Probabilidade (PPGECD000000001)

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

Sessão 11

Raydonal Ospina

Departamento de Estatística
Universidade Federal da Bahia
Salvador/BA

Motivação

Entre outras coisas, a Lei dos grandes Números nos permite formalizar a idéia que à medida que o número de repetições de um experimento cresce, a frequência relativa f_A de algum evento A converge (quase certamente) para a probabilidade teórica $P(A)$. É este fato que nos permite estimar o valor da probabilidade de um evento A, baseado na frequência relativa de A em um grande número de repetições de um experimento. É também este fato que justifica a intuição que temos que eventos com probabilidade próximas de 1, quase sempre ocorrem; e que eventos com probabilidade próximas de 0 quase sempre não ocorrem.

Motivação

Por exemplo, se uma nova peça for produzida e não tivermos conhecimento anterior sobre quão provável será que a peça seja defeituosa, poderemos proceder à inspeção de um grande número dessas peças, digamos N, contarmos o número de peças defeituosas dentre elas, por exemplo n, e depois empregarmos n/N com uma aproximação da probabilidade de que uma peça seja defeituosa. O número n/N é uma variável aleatória, e seu valor depende essencialmente de duas coisas. Primeira, o valor de n/N depende da probabilidade básica, mas desconhecida, p de que uma peça seja defeituosa. Segunda, depende daquelas N peças que tenham sido inspecionadas. O que a Lei dos Grandes Números mostra é que se a técnica de selecionar as N peças for aleatória, então o quociente n/N convergirá quase certamente para p. (Evidentemente, a seleção das N peças é importante. Se fôssemos escolher somente aquelas peças que exibissem algum defeito físico externo, por exemplo, poderíamos prejudicar seriamente nossos cálculos.)

Mais formalmente, considere um experimento básico, com a variável aleatória X representando o valor de um característico numérico do resultado (no caso anterior, temos que X seria a função indicadora do evento A). Pensemos na realização deste experimento N vezes (N grande), de tal maneira que as realizações sejam independentes. Suponhamos que depois de cada realização do experimento registre-se o valor do característico numérico do resultado; chamemos este um valor observado. A Lei dos Grandes Números afirma que a média aritmética dos N valores observados converge, em certo sentido, para a média EX, quando $N \to \infty$.

Motivação

Vamos agora construir um modelo para o experimento repetido que apresentamos acima. Para experimentos dessa natureza, um resultado possível é uma sequência de N resultados possíveis do experimento básico. Como estamos interessados em analisar a convergência para N grande, se Ω_0 é o espaço amostral do experimento básico, o espaço amostral do experimento global consiste nas sequências infinitas de elementos de Ω_0, ou seja,

$$\begin{aligned}\Omega &= \{(w_1, w_2, \ldots) : w_i \in \Omega_0, i = 1, 2, \ldots\}\\ &= \Omega_0 \times \Omega_0 \times \ldots = \Omega^{\infty},\end{aligned}$$

onde w_i é o resultado do i-ésimo ensaio do experimento básico. Podemos completar o modelo utilizando a σ-álgebra produto para $\mathcal{A}$ e a probabilidade produto para P, pois os ensaios são independentes.

Motivação

Já que vamos registrar um certo característico do i-ésimo resultado para todo i, estaremos registrando os valores de uma sequência de variáveis aleatórias. Intuitivamente, $X(w_0)$ representa o valor do característico numérico do experimento básico ($w_0 \in \Omega_0$), então, quando o resultado da sequência de realizações for $w = (w_1, w_2, \ldots)$, os valores observados serão $X(w_1), X(w_2), \ldots$ É conveniente representar por X_n o resultado observado na n-ésima realização. Assim, X_n é função do resultado w do experimento global, com $X_n(w) = X(w_n)$, e no decorrer serão registrados os valores das variáveis aleatórias $X_1, X_2, \ldots$ Notemos que X_n tem a mesma distribuição de X, pois trata-se de uma sequência de repetições do mesmo experimento. Como as X_n dependem de realizações independentes, elas são independentes, onde $X_1, X_2, \ldots$ são independentes se para todo $n \geq 2$, $X_1, \ldots, X_n$ são independentes.

Uma versão da Lei dos Grandes Números diz que se $X_1, X_2, \ldots$ são i.i.d. e integráveis, então

$$\frac{X_1 + \ldots + X_n}{n} \to EX_1.$$

Quando o tipo de convergência é convergência em probabilidade, chamamos de Lei Fraca dos Grandes Números, e quando temos convergência quase certa, chamamos de Lei Forte dos Grandes Números. Como vimos em capítulo anterior, convergência quase-certa implica convergência em probabilidade, portanto se uma sequência de variáveis aleatórias satisfaz a Lei Forte dos Grandes Números, então ela também satisfaz a Lei Fraca.

Motivação

Para esclarecer as diferenças entre as Leis Fraca e Leis Fortes, considere o caso em que $X_i \sim Bernoulli(p)$ é a função indicadora de certo evento A e n_A é o número de vezes que o evento A ocorre em n realizações do experimento. Então, a Lei Fraca afirma que $\frac{n_A}{n} \rightarrow^P p$, o que é equivalente a dizer que para todo $\epsilon > 0$ podemos encontrar um n suficientemente grande tal que, a probabilidade de $\frac{n_A}{n}$ estar entre $p - \epsilon$ e $p + \epsilon$, é maior que $1 - \delta$ para qualquer $\delta > 0$ especificado. Em outras palavras, se realizarmos muitas sequências $Bernoulli(p)$ de tamanho n, espera-se que apenas em uma fração delas menor que δ, temos que $\frac{n_A}{n}$ está fora do intervalo $(p - \epsilon, p + \epsilon)$. Note que a Lei Fraca não dá nenhuma informação sobre a existência ou o valor do limite de $\frac{n_A}{n}$. Em contraste, a Lei Forte garante que o conjunto de todas as realizações do experimento, para as quais $\lim_n \frac{n_A}{n} = p$, é um evento com probabilidade 1. Se fixarmos $\epsilon > 0$, o conjunto das realizações dos experimentos para os quais $p - \epsilon < \frac{n_A}{n} < p + \epsilon$, para n suficientemente grande é um evento com probabilidade 1. A Lei Forte assegura que dado $\epsilon > 0$, com probabilidade 1, os termos da sequência de frequência relativas de uma particular realização do experimento realmente estarão no intervalo $(p - \epsilon, p + \epsilon)$.

Lei Fraca dos Grandes Números

Nos slides anteriores, motivamos o resultado da Leis dos Grandes Números para variáveis aleatórias independentes e identicamente distribuídas. Nesta seção, analisaremos duas versões da Lei Fraca dos Grandes Números, na primeira delas não é necessário assumir que as variáveis aleatórias são identicamente distribuídas. Vamos usar a desigualdade de Chebyshev para provar a Lei Fraca dos Grandes Números de Chebyshev.

Teorema 1

Lei Fraca de Chebyshev *Sejam* $X_1, X_2, \ldots$ *variáveis aleatórias independentes 2 a 2 com variâncias finitas e uniformemente limitadas (ou seja, existe c finito tal que para todo n,* $\mathrm{Var}X_n \leq c$*). Então,* $X_1, X_2, \ldots$ *satisfazem a Lei Fraca dos Grandes Números:*

$$\frac{S_n - ES_n}{n} \to^P 0.$$

Demonstração.

Precisamos provar que para todo $\epsilon > 0$,

$$P(\frac{|S_n - ES_n|}{n} \geq \epsilon) \rightarrow 0 \text{ quando } n \rightarrow \infty.$$

Como as variáveis aleatórias são independentes 2 a 2, temos que

$$Var(S_n) = \sum_{i=1}^{n} Var(X_i) \leq nc.$$

Pela desigualdade de Chebyshev, temos que

$$P(|S_n - ES_n| \geq n\epsilon) \leq \frac{Var(S_n)}{\epsilon^2 n^2} \leq \frac{c}{\epsilon^2 n} \rightarrow 0.$$

$\square$

Lei Fraca de Bernoulli

Corolário 1 (Lei Fraca dos Grandes Números de Bernoulli.)

Consideremos uma sequência de ensaios binomiais independentes, tendo a mesma probabilidade p de "sucesso" em cada ensaio. Se S_n é o número de sucessos nos primeiros n ensaios, então

$$\frac{S_n}{n} \to^P p$$

Demonstração.

Seja $X_n = 1$ se o n-ésimo ensaio é sucesso, $X_n = 0$ caso contrário. Então, $X_1, X_2, \ldots$ são i.i.d. e integráveis com média $\mu = p$. Como $VarX_n = p(1-p)$, a Lei Fraca de Chebyshev implica que $\frac{S_n - np}{n} \to^P 0$, ou, equivalentemente, $\frac{S_n}{n} \to^P p$. $\square$

Aplicação da Lei Fraca

Podemos utilizar a Lei Fraca dos Grandes Números para responder a seguinte questão: quantas repetições de um experimento devemos realizar a fim de termos uma probabilidade ao menos $0,95$ para que a frequência relativa difira de $p = P(A)$ por menos do que, digamos, 0,01? Utilizando a equação (1), onde S_n é o número de ocorrências do evento A em n realizações do experimento temos que $S_n/n = f_A$, $ES_n = np$, $VarS_n = np(1-p)$, e:

$$P(|f_A - p| \geq 0,01) \leq \frac{p(1-p)}{n(0,01)^2},$$

ou seja, queremos que $\frac{p(1-p)}{n(0,01)^2} \leq 0,05$, o que é equivalente a $n \geq \frac{p(1-p)}{0,05(0,01)^2}$. Substituindo os valores específicos de $0,05$ e $0,01$ por δ e ϵ, respectivamente, teremos

$$P(|f_A - p| < \epsilon) \geq 1 - \delta \text{ sempre que } n \geq \frac{p(1-p)}{\delta(\epsilon)^2}.$$

Aplicação da Lei Fraca

Em muitos problemas, não conhecemos o valor de $p = P(A)$ e, por isso, não poderemos empregar o limite acima. Nesse caso, poderemos empregar o fato de que $p(1-p)$ toma seu valor máximo quando $p = 1/2$, e esse valor máximo é igual a 1/4.
Consequentemente, estamos certamente seguros se afirmamos que para $n \geq \frac{1}{4\epsilon^2\delta}$ teremos

$$P(|f_A - p| < \epsilon) \geq 1 - \delta.$$

Exemplo 1

Peças são produzidas de tal maneira que a probabilidade de uma peça ser defeituosa é p (admitida desconhecida). Um grande número de peças, digamos n, são classificadas como defeituosas ou perfeitas. Que valor deverá ter n de maneira que possamos estar 99% certos de que a frequência relativa de defeituosas difere de p por menos de $0,05$?
Solução: Porque não conhecemos o valor de p, deveremos aplicar a última fórmula com $\epsilon = 0,05$, $\delta = 0,01$. Deste modo encontraremos que se $n \geq \frac{1}{4(0{,}05)^2 0{,}01} = 10.000$, a condição exigida será satisfeita.

Lei Fraca de Khintchin

A hipótese de variâncias finitas pode ser eliminada e o próximo teorema prova uma versão da Lei Fraca dos Grandes Números para variáveis aleatórias i.i.d. e integráveis.

Teorema 2

Lei Fraca de Khintchin. *Se $X_1, X_2, \ldots$ são i.i.d. e integráveis com média comum μ, então*

$$\frac{S_n}{n} \to^P \mu.$$

Demonstração.

É consequência da Lei Forte de Kolmogorov e do fato que convergência quase certa implica convergência em probabilidade. □

Exemplo 2

Sejam $\{X_n : n \geq 1\}$ variáveis i.i.d. com média μ e variância σ^2, ambas finitas. Prove que $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \to^P \sigma^2$.

Solução: $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$ Pela Lei Fraca de Kintchin, temos que

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \to^P E(X_i^2) = \sigma^2 + \mu^2, \qquad , \overline{X} \to^P E(X_i) = \mu.$$

Como funções contínuas preservam convergência, temos que

$$\overline{X}^2 \to^P \mu^2.$$

Logo, temos que

$$(\frac{1}{n}\sum_{i=1}^{n} X_i^2, \overline{X}^2) \to^P (\sigma^2 + \mu^2, \mu^2).$$

Finalmente, como funções contínuas preservam convergência

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 \to^P \sigma^2.$$

Lei Forte dos Grandes Números

Antes de discutir sobre a Lei Forte dos Grandes Números, vamos apresentar uma extensão da desigualdade de Chebyshev.

Lema 1

Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes tais que $EX_k = 0$ e $VarX_k < \infty, k = 1, \ldots, n$. Então, para todo $\lambda > 0$,

$$P(\max_{1 \leq k \leq n} |S_k| \geq \lambda) \leq \frac{1}{\lambda^2} VarS_n = \frac{1}{\lambda^2} \sum_{k=1}^{n} VarX_k,$$

onde $S_k = X_1 + \ldots + X_k$.

Teorema 3 (Primeira Lei Forte de Kolmogorov)

Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e integráveis, e suponha que

$$\sum_{n=1}^{\infty} \frac{VarX_n}{n^2} < \infty.$$

Então, as X_n satisfazem a Lei Forte dos Grandes Números, ou seja,

$$\frac{X_1 + \ldots + X_n}{n} - \frac{(EX_1 + \ldots + EX_n)}{n} \to 0 \text{ cp1}.$$

Aplicação da Primeira Lei Forte de Kolmogorov

Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes com $X_n \sim Poisson(\sqrt{n})$, para cada $n \geq 1$. Calcule o limite quase-certo de $\overline{X}$.

Solução: Como $VarX_n = \sqrt{n}$, temos que

$$\sum_{n=1}^{\infty} \frac{VarX_n}{n^2} = \sum_{n=1}^{\infty} \frac{\sqrt{n}}{n^2} < \infty.$$

Logo, a primeira Lei Forte de Kolmogorov implica que

$$\overline{X} - \frac{EX_1 + \cdots + EX_n}{n} \to 0 \text{ cp1, ou seja}$$

$$\overline{X} - \frac{\sqrt{1} + \sqrt{2} + \cdots + \sqrt{n}}{n} \to 0 \text{ cp1.}$$

Pelo teste da integral, pode-se verificar que

$$\sqrt{1} + \sqrt{2} + \cdots + \sqrt{n} \geq \frac{2n^{3/2}}{3}.$$

Portanto,

$$\frac{\sqrt{1} + \sqrt{2} + \cdots + \sqrt{n}}{n} \geq \frac{2n^{1/2}}{3} \to \infty.$$

Logo, $\overline{X} \to \infty$ cp1.

Segunda Lei Forte de Kolmogorov

Antes de enunciarmos a Segunda Lei Forte de Kolmogorov, considere o seguinte lema:

Lema 2

Seja X uma variável aleatória integrável com função de distribuição F. Então,

$$\sum_{n=1}^{\infty}\left(\frac{1}{n^2}\int_{-n}^{n} x^2 dF(x)\right) < \infty.$$

Teorema 4 (Segunda Lei Forte de Kolmogorov)

Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes, identicamente distribuídas e integráveis, com $EX_n = \mu$. Então,

$$\frac{X_1 + \ldots + X_n}{n} \to \mu \text{ quase certamente.}$$

Aplicação da Segunda Lei Forte de Kolmogorov

Exemplo 3

As variáveis X_n, $n \geq 1$, são independentes e todas têm distribuição Exponencial de parâmetro λ. Mostre que a sequência $\{X_n^2 : n \geq 1\}$ satisfaz a Lei Forte dos Grandes Números.

Solução: De acordo com a Segunda Lei Forte de Kolmogorov, precisamos mostrar que EX_n^2 é finita para todo n. Como $EX_n^2 = VarX_n + (EX_n)^2 = \frac{2}{\lambda^2} < \infty$, temos que a sequência $\{X_n^2 : n \geq 1\}$ satisfaz a Lei Forte dos Grandes Números.

Exemplo 4

Seja $\{X_n : n \geq 1\}$ uma sequência de variáveis aleatórias i.i.d., seguindo o modelo Uniforme contínuo em $(0, 1)$. Calcule o limite, quase certo, para $\frac{1}{n}\sum_{k=1}^{n}(-\log(X_k))$ quando $n \to \infty$.

Solução: Vamos tentar usar a Lei Forte dos Grandes Números. Para isso, precisamos calcular $E(-\log X_k)$.

$$E(-\log X_k) = \int_0^1 -\log x dx = -x \log x|_0^1 + \int_0^1 dx = 1.$$

Portanto, temos que $\frac{1}{n}\sum_{k=1}^{n}(-\log(X_k)) \to 1$ cp1.

Convergência da Função de Distribuição Empírica

A seguir veremos uma importante consequência da Lei Forte dos Grandes Números para a área de Estatística Aplicada. Sejam $X_1, X_2, , \ldots, X_n$ variáveis aleatórias em $(\Omega, \mathcal{A}, P)$ independentes e identicamente distribuídas com função de distribuição F. Essas variáveis podem representar a amostra observada de uma certa quantidade de interesse. A *função de distribuição empírica* ou *amostral*, denotada por F_n^e, é definida para todo $x \in \mathbb{R}$ e $w \in \Omega$ por:

$$\begin{aligned} & F_n^e(x, w) \\ & = \frac{1}{n}[\text{no. de } i\text{'s tais que } X_i(w) \leq x, i = 1, 2, \ldots, n]. \end{aligned}$$

Para uma particular trajetória $w_0 \in \Omega$, obtemos o conjunto de valores fixados $X_1(w_0) = x_1, \ldots, X_n(w_0) = x_n$. Se os x_i's são todos diferentes, então $F_n^e(x, w_0)$ é uma função de distribuição com saltos $1/n$ em cada um desses valores.

Convergência da Função de Distribuição Empírica

Considere um $x_0 \in \mathbb{R}$ fixo. Então $F_n^e(x_0, w)$ é uma variável aleatória, pois é uma função das variáveis aleatórias $X_1, X_2, \ldots, X_n$. Se $Y_i = I_{X_i \leq x_0}$, $i = 1, 2, \ldots, n$, então $F_n^e(x_0, w) = \frac{1}{n}\sum_{i=1}^n Y_i(w)$. Como as variáveis aleatórias Y_i são funções de famílias disjuntas de variáveis aleatórias independentes, elas também são independentes. Além disso, temos que $Y_i \sim \mathit{Bernoulli}(p)$ com

$$p = P(Y_i = 1) = P(X_i \leq x_0) = F(x_0).$$

Portanto, concluímos que pela Lei Forte de Kolmogorov, para cada valor $x_0 \in \mathbb{R}$ fixo, temos $F_n^e(x_0, w) \to F(x_0)$ cp1. O *Teorema de Glivenko-Cantelli* também conhecido como *Teorema Fundamental da Estatística* afirma que a função de distribuição empírica converge para a função de distribuição populacional, quase certamente em Ω e uniformemente em $\mathbb{R}$.

Convergência da Função de Distribuição Empírica

Teorema 5

Sejam $X_1, X_2, \dots, X_n$ variáveis aleatórias em $(\Omega, \mathcal{A}, P)$, independentes e identicamente distribuídas com função de distribuição F. Seja F_n^e a correspondente função de distribuição empírica, então:

$$P(\lim_n \sup_{x \in \mathbb{R}} |F_n^e(x, w) - F(x)| = 0) = 1.$$

Demonstração.

Para cada x fixo, os argumentos anteriores garantem convergência quase certa. A prova de que este resultado pode ser estendido, usa técnicas de Análise Matemática e será omitida. $\square$

Recíproca da Lei Forte de Kolmogorov

A Lei Forte afirma que se as variáveis aleatórias X_n são integráveis, então $\frac{S_n}{n}$ converge para um limite finito ($= EX_1$) com probabilidade 1. A recíproca diz que se as X_n não forem integráveis, então com probabilidade 1, $\frac{S_n}{n}$ não convergirá para um limite finito.

Teorema 6

Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas. Se $E|X_1| = \infty$, então, com probabilidade 1, a sequência $\frac{|S_n|}{n}$ não é limitada.

Um Exemplo de Divergência das Médias

Uma variável aleatória tem distribuição de Cauchy de parâmetro a se, para $a > 0$

$$f_X(x) = \frac{1}{\pi} \cdot \frac{a}{a^2 + x^2}.$$

Assuma que X_n são i.i.d. segundo uma distribuição de Cauchy de parâmetro a. Seja $S_n = \frac{1}{n} \sum_{i=1}^{n} X_n$. Utilizando a definição e as propriedades da função característica pode-se provar que

$$\phi_{X_n}(u) = e^{-a|u|}, \text{ e } \phi_{S_n}(u) = e^{-a|u|}.$$

Então, as médias S_n são distribuídas exatamente como uma das parcelas da soma. Para $n \geq m$, após alguma manipulação algébrica, temos que

$$S_n - S_m = (1 - \frac{m}{n})([Z_{n,m}] - [Y_{n,m}]),$$

onde $Z_{n,m} = \frac{1}{n-m} \sum_{i=m+1}^{n} X_i$ e $Y_{n,m} = \frac{1}{m} \sum_{i=1}^{m} X_i$. Observe que como $Z_{n,m}$ e $Y_{n,m}$ são médias de conjuntos disjuntos de variáveis aleatórias independentes, elas são independentes uma da outra. Ainda mais, pelo resultado para ϕ_{S_n}, é o caso que elas são identicamente distribuídas com função característica igual a $e^{-a|u|}$. Seja $W_{n,m} = Z_{n,m} - Y_{n,m}$, nós vemos que $S_n - S_m = (1 - \frac{m}{n})W_{n,m}$.

Contudo,

$$\phi_{W_{n,m}}(u) = \phi_{Z_{n,m}}(u)\phi_{Y_{n,m}}(-u) = e^{-2a|u|}.$$

Então, $W_{n,m}$ tem uma distribuição fixa, não degenerada que é independente de n e m. Fixando, $n = 2m$, temos que

$$\phi_{S_{2m}-S_m}(u) = e^{-a|u|}.$$

Portanto, quando $m \to \infty$, $S_{2m} - S_m$ não converge para zero, mas para todo m, tem uma distribuição Cauchy de parâmetro a. Portanto, S_n não satisfaz o critério de convergência de Cauchy e não é convergente.

Observação

Observe que isto não é um contra-exemplo a Lei Forte de Kolmogorov, tendo em vista que uma variável aleatória que tem distribuição de acordo com uma Cauchy não tem valor esperado definido, ou seja

$$EX = -\int_{-\infty}^{0} \frac{1}{\pi} \frac{a|x|}{a^2 + x^2} dx + \int_{0}^{\infty} \frac{1}{\pi} \frac{ax}{a^2 + x^2} dx,$$

é indefinido, visto que ambas as integrais são infinitas. Este exemplo serve para ilustrar que a suposição da existência de EX é necessária para a Lei Forte dos Grandes Números.