_2 = \left[0, \frac{1}{2}\right]$, $I_3 = \left[\frac{1}{2}, 1\right]$, $I_4 = \left[0, \frac{1}{4}\right]$, ..., $I_7 = \left[\frac{3}{4}, 1\right]$, $I_8 = \left[0, \frac{1}{8}\right]$ e assim por diante tal que

$$I_{2^m+i} = \left[\frac{i}{2^m}, \frac{i+1}{2^m}\right]$$

para $m = 0, 1, 2, \ldots$ e $i = 0, 1, \ldots, 2^m - 1$. Logo, os 2^m intervalos de tamanho 2^{-m} cobrem o intervalo $[0,1]$.

- Seja $Y_n = 1$ se $X \in I_n$ e $Y_n = 0$, caso contrário. A sequência Y_1, Y_2, ... converge *em probabilidade* para $Y = 0$ pois, para todo $0 < \epsilon \leq 1$, $P(|Y_n - 0| \geq \epsilon) = P(Y_n = 1) = P(X \in I_n) =$ tamanho do intervalo $I_n \to 0$, quando $n \to \infty$.

Em probabilidade mas não quase certamente (cont.)

- Y_n não converge quase certamente para zero pois qualquer X está somente em um dos 2^n intervalos de tamanho 2^{-n}. Ou seja, para todo $\omega \in \Omega$, $Y(\omega)$ assume o valor 1 para um número infinito de n's e assume o valor 0 para um número infinito de n's. Logo, para cada $\omega \in \Omega$, $Y(\omega)$ não converge.

a) Considere a variável aleatória X_n tal que $P(X_n = 0) = 1 - \frac{1}{n}$ e $P(X_n = n) = \frac{1}{n}$. Neste exemplo pode-se ver que $X_n \xrightarrow{P} 0$ pois

$$\begin{aligned} P(|X_n| > \epsilon) &= P(X_n = n) \\ &= \frac{1}{n} \\ &\rightarrow 0 \end{aligned}$$

Entretanto

$$E(X_n) = 1$$

e consequentemente X_n não converge em média para zero.

Quase certamente e em p-média implica em probabilidade (cont.)

b) Suponha o modelo para X_n do exemplo anterior mas com ω colhido de uma distribuição uniforme em $[0, 1]$, ou seja

$$X_n(\omega) = \begin{cases} n, & \omega \in \left[0, \frac{1}{n}\right), \\ 0, & \omega \in \left[\frac{1}{n}, 1\right] \end{cases}$$

tal que $P(X_n = 0) = 1 - \frac{1}{n}$ e $P(X_n = n) = \frac{1}{n}$ como antes. Para mostrar convergência quase certamente vamos utilizar a seguinte expressão

$$\lim_{n\to\infty} P(sup_{m\geq n}|X_m - X_n| \leq \epsilon) = 1 \quad \text{para todo } \epsilon > 0$$

Quase certamente e em p-média implica em probabilidade (cont.)

Temos o seguinte com o limite $X = 0$

$$
\begin{aligned}
\lim_{n\to\infty} P(sup_{m\geq n}|X_m - X_n| \leq \epsilon) &= \lim_{n\to\infty} P(sup_{m\geq n}|X_m| \leq \epsilon) \\
&= \lim_{n\to\infty} P(|X_n| \leq \epsilon, |X_{n+1}| \leq \epsilon, \ldots) \\
&\geq \lim_{n\to\infty} P(X_n = 0, X_{n+1} = 0, \ldots) \\
&= \lim_{n\to\infty} P\left(\omega \geq \frac{1}{n}, \omega \geq \frac{1}{n+1}, \ldots\right) \\
&= \lim_{n\to\infty} P\left(\omega \geq \frac{1}{n}\right) \\
&= \lim_{n\to\infty} \left(1 - \frac{1}{n}\right) \\
&= 1
\end{aligned}
$$

Quase certamente e em p-média implica em probabilidade (cont.)

portanto converge quase certamente.

c) Suponhamos que os elementos de X_n sejam independentes. Para todos os valores positivos inteiros de n, temos

Quase certamente e em p-média implica em probabilidade (cont.)

$$
\begin{aligned}
P(X_n = 0, X_{n+1} = 0, \ldots) &= \prod_{m=n}^{\infty}\left(1 - \frac{1}{m}\right) \\
&= \lim_{N\to\infty}\prod_{m=n}^{N}\left(1 - \frac{1}{m}\right) \\
&= \lim_{N\to\infty}\prod_{m=n}^{N}\frac{m-1}{m} \\
&= \lim_{N\to\infty}\frac{n-1}{n}\frac{n}{n+1}\cdots\frac{N-2}{N-1}\frac{N-1}{N} \\
&= \lim_{N\to\infty}\frac{n-1}{N} \\
&= 0
\end{aligned}
$$

Quase certamente e em *p*-média implica em probabilidade (cont.)

e portanto, para todo $\epsilon > 0$

$$\lim_{n \to \infty} P(sup_{m \geq n} |X_m| \leq \epsilon) = 0$$

que viola a condição, ou seja, a convergência quase certamente de X_n para zero falha, assim como a convergência na média. Assim mesmo temos a convergência em probabilidade.

Convergência em Distribuição

Seja X_n uma sequência de variáveis aleatórias e seja X uma variável aleatória. Suponha que X_n tenha distribuição F_n e que X tenha distribuição F. X_n converge em distribuição para a variável X se

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

para cada valor de t onde F é contínua. Notação: $X_n \xrightarrow{d} X$.

Exemplo: Seja $X_n = 1 + \frac{1}{n}$ uma variável aleatória constante. Podemos facilmente ver que a distribuição de X_n é

$$F_n(t) = \begin{cases} 0, & \text{se } t < 1 + \frac{1}{n} \\ 1, & \text{se } t \geq 1 + \frac{1}{n} \end{cases}$$

Note que $\lim_n F_n(t) = \tilde{F}(t)$ onde

$$\tilde{F}(t) = \begin{cases} 0, & \text{se } t \leq 1 \\ 1, & \text{se } t > 1 \end{cases}$$

que não é uma distribuição, pois não é contínua a direita. Podemos, entretanto, definir

$$F(t) = \begin{cases} 0, & \text{se } t < 1 \\ 1, & \text{se } t \geq 1 \end{cases}$$

que é uma distribuição para a variável $X = 1$. Logo $X_n \xrightarrow{d} X$, mesmo que $F_n(1) \nrightarrow F(1)$.