

Probabilidade (PPGECD000000001)

Programa de Pós-Graduação em Estatística e Ciência de Dados (PGECD)

Sessão 9

Raydonal Ospina

Departamento de Estatística
Universidade Federal da Bahia
Salvador/BA

Sequência de Eventos

A definição de conceitos de convergência de variáveis aleatórias depende de manipulações de sequências de eventos. Seja $A_n \subseteq \Omega$, define-se:

$$\inf_{k \geq n} A_k = \cap_{k=n}^{\infty} A_k, \quad \sup_{k \geq n} A_k = \cup_{k=n}^{\infty} A_k$$

$$\liminf_n A_n = \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k$$

$$\limsup_n A_n = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k.$$

O limite de uma sequência de eventos é definido da seguinte maneira: se para alguma sequência (B_n) de eventos $\liminf_n B_n = \limsup_n B_n = B$, então B é chamado de limite de (B_n) e nós escrevemos $\lim_n B_n = B$ ou $B_n \rightarrow B$.

Exemplo 1

$$\liminf [0, \frac{n}{n+1}) = \limsup [0, \frac{n}{n+1}) = [0, 1)$$

Teorema 1

Seja (A_n) uma sequência de eventos de Ω .

- (a) $\omega \in \limsup A_n$ se, e somente se, $\omega \in A_k$ para um número infinito de índices k .
- (b) $\omega \in \liminf A_n$ se, e somente se, $\omega \notin A_k$ para um número finito de índices k .

Demonstração.

Para parte (a), note que $\omega \in \limsup A_n$, se, e somente se, para todo n , $\omega \in \bigcup_{k=n}^{\infty} A_k$, ou seja, se, e somente se, para todo n existe $n' \geq n$ tal que $\omega \in A_{n'}$. Como isto é válido para todo n , temos que isto é equivalente a existência de um número infinito de índices k tais que $\omega \in A_k$.

A prova da parte (b) é similar.



Propriedades do \liminf e do \limsup

A seguir descreveremos algumas propriedades do \liminf e \limsup de uma sequência de eventos.

① $\liminf A_n \subseteq \limsup A_n$

Este fato é uma simples consequência do Teorema 1, pois se $\omega \in \liminf A_n$, ω não pertence apenas a um número finito de eventos A_k 's, e consequentemente pertence a um número infinito deles. Logo, $\omega \in \limsup A_n$.

② $(\liminf A_n)^c = \limsup A_n^c$

Este fato decorre aplicando a Lei de De Morgan duas vezes:

$$(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k)^c = \cap_{n=1}^{\infty} (\cap_{k=n}^{\infty} A_k)^c = \cap_{n=1}^{\infty} (\cup_{k=n}^{\infty} A_k^c).$$

Sequências Monotônicas

Uma sequência de eventos (A_n) é monotônica não-decrescente (resp., não-crescente) se $A_1 \subseteq A_2 \subseteq \dots$ (resp, $A_1 \supseteq A_2 \supseteq \dots$). Denotaremos por $A_n \uparrow$ (resp., $A_n \downarrow$) uma sequência não-decrescente (resp. não-crescente) de eventos.

Teorema 2

Suponha que (A_n) é uma sequência monotônica de eventos. Então,

① Se $A_n \uparrow$, então $\lim_n A_n = \bigcup_{n=1}^{\infty} A_n$.

② Se $A_n \downarrow$, então $\lim_n A_n = \bigcap_{n=1}^{\infty} A_n$.

Consequentemente, como para qualquer sequência B_n , temos $\inf_{k \geq n} B_k \uparrow$ e $\sup_{k \geq n} B_k \downarrow$, segue que:

$$\liminf B_n = \lim_n (\inf_{k \geq n} B_k), \quad \limsup B_n = \lim_n (\sup_{k \geq n} B_k)$$

Demonstração.

Para provar (1), precisamos mostrar que $\liminf A_n = \limsup A_n = \bigcup_{n=1}^{\infty} A_n$. Como $A_j \subseteq A_{j+1}$, temos $\bigcap_{k \geq n} A_k = A_n$, e portanto,

$$\liminf A_n = \bigcup_{n=1}^{\infty} (\bigcap_{k \geq n} A_k) = \bigcup_{n=1}^{\infty} A_n.$$

Por outro lado, temos, $\limsup A_n = \bigcap_{n=1}^{\infty} (\bigcup_{k \geq n} A_k) \subseteq \bigcup_{k=1}^{\infty} A_k = \liminf A_n \subseteq \limsup A_n$. Logo, temos igualdade acima, ou seja,

$$\limsup A_n = \bigcup_{k=1}^{\infty} A_k.$$

A prova de (2) é similar. □

Exemplo 2

1. $\lim_n [0, 1 - \frac{1}{n}] = \cup_{n=1}^{\infty} [0, 1 - \frac{1}{n}] = [0, 1).$
2. $\lim_n [0, 1 + \frac{1}{n}] = \cap_{n=1}^{\infty} [0, 1 + \frac{1}{n}] = [0, 1].$
3. $\lim_n (\frac{n}{n+1}, \frac{n}{n-1}) = \cap_{n=1}^{\infty} (\frac{n}{n+1}, \frac{n}{n-1}) = \{1\}.$

Exemplo

Sejam A_n, A, B_n, B eventos em Ω . Mostre que:

1. se $\lim_n A_n = A$, então $\lim_n A_n^c = A^c$.

Solução: $\liminf A_n^c = (\limsup A_n)^c = A^c$ e
 $\limsup A_n^c = (\liminf A_n)^c = A^c$.

2. $\limsup(A_n \cup B_n) = \limsup A_n \cup \limsup B_n$.

Solução: Se $\omega \in \limsup(A_n \cup B_n)$, então $\omega \in (A_k \cup B_k)$ para infinitos índices k . Logo, temos que $\omega \in A_k$ para infinitos índices k , ou $\omega \in B_k$ para infinitos índices k . Portanto, temos $\omega \in \limsup A_n$ ou $\omega \in \limsup B_n$, ou seja, $\omega \in \limsup A_n \cup \limsup B_n$. Reciprocamente, se $\omega \in \limsup A_n \cup \limsup B_n$, então $\omega \in \limsup A_n$ ou $\omega \in \limsup B_n$. Logo, temos que $\omega \in A_k$ para infinitos índices k , ou $\omega \in B_k$ para infinitos índices k , ou seja, $\omega \in (A_k \cup B_k)$ para infinitos índices k . Portanto, $\omega \in \limsup(A_n \cup B_n)$.

3. Não é verdade que $\liminf(A_n \cup B_n) = \liminf A_n \cup \liminf B_n$.

Solução: Vamos construir um contra-exemplo: Suponha que $A \cap B = \emptyset$, $A_n = A \neq \emptyset$ e $B_n = B \neq \emptyset$ para n par; e $A_n = B$ e $B_n = A$ para n ímpar. Como $A_n \cup B_n = A \cup B$ para todo n , é fácil ver que $\liminf(A_n \cup B_n) = A \cup B$. Também é fácil ver que $\liminf A_n = \liminf B_n = A \cap B = \emptyset$, pois somente os ω 's em $A \cap B$ não ocorrem para um número finito de índices n tanto na sequência A_n quanto na sequência B_n . Então, $A \cup B = \liminf(A_n \cup B_n) \neq \emptyset = \liminf A_n \cup \liminf B_n$.

4. se $A_n \rightarrow A$ e $B_n \rightarrow B$, então $A_n \cup B_n \rightarrow A \cup B$ e $A_n \cap B_n \rightarrow A \cap B$.

Solução: Pela parte (2), temos que

$$\limsup A_n \cup B_n = \limsup A_n \cup \limsup B_n = A \cup B,$$

e pela propriedade (1) de \liminf e \limsup , temos

$$\liminf A_n \cup B_n \subseteq \limsup A_n \cup B_n = A \cup B.$$

Resta-nos provar que $A \cup B \subseteq \liminf A_n \cup B_n$. Suponha que $\omega \in A \cup B$, então $\omega \in \liminf A_n$ ou $\omega \in \liminf B_n$, ou seja, ω não pertence a um número finito de A_k 's, ou ω não pertence a um número finito de B_k 's. Logo, ω não pertence a um número finito de $A_k \cup B_k$'s. Portanto, $\omega \in \liminf A_n \cup B_n$. Então, $A_n \cup B_n \rightarrow A \cup B$.

Utilizando os itens anteriores e a Lei de De Morgan, temos:

$$\begin{aligned} A \cap B &= (A^c \cup B^c)^c = (\lim A_n^c \cup \lim B_n^c)^c = \\ &= (\lim A_n^c \cup B_n^c)^c = \lim(A_n^c \cup B_n^c)^c = \lim A_n \cap B_n. \end{aligned}$$

Lema 1

Sejam A_1, A_2, \dots eventos aleatórios em (Ω, \mathcal{A}, P) , ou seja, $A_n \in \mathcal{A}, \forall n$.

(a) Se $\sum_{n=1}^{\infty} P(A_n) < \infty$, então $P(A_n \text{ infinitas vezes}) = 0$.

(b) Se $\sum_{n=1}^{\infty} P(A_n) = \infty$ e os eventos A_n 's são independentes, então

$$P(A_n \text{ infinitas vezes}) = 1.$$

Observação: O item (b) não vale necessariamente sem independência. Por exemplo, seja $A_n = A, \forall n$, onde $0 < P(A) < 1$. Então, $\sum P(A_n) = \infty$ mas o evento $[A_n \text{ infinitas vezes}] = A$ e $P(A_n \text{ infinitas vezes}) = P(A) < 1$.

Demonstração

Para parte (a), se $\sum P(A_n) < \infty$, então

$$\sum_{k=j}^{\infty} P(A_k) \rightarrow 0,$$

quando $j \rightarrow \infty$. Mas

$$[A_n \text{ infinitas vezes}] \subseteq \cup_{k=j}^{\infty} A_k, \forall j,$$

logo

$$P(A_n \text{ infinitas vezes}) \leq P(\cup_{k=j}^{\infty} A_k) \leq \sum_{k=j}^{\infty} P(A_k) \rightarrow 0.$$

Para parte (b), basta provar que

$$P(\cup_{k=n}^{\infty} A_k) = 1, \forall n$$

(pois sendo $[A_n \text{ infinitas vezes}] = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k$ a intersecção de um número enumerável de eventos de probabilidade 1, é também de probabilidade 1). Para tanto, seja $B_n = \cup_{k=n}^{\infty} A_k$. Então B_n contém $\cup_{k=n}^{n+m} A_k$ para todo m , e

$$B_n^c \subseteq (\cup_{k=n}^{n+m} A_k)^c = \cap_{k=n}^{n+m} A_k^c.$$

Logo para todo m ,

$$\begin{aligned} 1 - P(B_n) &= P(B_n^c) \leq P(\cap_{k=n}^{n+m} A_k^c) \\ &= \prod_{k=n}^{n+m} P(A_k^c) = \prod_{k=n}^{n+m} (1 - P(A_k)). \end{aligned}$$

Como $1 - p \leq e^{-p}$ para $0 \leq p \leq 1$, temos

$$1 - P(B_n) \leq \prod_{k=n}^{n+m} e^{-P(A_k)} = \exp(-\sum_{k=n}^{n+m} P(A_k)) \rightarrow 0$$

quando $m \rightarrow \infty$, pois $\sum_{k=n}^{n+m} P(A_k) \rightarrow \infty$ quando $m \rightarrow \infty$. Logo $P(B_n) = 1, \forall n$.

Exemplo 3

Se sabemos que para uma dada coleção de eventos $\{A_k\}$, as suas probabilidades individuais satisfazem $P(A_k) \leq \frac{1}{k^2}$, então podemos concluir que infinitos desses eventos ocorrem com probabilidade zero ou, que apenas um número finito deles ocorrem com probabilidade 1. Podemos reescrever isso da seguinte forma: existe um instante aleatório N tal que, com probabilidade 1, nenhum dos A_k ocorrem para $k > N$. É importante ressaltar que nós podemos chegar a essa conclusão sem saber nada sobre as interações entre esses eventos como as que são expressas por probabilidades de pares de eventos $P(A_i \cap A_j)$. Contudo, se apenas sabemos que $P(A_k) > 1/k$, então não podemos concluir nada baseados no Lema de Borel-Cantelli. Se soubermos que os eventos são mutuamente independentes, então sabendo que $P(A_k) > 1/k$, podemos concluir que infinitos A_k ocorrem com probabilidade 1.

Exemplo 4

Considere uma sequência de variáveis aleatórias X_1, X_2, X_3, \dots . Podemos usar o Lema de Borel-Cantelli para determinar a probabilidade que $X_k > b_k$ infinitas vezes para qualquer sequência de números reais $\{b_k\}$. Note que $P(X_k > b_k) = 1 - F_{X_k}(b_k)$. Logo, se

$$\sum_{k=1}^{\infty} P(X_k > b_k) = \sum_{k=1}^{\infty} 1 - F_{X_k}(b_k) < \infty,$$

então, não importa qual a distribuição conjunta das variáveis aleatórias $\{X_k\}$, temos que o evento $\{X_k > b_k\}$ só ocorrerá para um número finito de índices k .

Por outro lado, se

$$\sum_{k=1}^{\infty} P(X_k > b_k) = \sum_{k=1}^{\infty} 1 - F_{X_k}(b_k) = \infty,$$

então precisaríamos de informação adicional sobre a distribuição conjunta das variáveis aleatórias $\{X_k\}$ para determinar se os eventos $\{X_k > b_k\}$ ocorrem um número finito ou infinito de vezes.

Exemplo 5

Considere uma moeda não necessariamente honesta com probabilidade de cara igual a p , onde $0 < p < 1$. Se esta moeda for jogada um número infinito de vezes de maneira independente, qual a probabilidade da sequência (*cara, cara, coroa, coroa*) aparecer um número infinito de vezes? Justifique sua resposta.

Solução: Seja X_i o resultado do i -ésimo lançamento da moeda. Defina o evento $A_i = \{X_i = \text{cara}, X_{i+1} = \text{cara}, X_{i+2} = \text{coroa}, X_{i+3} = \text{coroa}\}$, queremos calcular $P(A_i \text{ infinitas vezes})$. Note que para todo i , temos $P(A_i) = p^2(1-p)^2 > 0$. Não podemos aplicar diretamente o lema de Borel Cantelli, pois os eventos A_i 's não são independentes, visto que, por exemplo, ambos A_1 e A_2 dependem de X_2, X_3, X_4 . Considere a seguinte subsequência da sequência de eventos (A_i) tal que $B_i = A_{4i-3}$. Como os eventos B_i 's dependem de famílias disjuntas de variáveis aleatórias independentes, eles são independentes. Além disso temos que $P(B_i) = p^2(1-p)^2 > 0$. Logo, $\sum_i P(B_i) = \infty$. Portanto, Borel-Cantelli implica que $P(B_i \text{ infinitas vezes}) = 1$. Como (B_i) é uma subsequência de (A_i) , temos que

$$[B_i \text{ infinitas vezes}] \subseteq [A_i \text{ infinitas vezes}].$$

Portanto, $P(A_i \text{ infinitas vezes}) = 1$.

Convergência de Variáveis Aleatórias

Seguindo uma interpretação frequentista, probabilidade está relacionada com a frequência relativa de eventos no longo prazo. A matemática para estudar o longo prazo é a dos limites. Mas quando se trata de funções, existem vários tipos de limites (por exemplo, pontual, uniforme, em quase todo lugar). O mesmo ocorre quando consideramos limites de variáveis aleatórias definidas em um mesmo espaço de probabilidade (Ω, \mathcal{A}, P) , visto que variáveis aleatórias são funções reais cujo domínio é Ω .

Relembrando: Seja (Ω, \mathcal{A}) um espaço mensurável. Uma função $X : \Omega \rightarrow R$ é chamada de variável aleatória se para todo evento Boreliano B , $X^{-1}(B) \in \mathcal{A}$. Nós recordamos que um evento Boreliano é qualquer evento pertencente à σ -álgebra de Borel, onde a σ -álgebra de Borel é a menor σ -álgebra contendo intervalos da forma $(-\infty, x]$ para todo $x \in R$.

Definição 1 (Convergência Quase Certa)

A sequência de variáveis aleatórias Y_1, Y_2, \dots converge quase certamente (ou com probabilidade 1) para a variável aleatória Y se

$$P(\{w : \lim_{n \rightarrow \infty} Y_n(w) = Y(w)\}) = 1.$$

Notação: $Y_n \rightarrow Y$ cp1.

Então se uma sequência de variáveis aleatórias Y_1, Y_2, \dots converge quase certamente para Y não significa que para todo $w \in \Omega$, $Y_n(w) \rightarrow Y(w)$, apenas o que se sabe é que a probabilidade do evento $D = \{w : Y_n(w) \nrightarrow Y(w)\}$ é nula. D é chamado de conjunto de exceção.

Exemplo 6

Considere uma variável aleatória Z tal que $P(\{w : 0 \leq |Z(w)| < 1\}) = 1$. Seja $X_n(w) = Z^n(w)$, então $X_n(w) \rightarrow 0$ cp1; note que o conjunto de exceção é $D = \{w \in \Omega : |Z(w)| \geq 1\}$ e que $P(D) = 0$. \square

Teorema 3

Se $\sum_n P(|X_n - X| > \epsilon) < \infty, \forall \epsilon > 0$, então $X_n \rightarrow X$ cp1. Se $\{X_n\}$ for uma sequência de variáveis aleatórias independentes e $\sum_n P(|X_n - X| > \epsilon) = \infty$ para algum $\epsilon > 0$, então, com probabilidade 1, $X_n \nrightarrow X$.

Demonstração.

Se $\sum_n P(|X_n - X| > \epsilon) < \infty, \forall \epsilon > 0$, então pelo Lema de Borel-Cantelli,

$$P(|X_n - X| > \epsilon \text{ infinitas vezes}) = 0 = 1 - P(|X_n - X| > \epsilon \text{ finitas vezes}).$$

Como isto vale para todo $\epsilon > 0$, temos que $X_n \rightarrow X$ cp1.

Por outro lado, se $\{X_n\}$ for uma sequência de variáveis aleatórias independentes e $\sum_n P(|X_n - X| > \epsilon) = \infty$ para algum $\epsilon > 0$, então pelo Lema de Borel-Cantelli,

$$P(|X_n - X| > \epsilon \text{ infinitas vezes}) = 1.$$

Logo, com probabilidade 1, $X_n \nrightarrow X$.



Exemplo 7

Seja $\{X_n\}_{n \geq 2}$ uma sequência de variáveis aleatórias independentes com distribuição de probabilidade dada por:

$$P(X_n = 0) = 1 - \frac{1}{\log n} \text{ e } P(X_n = n) = \frac{1}{\log n}, \forall n \geq 2.$$

Mostre que $X_n \not\rightarrow 0$ cp1.

Solução: Para qualquer ϵ tal que $0 < \epsilon < 1$, temos que

$$P(|X_n| > \epsilon) = P(X_n = n) = \frac{1}{\log n}.$$

Logo, $\sum_n P(|X_n| > \epsilon) = \sum_n \frac{1}{\log n} = \infty$. Então, o Lema de Borel-Cantelli implica que $P(|X_n| > \epsilon \text{ infinitas vezes}) = 1$, portanto com probabilidade 1, $X_n \not\rightarrow 0$.

Exemplo 8

Considere $\{X_n : n \geq 1\}$ uma sequência de variáveis aleatórias i.i.d. com função de distribuição F . Suponha que $F(x) < 1$, para todo $x < \infty$. Defina $Y_n = \max(X_1, X_2, \dots, X_n)$. Vamos verificar que $Y_n \rightarrow \infty$ cp1.

Inicialmente, observe que para cada $\omega \in \Omega$, as variáveis Y_n formam uma sequência não-decrescente de números reais. Seja M um número real, temos

$$\begin{aligned}
& P(Y_n \leq M : n = 1, 2, \dots) \\
& \leq P(Y_n \leq M : n = 1, 2, \dots, k) = P(Y_k \leq M) \\
& = P(\max(X_1, X_2, \dots, X_k) \leq M) \\
& = P(X_1 \leq M, X_2 \leq M, \dots, X_k \leq M) \\
& = \prod_{n=1}^k P(X_n \leq M) = F^k(M), \forall k \geq 1.
\end{aligned}$$

Fazendo $k \rightarrow \infty$, temos que para todo M finito,

$$P(\lim_n Y_n \leq M) = P(Y_n \leq M : n = 1, 2, \dots) = 0;$$

pois $F^k(M)$ tende a zero, quando $k \rightarrow \infty$. Dessa forma, o conjunto dos $w \in \Omega$, em que $\lim_n Y_n(w)$ é finito, tem probabilidade zero e, portanto, $Y_n \rightarrow \infty$ cp1.

Tipos de Convergência

Definição 2 (Convergência na r-ésima Média)

A sequência de variáveis aleatórias Y_1, Y_2, \dots converge na r-ésima Média, onde $r > 0$, para a variável aleatória Y se

$$\lim_{n \rightarrow \infty} E|Y_n - Y|^r = 0.$$

Notação: $Y_n \rightarrow^r Y$.

Se $r = 2$ este tipo de convergência é frequentemente chamado de *convergência em média quadrática*.

Exemplo 9

Sejam Z, X_1, X_2, \dots variáveis aleatórias tais que

$$X_n = \frac{n}{n+1} Z,$$

então $X_n \rightarrow^2 Z$ se $EZ^2 < \infty$, mas não em caso contrário. \square

Exemplo 10

Considere a sequência de variáveis aleatórias definidas no Exemplo 7. Mostre que $X_n \not\rightarrow^r 0$, para todo $r > 0$.

Solução: Temos que

$$E|X_n|^r = n^r P(X_n = n) = \frac{n^r}{\log n} \rightarrow \infty.$$

Logo, $X_n \not\rightarrow^r 0$.

O próximo teorema afirma que se $X_n \rightarrow^r X$, então $X_n \rightarrow^s X$ para $s < r$.

Teorema 4

Se $X_n \rightarrow^r X$, então $X_n \rightarrow^s X$ para $0 < s < r$

Demonstração.

Como $\frac{r}{s} > 1$, a função $\phi(x) = x^{\frac{r}{s}}$ é convexa, logo a desigualdade de Jensen implica que

$$E(|X|^r) = E\phi(|X|^s) \geq \phi(E(|X|^s)) = (E(|X|^s))^{\frac{r}{s}}$$

Substituindo X por $X_n - X$, temos

$$E(|X_n - X|^r) \geq (E|X_n - X|^s)^{\frac{r}{s}}.$$

Portanto, se $\lim_n E|X_n - X|^r = 0$, então $\lim_n E|X_n - X|^s = 0$.



Definição 3 (Convergência em Probabilidade)

A sequência de variáveis aleatórias Y_1, Y_2, \dots converge em probabilidade para a variável aleatória Y se $\forall \epsilon > 0$

$$\lim_{n \rightarrow \infty} P(\{w : |Y_n(w) - Y(w)| > \epsilon\}) = 0.$$

Notação: $Y_n \xrightarrow{P} Y$.

A intuição por trás desta definição é que para n muito grande a probabilidade de que Y_n e Y sejam bem próximas é bastante alta.

Exemplo 11

Considere a sequência de variáveis aleatórias definidas no Exemplo 7. Mostre que $X_n \xrightarrow{P} 0$.

Solução: Temos que para $0 < \epsilon < 1$, $P(|X_n| > \epsilon) = P(X_n = n)$ e para $\epsilon \geq 1$, $P(|X_n| > \epsilon) \leq P(X_n = n)$. Como $P(X_n = n) = \frac{1}{\log n} \rightarrow 0$, temos que $\forall \epsilon > 0$, $\lim P(|X_n| > \epsilon) = 0$. Portanto, $X_n \xrightarrow{P} 0$.

Exemplo 12

Considere X, X_1, X_2, \dots onde as variáveis aleatórias têm distribuição normal conjunta, todas com média 0 e matriz de covariância parcialmente descrita por

$$\text{COV}(X, X) = \text{COV}(X_n, X_n) = 1, \text{ e}$$

$$\text{COV}(X, X_n) = 1 - \frac{1}{n}.$$

Seja $Y_n = X_n - X$, como Y_n é uma combinação linear de variáveis aleatórias com distribuição normal, ela também possui distribuição normal. Precisamos determinar então sua média e sua variância. Mas $EY = E(X_n - X) = EX_n - EX = 0$ e

$$\text{Var}Y = EY^2 = E(X_n - X)^2 = EX_n^2 - 2EX_nX + EX^2 = 1 - 2\left(1 - \frac{1}{n}\right) + 1 = \frac{2}{n}.$$

Portanto, $Y_n \sim \mathcal{N}(0, \frac{2}{n})$. Então,

$$\begin{aligned} P(|X_n - X| > \epsilon) &= P(|Y_n| > \epsilon) \\ &= 2P(Y_n > \epsilon) = 2 \int_{\epsilon}^{\infty} \frac{\sqrt{n}}{\sqrt{4\pi}} e^{-\frac{ny^2}{4}} dy = 2 \int_{\epsilon\sqrt{\frac{n}{2}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx. \end{aligned}$$

Logo, $\forall \epsilon > 0, \lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$, ou seja, $X_n \xrightarrow{P} X$. \square

O último tipo de convergência estocástico que mencionamos não é exatamente uma noção de convergência das variáveis aleatórias propriamente ditas, mas uma noção de convergência de suas respectivas funções de distribuição acumuladas.

Definição 4 (Convergência em Distribuição)

A sequência de variáveis aleatórias Y_1, Y_2, \dots , converge em distribuição para a variável aleatória Y (denotado por $Y_n \rightarrow^D Y$) se para todo ponto x de continuidade de F_Y

$$\lim_{n \rightarrow \infty} F_{Y_n}(x) = F_Y(x).$$

Exemplo 13

Seja $\{X_n : n \geq 1\}$ uma sequência de variáveis aleatórias independentes com distribuição Uniforme em $(0, b)$, $b > 0$. Defina $Y_n = \max(X_1, X_2, \dots, X_n)$ e $Y = b$. Vamos verificar que $Y_n \rightarrow^D Y$. Temos

$$F_{Y_n}(y) = P(\max(X_1, X_2, \dots, X_n) \leq y) = F_{X_1}^n(y) = \begin{cases} 0 & \text{se } y < 0, \\ (\frac{y}{b})^n & \text{se } 0 \leq y < b, \\ 1 & \text{se } y \geq b. \end{cases}$$

Fazendo n tender ao infinito, temos que

$$\lim_n F_{Y_n}(y) = \begin{cases} 0 & \text{se } y < b, \\ 1 & \text{se } y \geq b, \end{cases}$$

que corresponde à função de distribuição de Y e, portanto, $Y_n \rightarrow^D Y$.

Deve-se ficar atento que convergência em distribuição não implica nada em relação aos outros tipos de convergência. Uma sequência convergindo em distribuição para uma variável aleatória X também converge em distribuição para qualquer outra variável aleatória Y tal que $F_Y = F_X$. O próximo exemplo serve para ilustrar melhor este fato.

Exemplo 14

Se uma sequência de variáveis aleatórias Y_1, Y_2, \dots é independente e identicamente distribuída de acordo com F , então para todo n tem-se que $F_{Y_n} = F$, logo a sequência converge em distribuição para qualquer variável aleatória X tal que $F_X = F$. Claro, como a sequência é independente, os valores de termos sucessivos são independentes e não exibem nenhum comportamento usual de convergência. \square

O requisito de continuidade, mencionado na definição acima, se justifica para evitar algumas anomalias. Por exemplo, para $n \geq 1$ seja $X_n = \frac{1}{n}$ e $X = 0$, para todo Ω . Parece aceitável que deveríamos ter convergência de X_n para X , qualquer que fosse o modo de convergência. Observe que

$$F_n(x) = \begin{cases} 0 & \text{se } x < \frac{1}{n}, \\ 1 & \text{se } x \geq \frac{1}{n}, \end{cases} \quad e \quad F(x) = \begin{cases} 0 & \text{se } x < 0, \\ 1 & \text{se } x \geq 0. \end{cases}$$

Portanto, como $\lim_n F_n(0) = 0 \neq F(0) = 1$, não temos $\lim_n F_n(x) = F(x)$ para todo $x \in \mathbb{R}$. Desse modo se houvesse a exigência de convergência em todos os pontos, não teríamos convergência em distribuição. Entretanto, note que para $x \neq 0$, temos $\lim_n F_n(x) = F(x)$ e, como o ponto 0 não é de continuidade de F , concluímos que $X_n \xrightarrow{D} X$.

Um exemplo mais complexo de convergência em distribuição pode ser visto na análise do limite de

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - EX_i),$$

onde X_i 's são variáveis aleatórias independentes e identicamente distribuídas. Neste, o Teorema Central do Limite afirma que se $VAR(X_i) = \sigma^2 < \infty$, então S_n converge em distribuição para qualquer variável aleatória com distribuição $\mathcal{N}(0, \sigma^2)$. O próximo teorema estabelece duas condições suficientes para que uma sequência de variáveis aleatórias convirja em distribuição.

Teorema 5

Seja X, X_1, X_2, \dots uma sequência de variáveis aleatórias:

- (a) Se X, X_1, X_2, \dots são variáveis aleatórias discretas com $P(X_n = x_i) = p_n(i)$ e $P(X = x_i) = p(i)$, onde $p_n(i) \rightarrow p(i)$ quando $n \rightarrow \infty$ para todo $i = 0, 1, 2, 3, \dots$, então $X_n \rightarrow^D X$.*
- (b) Se X, X_1, X_2, \dots são variáveis aleatórias absolutamente contínuas com densidades dadas respectivamente por f, f_1, f_2, f_3, \dots , onde $f_n(x) \rightarrow f(x)$ quando $n \rightarrow \infty$ em quase todo lugar, então $X_n \rightarrow^D X$.*

Demonstração.

Se $p_n(i) \rightarrow p(i)$ para todo i , então

$$F_{X_n}(x) = \sum_{i: x_i \leq x} p_n(i) \rightarrow \sum_{i: x_i \leq x} p(i) = F_X(x).$$

Onde a convergência acima segue do Teorema da Convergência Dominada, visto que $F_{X_n}(x) \leq 1, \forall x \in \mathbb{R}$.

A prova da parte (b) usa conceitos de Teoria da Medida e será omitida. □

O próximo exemplo mostra que se uma sequência de variáveis aleatórias discretas converge em distribuição, não necessariamente sua função probabilidade de massa converge.

Exemplo 15

Sejam X, X_1, X_2, \dots variáveis aleatórias tais que $P(X = 0) = 1$ e $P(X_n = 1/n) = 1$. Então, temos $F_X(x) = 1$ se $x \geq 0$, e $F_X(x) = 0$ caso contrário; e $F_{X_n}(x) = 1$ se $x \geq 1/n$ e $F_{X_n}(x) = 0$ caso contrário. Logo, $F_{X_n}(x) \rightarrow F_X(x), \forall x \neq 0$, ou seja, $X_n \xrightarrow{D} X$. Porém, $p(0) = 1 \neq 0 = \lim_n p_n(0)$.

O próximo exemplo mostra que se uma sequência de variáveis aleatórias absolutamente contínuas converge em distribuição, não necessariamente sua função densidade de probabilidade converge.

Exemplo 16

Considere uma sequência de variáveis aleatórias X, X_1, X_2, \dots com função de distribuição acumuladas dadas respectivamente por F, F_1, F_2, F_3, \dots , onde

$$F_n(x) = \begin{cases} 0 & , \text{ se } x \leq 0 \\ x(1 - \frac{\sin 2n\pi x}{2n\pi x}) & , \text{ se } 0 < x \leq 1 \\ 1 & , \text{ se } x > 1; \end{cases} \quad e \quad F(x) = \begin{cases} 0 & , \text{ se } x \leq 0 \\ x & , \text{ se } 0 < x \leq 1 \\ 1 & , \text{ se } x > 1. \end{cases}$$

Então F_n e F são absolutamente contínuas com densidade dada por

$$f_n(x) = \begin{cases} 1 - \cos 2n\pi x & , \text{ se } 0 \leq x \leq 1 \\ 0 & , \text{ caso contrário;} \end{cases} \quad e \quad f(x) = \begin{cases} 1 & , \text{ se } 0 < x \leq 1 \\ 0 & , \text{ caso contrário.} \end{cases}$$

É fácil ver que $F_n(x) \rightarrow F(x), \forall x \in \mathbb{R}$. Contudo, $f_n(x) \not\rightarrow f(x)$.

Relação Entre os Tipos de Convergência

A primeira relação que iremos provar é que convergência quase certa implica convergência em probabilidade.

Teorema 6

$$X_n \rightarrow X \text{ cp1} \Rightarrow X_n \rightarrow^P X.$$

Demonstração

Para provar que convergência quase certa implica em convergência em probabilidade, considere a seguinte família de eventos

$$A_{n,\epsilon} = \{w : |X_n(w) - X(w)| \leq \epsilon\}.$$

Logo, pela interpretação de convergência pontual,

$$C = \{w : X_n(w) \rightarrow X(w)\} = \bigcap_{\epsilon > 0} \bigcup_{N=1}^{\infty} \bigcap_{n \geq N} A_{n,\epsilon}.$$

Se $X_n \rightarrow X$ cp1, então $P(C) = 1$. Equivalentemente, pela Lei de De Morgan,

$$D = C^c = \bigcup_{\epsilon > 0} D_{\epsilon}, \text{ onde } D_{\epsilon} = \bigcap_{N=1}^{\infty} \bigcup_{n \geq N} A_{n,\epsilon}^c,$$

e

$$P(\bigcup_{\epsilon > 0} D_{\epsilon}) = 0.$$

Portanto, convergência quase certa implica que $\forall \epsilon > 0, P(D_\epsilon) = 0$. Seja $F_N = \cup_{n \geq N} B_n$. Note que $F_N \downarrow$. Logo, $\lim_N F_N = \cap_{N=1}^\infty \cup_{n \geq N} B_n$. Portanto, pelo axioma da continuidade monotônica da probabilidade, tem-se que

$$P(\cap_{N=1}^\infty \cup_{n \geq N} B_n) = \lim_{N \rightarrow \infty} P(\cup_{n \geq N} B_n).$$

Então,

$$0 = P(D_\epsilon) = \lim_{N \rightarrow \infty} P(\cup_{n \geq N} A_{n,\epsilon}^c) \geq \lim_{N \rightarrow \infty} P(A_{N,\epsilon}^c) = \lim_{N \rightarrow \infty} P(|X_N(w) - X(w)| > \epsilon).$$

Portanto, $X_n \rightarrow^P X$.

O próximo teorema prova que convergência na r -ésima média implica convergência em probabilidade.

Teorema 7

$$X_n \rightarrow^r X \Rightarrow X_n \rightarrow^P X.$$

Demonstração.

Primeiro note que $\frac{|X_n - X|^r}{\epsilon^r} \geq I_{\{w: |X_n - X| > \epsilon\}}$. Logo, tem-se que

$$E\left(\frac{|X_n - X|^r}{\epsilon^r}\right) \geq E(I_{\{w: |X_n - X| > \epsilon\}}),$$

ou seja,

$$\frac{E(|X_n - X|^r)}{\epsilon^r} \geq P(\{w : |X_n - X| > \epsilon\}).$$

Se $X_n \rightarrow^r X$, tem-se que $\lim_{n \rightarrow \infty} E(|X_n - X|^r) = 0$. Então, para todo $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(\{w : |X_n - X| > \epsilon\}) = 0,$$

ou seja, $X_n \rightarrow^P X$. □

O próximo exemplo prova que nem convergência em probabilidade, nem convergência na r -ésima média implicam convergência quase certa.

Exemplo 17

Seja X uma variável aleatória com distribuição uniforme no intervalo $[0, 1]$, e considere a sequência de intervalos definida por

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

para $m = 0, 1, 2, \dots$ e $i = 0, 1, \dots, 2^m - 1$.

Note que tem-se 2^m intervalos de comprimento 2^{-m} que cobrem todo o intervalo $[0, 1]$, e o comprimento dos intervalos fica cada vez menor tendendo a 0. Definamos

$$Y_n(w) = \begin{cases} 1 & \text{se } X(w) \in I_n, \\ 0 & \text{se } X(w) \notin I_n. \end{cases}$$

A sequência Y_1, Y_2, \dots converge em probabilidade para 0, pois para $0 < \epsilon \leq 1$,

$$P(|Y_n| \geq \epsilon) = P(Y_n = 1) = P(X \in I_n),$$

e esta probabilidade, que é igual ao comprimento de I_n , converge para zero quando $n \rightarrow \infty$.

Esta sequência também converge na r -ésima média para todo $r > 0$, visto que

$E(|Y_n|^r) = P(Y_n = 1) \rightarrow 0$ quando $n \rightarrow \infty$. Logo, Y_n converge na r -ésima média para 0.

Porém para todo $w \in \Omega$, $Y_n(w) = 1$ para um número infinito de n 's e $Y_n(w) = 0$ para um número infinito de n 's. Portanto, $Y_n(w)$ não converge para todo w , o que implica que Y_n não converge quase certamente. \square

O próximo teorema estabelece mais uma relação entre convergência quase certa e convergência em probabilidade.

Teorema 8

$X_n \rightarrow^P X$ se, e somente se, toda subsequência $\{X_{n_k}\}$ possui uma outra subsequência $\{X_{n_{k(i)}}\}$ tal que $X_{n_{k(i)}} \rightarrow X$ cp1 para $i \rightarrow \infty$.

Demonstração.

Suponha que $X_n \rightarrow^P X$, então dada qualquer subsequência $\{X_{n_k}\}$, escolha uma outra subsequência $\{X_{n_{k(i)}}\}$ tal que $j \geq k(i)$ implica que $P(|X_{n_j} - X| \geq i^{-1}) < 2^{-i}$. Em particular, temos que $P(|X_{n_{k(i)}} - X| \geq i^{-1}) < 2^{-i}$. Seja $A_i = \{|X_{n_{k(i)}} - X| \geq i^{-1}\}$, então $\sum_{i=1}^{\infty} P(A_i) < \sum_{i=1}^{\infty} 2^{-i} = 1 < \infty$. Logo, pelo Lema de Borel-Cantelli, temos que $P(A_i \text{ infinitas vezes}) = 0$, ou seja, $P(A_i \text{ finitas vezes}) = 1$. Portanto, $|X_{n_{k(i)}} - X| < i^{-1}$ exceto para um número finito de i 's com probabilidade 1. Portanto, $X_{n_{k(i)}} \rightarrow X$ cp1. Se X_n não converge para X em probabilidade, existe um $\epsilon > 0$ e uma subsequência $\{X_{n_k}\}$ tal que $P(|X_{n_k} - X| > \epsilon) > \epsilon$. Logo nenhuma subsequência de $\{X_{n_k}\}$ pode convergir para X em probabilidade, logo pelo Teorema 6, nenhuma subsequência converge para X quase certamente. □

O próximo exemplo mostra que convergência em probabilidade não implica convergência na r -ésima média

Exemplo 18

Seja X uma variável aleatória com distribuição uniforme no intervalo $[0, 1]$. Considere a seguinte sequência de variáveis aleatórias

$$Y_n(w) = \begin{cases} 2^n & \text{se } X(w) \in (0, \frac{1}{n}), \\ 0 & \text{se } X(w) \notin (0, \frac{1}{n}). \end{cases}$$

Então, $P(|Y_n| > \epsilon) = P(X(w) \in (0, \frac{1}{n})) = \frac{1}{n} \rightarrow 0$, mas $E(|Y_n|^r) = 2^{nr} \frac{1}{n} \rightarrow \infty$.

O próximo exemplo mostra que convergência quase-certa não implica convergência na r -ésima média.

Exemplo

Seja $\{Y_n, n \geq 1\}$ uma sequência de variáveis aleatórias onde

$$P(Y_n = 0) = 1 - n^{-2} \text{ e } P(Y_n = e^n) = n^{-2}.$$

Portanto, para todo $\epsilon > 0$,

$$P(|Y_n| > \epsilon) = P(Y_n > \epsilon) \leq P(Y_n = e^n) = n^{-2}.$$

Logo,

$$\sum_{n=1}^{\infty} P(|Y_n| > \epsilon) \leq \sum_{n=1}^{\infty} n^{-2} < \infty.$$

Então, Borel-Cantelli implica que $|Y_n| > \epsilon$ infinitas vezes com probabilidade 0, o que por sua vez implica que $Y_n \rightarrow 0$ com probabilidade 1, ou seja, $Y_n \rightarrow 0$ cp1. Porém,

$$E|Y_n|^r = \frac{e^{nr}}{n^2} \rightarrow \infty,$$

para todo $r > 0$. Portanto, $Y_n \rightarrow 0$ cp1, mas $Y_n \not\rightarrow^r 0$ para todo $r > 0$.

O próximo teorema trata da relação entre convergência em distribuição e convergência em probabilidade.

Teorema 9

As seguintes relações entre os tipos de convergência são válidas:

- (a) $X_n \rightarrow^P X \Rightarrow X_n \rightarrow^D X$
- (b) *Se $X_n \rightarrow^D c$, onde c é uma constante, então $X_n \rightarrow^P c$.*

Demonstração

Para parte (a), suponha que $X_n \rightarrow^P X$ e seja x um ponto de continuidade de F_X . Queremos provar que $F_{X_n}(x) \rightarrow F_X(x)$ quando $n \rightarrow \infty$.

Como para $\epsilon > 0$, $X_n \leq x \Rightarrow X \leq x + \epsilon$ ou $|X - X_n| > \epsilon$, temos $\{w : X_n(w) \leq x\} \subseteq \{w : X(w) \leq x + \epsilon\} \cup \{w : |X_n(w) - X(w)| > \epsilon\}$. Logo,

$$F_{X_n}(x) = P(X_n \leq x) \leq F_X(x + \epsilon) + P(|X_n - X| > \epsilon).$$

Por outro lado, $X \leq x - \epsilon \Rightarrow X_n \leq x$ ou $|X_n - X| > \epsilon$ de modo que

$$F_X(x - \epsilon) \leq F_{X_n}(x) + P(|X_n - X| > \epsilon).$$

Juntando as duas desigualdades, temos que $\forall \epsilon > 0$, and $\forall n$,

$$\begin{aligned} F_X(x - \epsilon) - P(|X_n - X| > \epsilon) \\ \leq F_{X_n}(x) \leq F_X(x + \epsilon) + P(|X_n - X| > \epsilon). \end{aligned}$$

Como $X_n \rightarrow^P X$, para qualquer $\delta > 0$, existe N tal que para $n \geq N$, temos que

$$F_X(x - \epsilon) - \delta \leq F_{X_n}(x) \leq F_X(x + \epsilon) + \delta.$$

Finalmente, como x é ponto de continuidade de F_X , para ϵ suficientemente pequeno, temos que

$$F_X(x) - 2\delta \leq F_X(x - \epsilon) - \delta \leq F_{X_n}(x) \leq F_X(x + \epsilon) + \delta \leq F_X(x) + 2\delta.$$

Ou seja, $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$. Para parte (b), suponha que $X_n \rightarrow^D c$. Note que a função de distribuição de uma variável aleatória constante c é:

$$F_c(x) = \begin{cases} 1 & \text{se } x \geq c, \\ 0 & \text{se } x < c. \end{cases}$$

Pela convergência em distribuição, tem-se que

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = 0, \text{ se } x < c$$

e

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = 1, \text{ se } x > c.$$

Logo, para $\epsilon > 0$,

$$\begin{aligned} P(|X_n - c| \leq \epsilon) &= P(c - \epsilon \leq X_n \leq c + \epsilon) \\ &\geq P(c - \epsilon < X_n \leq c + \epsilon) = \\ &F_{X_n}(c + \epsilon) - F_{X_n}(c - \epsilon) \rightarrow 1 \text{ quando } n \rightarrow \infty. \end{aligned}$$

Ou seja, $\forall \epsilon > 0, \lim_{n \rightarrow \infty} P(|X_n - c| > \epsilon) = 0$.

Relação Entre os Tipos de Convergência

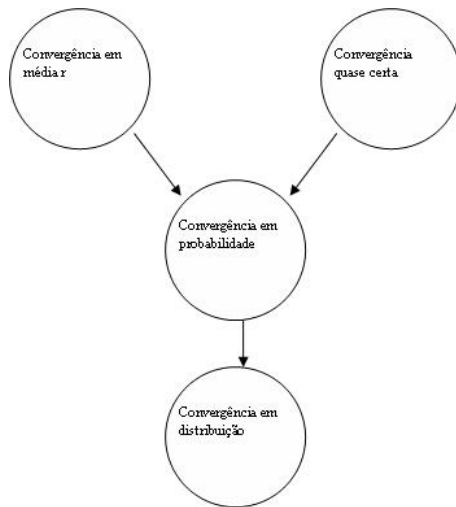


Figura: Relação entre os tipos de convergência.

Exemplo 19

Para $n \geq 1$, $X_n \sim U(0, 1)$ são variáveis aleatórias i.i.d. Defina $Y_n = \min(X_1, X_2, \dots, X_n)$ e $U_n = nY_n$. Mostre que

(a) $Y_n \xrightarrow{P} 0$,

(b) $U_n \xrightarrow{D} U$, sendo $U \sim \text{Exp}(1)$.

Solução: Para parte (a), note que

$$\begin{aligned} P(|Y_n| > \epsilon) &= P(Y_n > \epsilon) \\ &= P(X_1 > \epsilon, X_2 > \epsilon, \dots, X_n > \epsilon). \end{aligned}$$

Como os X_n são independentes temos que a última expressão é igual a

$$(P(X_1 > \epsilon))^n = (1 - \epsilon)^n.$$

Como $(1 - \epsilon)^n \rightarrow 0$ quando $n \rightarrow \infty$, temos que $Y_n \xrightarrow{P} 0$. Para parte (b), note que

$$\begin{aligned} F_{U_n}(x) &= P(U_n \leq x) = 1 - P(U_n > x) \\ &= 1 - P(nY_n > x) = 1 - P(Y_n > x/n) \end{aligned}$$

De acordo com a parte (a), esta expressão é igual a $1 - (1 - x/n)^n$, que por sua vez converge para $1 - e^{-x}$ quando $n \rightarrow \infty$, que é igual a $F_U(x)$.

Convergência de Vetores Aleatórios

- Para o caso vetorial as definições de convergência sofrem algumas adaptações.
- Para as convergências quase certa e em probabilidade, precisamos avaliar a proximidade entre os vetores aleatórios X_n e X pelo comportamento da norma da diferença entre eles.
- Em geral, essa norma é calculada por $\|X_n - X\| = (\sum_{j=1}^k (X_{nj} - X_j)^2)^{1/2}$, onde k é a dimensão dos vetores e X_{nj} a coordenada j do vetor X_n .
- Pode-se verificar que a convergência do vetor aleatório, quase certamente ou em probabilidade, ocorre se, e somente se, existir a mesma convergência em cada uma das variáveis que compõe o vetor aleatório. Dessa forma, o caso multidimensional pode ser estudado a partir de repetidas aplicações do caso univariado.
- Para convergência em distribuição de vetores aleatórios, requeremos que a função de distribuição conjunta $F_n(x)$ convirja para $F(x)$, em todos os pontos de continuidade da função F .
- Entretanto, lembremos que da função de distribuição conjunta podemos obter as marginais, mas o caminho inverso nem sempre é possível. Por essa razão, diferentemente das convergências quase certa e em probabilidade, não podemos reduzir o estudo da convergência em distribuição de vetores aleatórios, ao comportamento das suas respectivas coordenadas.

- Não temos equivalência, mas apenas implicação, em uma das direções. Ou seja, se o vetor converge em distribuição então cada componente também converge em distribuição, para a correspondente marginal da função de distribuição limite. Entretanto a recíproca não é em geral, verdadeira.
- Para convergência em média r de vetores aleatórios, requeremos que $E(\|X_n - X\|^r)$ convirja para zero quando n tender a infinito. Note que, se ϕ é uma função convexa e $\phi(0) = 0$, então para $0 \leq a \leq 1$,

$$\phi(ax) = \phi(ax + (1 - a)0) \leq a\phi(x) + (1 - a)\phi(0) = a\phi(x).$$

- Consequentemente, para x e y não negativos,

$$\phi(x + y) = \frac{x}{x + y}\phi(x + y) + \frac{y}{x + y}\phi(x + y) \geq \phi(x) + \phi(y).$$

Similarmente, se ϕ for côncava e $\phi(0) = 0$, então para x e y não negativos, $\phi(x + y) \leq \phi(x) + \phi(y)$.

- Então, como $\phi(x) = x^s$ é convexo (resp., côncavo) para $s \geq 1$ (resp., concava), temos que

$$\left(\sum_{j=1}^k (X_{nj} - X_j)^2\right)^{r/2} \geq \sum_{j=1}^k |X_{nj} - X_j|^r,$$

se $r \geq 2$ e

$$\left(\sum_{j=1}^k (X_{nj} - X_j)^2\right)^{r/2} \leq \sum_{j=1}^k |X_{nj} - X_j|^r,$$

se $r \leq 2$.

- Portanto, se $r \geq 2$, então se X_n converge para X em média r , o mesmo vale para as componentes correspondentes dos vetores. E se $r \leq 2$, então se X_{nj} converge para X_j em média r , para $j = 1, 2, \dots, k$, então X_n converge para X em média r .