Teorema do Limite Central

Elementos de Probabilidades e Estatística

Teorema do Limite Central

Em muitas situações, é difícil determinar a distribuição da soma de variáveis aleatórias, mesmo quando elas são independentes.

O teorema que veremos a seguir estabelece que, sob certas condições, a distribuição da soma ou da média de um grande número de variáveis aleatórias independentes e identicamente distribuídas tenderá a se aproximar de uma distribuição normal, independentemente da distribuição original das variáveis. Esta propriedade fundamental explica por que a distribuição normal é tão amplamente utilizada em análises estatísticas e em aplicações práticas.

Teorema do Limite Central (cont.)

Teorema do Limite Central (TLC)

Sejam $X_1, X_2 \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com $E(X_i) = \mu < \infty$ e $V(X_i) = \sigma^2 < \infty$. Considere-se $S_n = \sum_{i=1}^{n} X_i$ e $\overline{X_n} = \sum_{i=1}^{n} \frac{X_i}{n}$, então quando $n \to +\infty$

$$S_n \sim^{\text{aprox.}} N(n\mu, \sqrt{n\sigma^2}) \Leftrightarrow Z_n = \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \sim^{\text{aprox.}} N(0,1)$$

ou seja,

$$\lim_{n \to +\infty} P\left(\frac{S_n - E(S_n)}{\sqrt{V(S_n)}} \leq z\right) = \Phi(z),$$

onde $\Phi(.)$ é a função de distribuição da normal reduzida (padrão), ou seja, $N(0,1)$.

- $\sim^{aprox.}$ deve ler-se *tem distribuição aproximadamente*.
- Geralmente considera-se n suficientemente grande se $n \geq 30$.

Ciências ULisboa

Teorema do Limite Central (cont.)

Teorema do Limite Central (TLC)

De modo equivalente, se tem

$$\overline{X}_n \sim^{\text{aprox.}} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Leftrightarrow Z_n = \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{V(\overline{X}_n)}} = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim^{\text{aprox.}} N(0,1)$$

ou seja,

$$\lim_{n \to +\infty} P\left(\frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{V(\overline{X}_n)}} \leq z\right) = \Phi(z).$$

Note que:

- $E(S_n) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \mu = n\mu$
- $V(S_n) = V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) = \sum_{i=1}^{n} \sigma^2 = n\sigma^2$
- $E(\overline{X}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n} n\mu = \mu$
- $V(\overline{X}_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \left(\frac{1}{n}\right)^2 V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2$
  $= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$

Teorema do Limite Central (cont.)

Simulações:

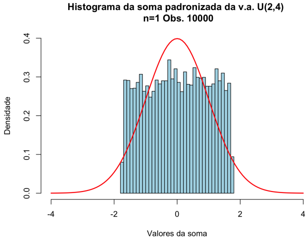
No R temos acesso as mais comuns distribuições univariadas. Todas as funções tem as seguintes formas:

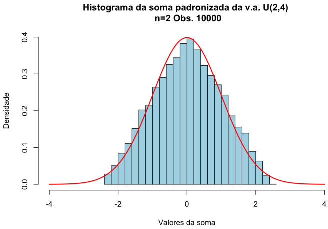
Função	Descrição
`pnome( ...)`	função de distribuição
`dnome( ...)`	função de probabilidade ou densidade de probabilidade
`qnome( ...)`	inversa da função de distribuição
`rnome( ...)`	geração de números aleatórios com a distribuição especificada

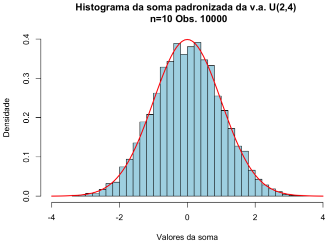
o **nome** é uma abreviatura do nome usual da distribuição (`binom`, `geom`, `pois`, `unif`, `exp`, `norm`, . . .)

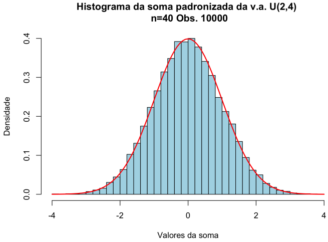
Ciências ULisboa

Teorema do Limite Central (cont.)



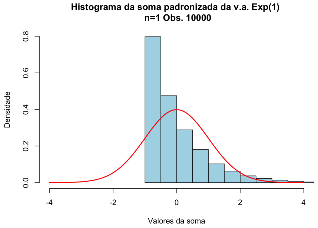


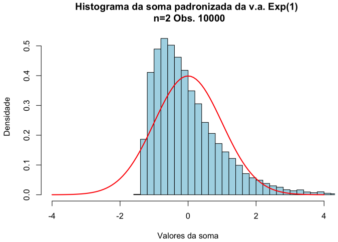


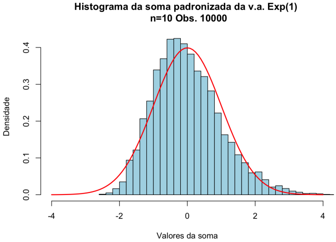


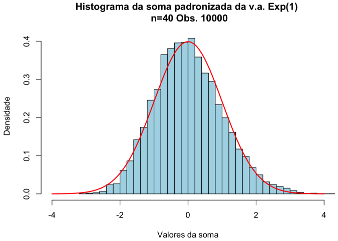
Ciências ULisboa

Teorema do Limite Central (cont.)









Teorema do Limite Central (cont.)

Aplicação do TLC para distribuições discretas

A aplicação do Teorema do Limite Central para distribuições discretas apresenta um desafio, pois envolve a aproximação de um fenómeno discreto por meio de uma distribuição contínua. Embora não haja uma solução perfeita para todas as situações, é comum, na prática, utilizar a correção de continuidade para melhorar a precisão da aproximação. Seja $\tilde{X}$ a variável aleatória normal que aproxima X. Assim

- $P(X = a) \simeq P(a - 0.5 < \tilde{X} < a + 0.5)$
- $P(a \leq X \leq b) \simeq P(a - 0.5 < \tilde{X} < b + 0.5)$
- $P(X \leq b) \simeq P(\tilde{X} < b + 0.5)$
- $P(X \geq a) \simeq P(\tilde{X} > a - 0.5)$

Teorema do Limite Central (cont.)

Aproximação da distribuição binomial pela distribuição normal

Sejam $X_1, X_2, \ldots X_n$ variáveis aleatórias independentes e identicamente distribuídas com $X_i \sim \text{Bernoulli}(p)$. Sabe-se que $X = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p)$. Assim, usando o TLC, tem-se que se $X \sim \text{Binomial}(n, p)$, com n suficientemente grande ($n \geq 30$), então

$$X \sim^{aprox.} N\left(np, \sqrt{np(1-p)}\right) \Leftrightarrow Z = \frac{X - np}{\sqrt{np(1-p)}} \sim^{aprox.} N(0, 1).$$

- Na prática, a aproximação é satisfatória quando $np > 5$ e $n(1-p) > 5$.

Teorema do Limite Central (cont.)

Aproximação da distribuição de Poisson pela distribuição normal

Sejam $X_1, X_2, \ldots X_n$ variáveis aleatórias independentes e identicamente distribuídas com $X_i \sim \text{Poisson}\left(\frac{\lambda}{n}\right)$. Sabe-se que $X = \sum_{i=1}^{n} X_i \sim \text{Poisson}(\lambda)$. Assim, usando o TLC, tem-se que se $X \sim \text{Poisson}(\lambda)$, com n suficientemente grande ($n \geq 30$), então

$$X \sim^{aprox.} N\left(\lambda, \sqrt{\lambda}\right) \Leftrightarrow Z = \frac{X - \lambda}{\sqrt{\lambda}} \sim^{aprox.} N(0,1).$$

- Na prática, a aproximação é satisfatória quando $\lambda > 5$.

Teorema do Limite Central (cont.)

Exemplo 1: O número de tremores de terra de baixa intensidade registados a partir de uma determinada região geológica durante uma semana tem distribuição de Poisson com parâmetro 300. Calcule a probabilidade de ocorrerem mais de 350 tremores de terra na próxima semana.

Resolução:

Variável aleatória de interesse

$X-$ "número de tremores de terra registados durante uma semana"
$X \sim \text{Poisson}(\lambda = 300)$

Probabilidade pedida

Como o valor de λ elevado não está presente nas tabelas disponíveis, e não é possível calcular a probabilidade manualmente sem o auxílio de um computador ou calculadora, podemos recorrer à distribuição normal para obter uma aproximação. Do TLC sabemos que $X \sim \text{Poisson}(300)$ pode ser aproximada por $\tilde{X} \sim N(300, \sqrt{300})$. Assim,

Teorema do Limite Central (cont.)

Resolução (cont.):

$$
\begin{aligned}
P(X > 350) &= P(X > 350 - 0.5) \\
&= 1 - P(X \leq 349.5) \\
&\simeq 1 - P(\tilde{X} < 349.5) \\
&= 1 - P\left(\frac{\tilde{X} - 300}{\sqrt{300}} < \frac{349.5 - 300}{\sqrt{300}}\right) \\
&= 1 - \Phi(2.86) \\
&= 1 - 0.9979 \\
&\simeq 0.0021
\end{aligned}
$$

Exemplo 2: O tempo (em horas) que João Pestana dorme por noite é uma variável aleatória com distribuição uniforme no intervalo (7,12).

(a) Calcule a probabilidade de João Pestana dormir mais de 11 horas numa noite.

(b) Qual a probabilidade de João Pestana dormir mais de 1100 horas em 100 noites?

(c) Qual a probabilidade do tempo médio em 100 noites ser inferior à 11 horas?

Resolução:
(a)

Variável aleatória de interesse

$X-$ "tempo (em horas) que João Pestana dorme por noite"
$X \sim \text{Uniforme}(7, 12)$

Função densidade de probabilidade

$$f_X(x) = \begin{cases} \frac{1}{5}, & 7 < x < 12 \\ 0, & \text{outros valores de x} \end{cases}$$

Resolução (cont.):

Função de distribuição

$$F_X(x) = \begin{cases} 0, & x \leq 7 \\ \frac{x-7}{5}, & 7 < x < 12 \\ 1, & x \geq 12 \end{cases}$$

Probabilidade pedida

$P(X > 11) = 1 - P(X \leq 11) = 1 - F_X(11) = 1 - \frac{11-7}{5} = 1 - \frac{4}{5} = \frac{1}{5} = 0.2$

(b)

Variável aleatória auxiliar

$X_i -$ "tempo (em horas) que JP dorme na i-ésima noite", $i = 1, 2, \ldots, 100$

$X_i \sim^{iid} X$

$E(X_i) = E(X) = \frac{7+12}{2} = \frac{19}{2} = 9.5$

$V(X_i) = V(X) = \frac{(12-7)^2}{12} = 2.0833$

Teorema do Limite Central (cont.)

Resolução (cont.):

Variável aleatória de interesse

$$S_{100} = \sum_{i=1}^{100} X_i = \text{“tempo (em horas) que JP dorme em 100 noites”}$$

$$E(S_{100}) = E\left(\sum_{i=1}^{100} X_i\right) = \sum_{i=1}^{100} E(X_i) = \sum_{i=1}^{100} 9.5 = 100 \times 9.5 = 950$$

$$V(S_{100}) = V\left(\sum_{i=1}^{100} X_i\right) = \sum_{i=1}^{100} V(X_i) = \sum_{i=1}^{100} 2.0833 = 100 \times 2.0833 = 208.33$$

Distribuição aproximada de S_{100}

Pelo TLC temos que

$$\frac{S_{100} - E(S_{100})}{\sqrt{V(S_{100})}} = \frac{S_{100} - 950}{\sqrt{208.33}} \sim^{aprox.} N(0,1)$$

Teorema do Limite Central (cont.)

Resolução (cont.):

Probabilidade pedida

$$
\begin{aligned}
P(S_{100} > 1100) &= 1 - P(S_{100} \leq 1100) \\
&= 1 - P\left(\frac{S_{100} - 950}{14.43} \leq \frac{1100 - 950}{14.43}\right) \\
&\overset{TLC}{\simeq} 1 - \Phi\left(\frac{1100 - 950}{14.43}\right) \\
&= 1 - \Phi(10.39) \\
&\simeq 1 - 1 \\
&= 0
\end{aligned}
$$

Resolução (cont.):

(c)

Variável aleatória de interesse

$\overline{X} = \frac{1}{100}\sum_{i=1}^{100} X_i =$ "tempo médio (em horas) que JP dorme em 100 noites"

$$E(\overline{X}) = E\left(\frac{1}{100}\sum_{i=1}^{100} X_i\right) = \frac{1}{100}\sum_{i=1}^{100} E(X_i) = \frac{100 \times 9.5}{100} = 9.5$$

$$V(\overline{X}) = V\left(\frac{1}{100}\sum_{i=1}^{100} X_i\right) = \left(\frac{1}{100}\right)^2 \sum_{i=1}^{100} V(X_i) = \frac{100 \times 2.0833}{100^2} = 0.020833$$

Distribuição aproximada de $\overline{X}$

Pelo TLC temos que

$$\frac{\overline{X} - E(\overline{X})}{\sqrt{V(\overline{X})}} = \frac{\overline{X} - 9.5}{\sqrt{0.020833}} = \frac{\overline{X} - 9.5}{0.14433} \sim^{aprox.} N(0,1)$$

Resolução (cont.):

Probabilidade pedida

$$P(\overline{X} < 11) = P\left(\frac{\overline{X} - 9.5}{0.14433} < \frac{11 - 9.5}{0.14433}\right) \overset{TLC}{\simeq} \Phi(10.39) \simeq 1$$

Referências

- Manuel Cabral Morais (2020): Probabilidade e Estatística: Teoria, Exemplos & Exercícios. IST Press, 1a edição.