Teorema do Limite Central

Bacharelado em Economia - FEA - Noturno

$1^{\underline{o}}$ Semestre 2014

Sumário

Objetivos da Aula

Soma de Variáveis Aleatórias

O objetivo principal desta aula é estudar empiricamente a distribuição da soma de variáveis aleatórias quantitativas e enunciar o principal teorema da Estatística Teorema do Limite Central.

Notação

Soma de Variáveis Aleatórias

Vamos supor $X_1, \ldots, X_n$ variáveis aleatórias independentes com mesma distribuição de média μ e variância σ^2 finitas. Vamos estudar a distribuição da soma

$$X = X_1 + \cdots + X_n$$

à medida que n cresce. Ou seja, vamos construir histogramas para a distribuição de X para diferentes valores de n.

Sumário

Distribuição Binomial

Distribuição Binomial

A distribuição binomial pode ser obtida através de n ensaios independentes de Bernoulli. Isto é, se $X_i \sim \mathrm{Be}(p)$ $(i = 1, \ldots, n)$, então

$$X = X_1 + \cdots + X_n \sim \mathrm{B}(n, p).$$

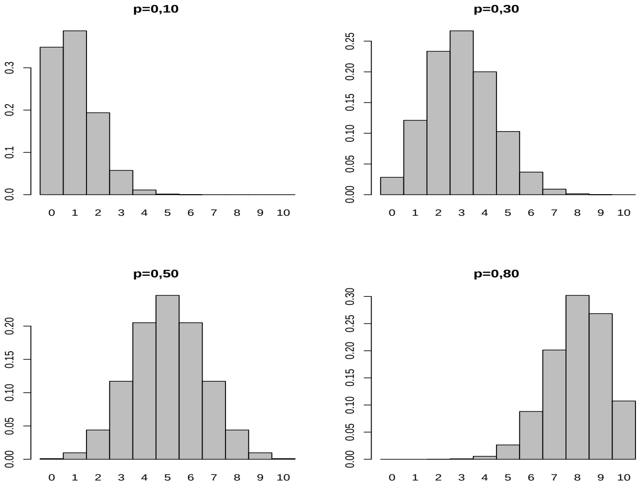
Temos ainda que $\mathrm{E}(X) = np$ e $\mathrm{Var}(X) = np(1 - p)$.

Histogramas Distribuição Binomial

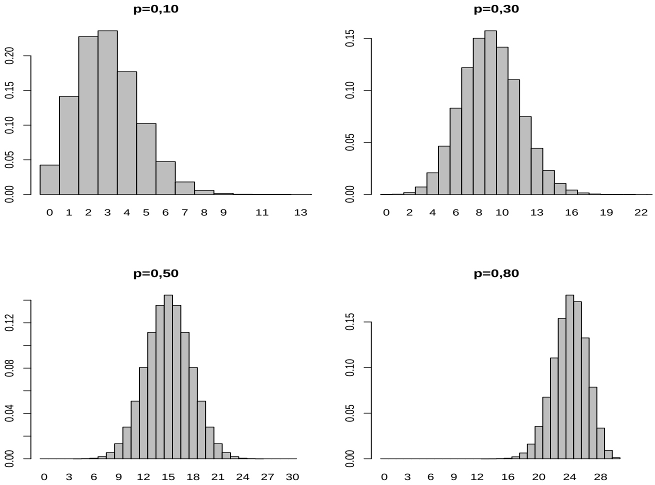
Descrição

A seguir serão construídos histogramas para a distribuição de $X \sim \mathrm{B}(n, p)$ variando-se o número de ensaios n e também a probabilidade de sucesso p.

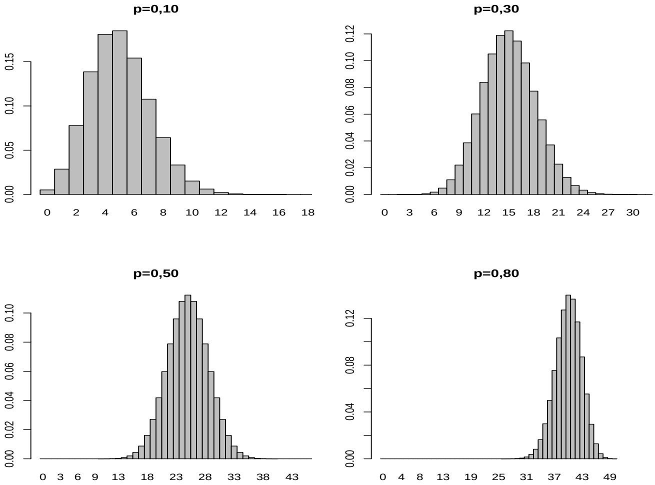
Histogramas B(n, p) para $n = 10$



Histogramas B(n, p) para $n = 30$



Histogramas B(n, p) para $n = 50$



Conclusões

Conclusões

Nota-se pelos gráficos que à medida que n cresce a distribuição de $X \sim \mathrm{B}(n, p)$ se aproxima da distribuição de $Y \sim \mathrm{N}(\mu_X, \sigma_X^2)$ em que $\mu_X = np$ e $\sigma_X^2 = np(1-p)$.

Sumário

Distribuição de Poisson

Definição

Se X segue distribuição de Poisson de parâmetro λ. Isto é, se $X \sim \mathrm{P}(\lambda)$, então a função de probabilidade de X fica dada por

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!},$$

em que $x = 0, 1, \ldots$. Temos ainda que $\mathrm{E}(X) = \lambda$ e $\mathrm{Var}(X) = \lambda$.

Histogramas Distribuição de Poisson

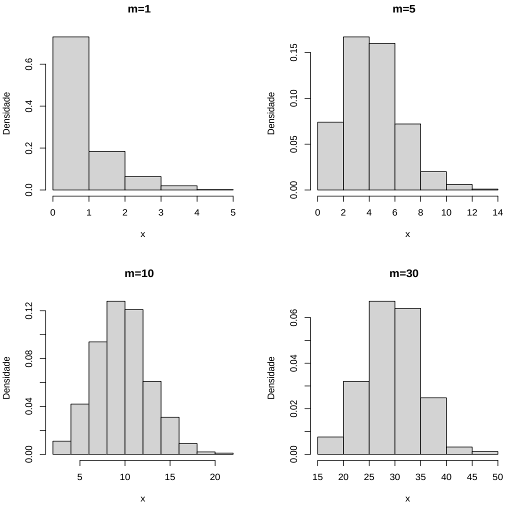
Descrição

A seguir serão construídos histogramas para a distribuição de

$$X = X_1 + \cdots + X_n \sim \mathrm{P}(n\lambda),$$

variando-se $m = n\lambda$, em que $X_i \sim \mathrm{P}(\lambda)$ independentes ($i = 1, \ldots, n$).

Histogramas P(m)



Conclusões

Conclusões

Nota-se pelos gráficos que à medida que *m* cresce a distribuição de $X \sim \mathrm{P}(m)$ se aproxima da distribuição de $Y \sim \mathrm{N}(\mu_X, \sigma_X^2)$ em que $\mu_X = m$ e $\sigma_X^2 = m$.

Sumário

Distribuição Uniforme

Definição

Vamos supor que X é uma variável aleatória com distribuição uniforme no intervalo [a,b] ($X \sim \mathrm{U}[a, b]$), então

$$f(x) = \frac{1}{(b-a)}, \quad a \leq x \leq b,$$

e $f(x) = 0$ em caso contrário.

Distribuição Uniforme

Definição

Vamos supor que X é uma variável aleatória com distribuição uniforme no intervalo [a,b] ($X \sim \mathrm{U}[a,b]$), então

$$f(x) = \frac{1}{(b-a)}, \quad a \leq x \leq b,$$

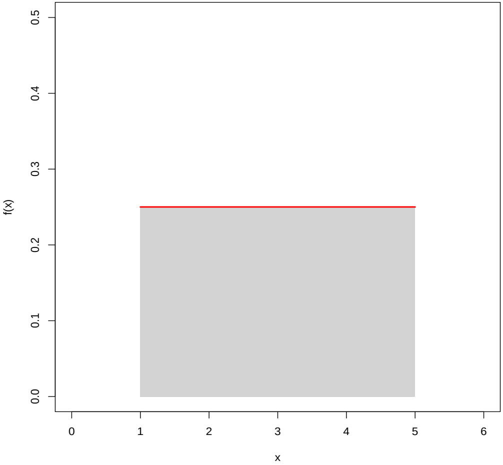
e $f(x) = 0$ em caso contrário.

Esperança e Variância

Temos que

$$\mathrm{E}(X) = \frac{a+b}{2} \text{ e } \mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

Distribuição Uniforme U$[1, 5]$

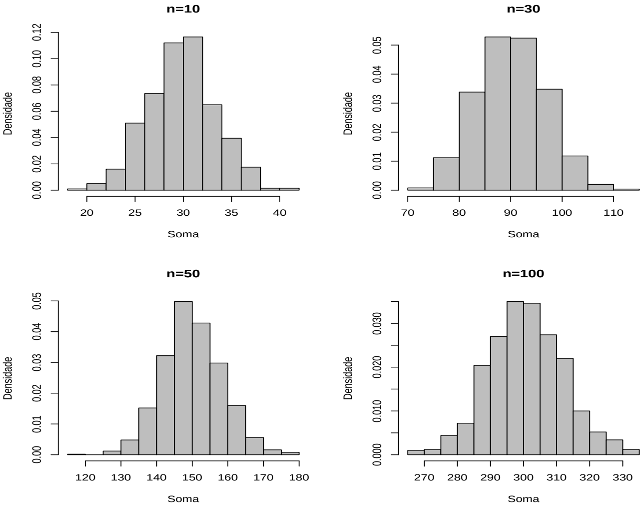


Histogramas Distribuição Uniforme

Descrição

Vamos supor que $X_i \sim \mathrm{U}[1, 5]$ independentes ($i = 1, \ldots, n$). A seguir serão construídos histogramas para a distribuição de $X = X_1 + \ldots + X_n$ variando-se o tamanho amostral n.

Histogramas Soma de Uniformes



Conclusões

Conclusões

Nota-se pelos gráficos que à medida que n cresce a distribuição de

$$X = X_1 + \cdots + X_n,$$

se aproxima da distribuição de $Y \sim \mathrm{N}(\mu_X, \sigma_X^2)$ em que

- $\mu_X = \frac{n(1+4)}{2} = \frac{5n}{2}$
- $\sigma_X^2 = \frac{n(4-1)^2}{12} = \frac{9n}{12}$

Sumário

Distribuição Exponencial

Definição

Se X é uma variável aleatória com distribuição exponencial de parâmetro $\lambda > 0$, a função densidade de probabilidade de X é definida por

$$f(x) = \lambda e^{-\lambda x},$$

em que $x > 0$. Notação $X \sim \text{Exp}(\lambda)$.

Distribuição Exponencial

Definição

Se X é uma variável aleatória com distribuição exponencial de parâmetro $\lambda > 0$, a função densidade de probabilidade de X é definida por

$$f(x) = \lambda e^{-\lambda x},$$

em que $x > 0$. Notação $X \sim \text{Exp}(\lambda)$.

Esperança e Variância

Temos que

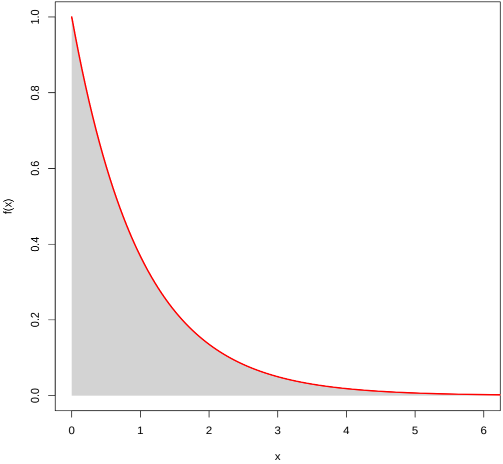
$$\text{E}(X) = \frac{1}{\lambda} \text{ e } \text{Var}(X) = \frac{1}{\lambda^2}.$$

Histogramas Distribuição Exponencial

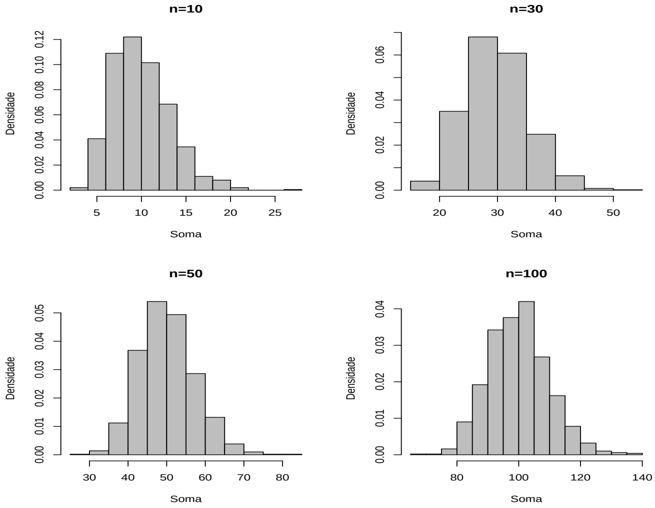
Definição

Vamos supor que $X_i \sim \text{Exp}(\lambda)$ independentes $(i = 1, \ldots, n)$. A seguir serão construídos histogramas para a distribuição de $X = X_1 + \ldots + X_n$ variando-se λ e o tamanho amostral n.

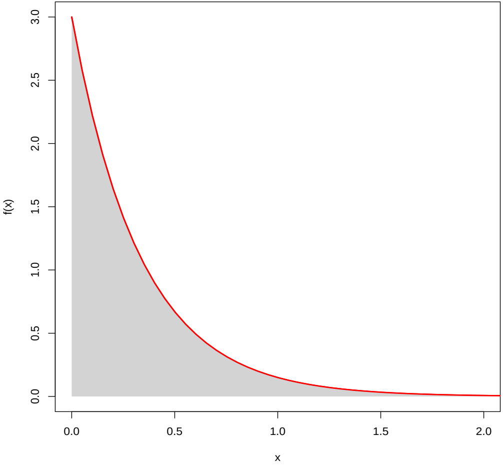
Distribuição Exponencail $\lambda = 1$



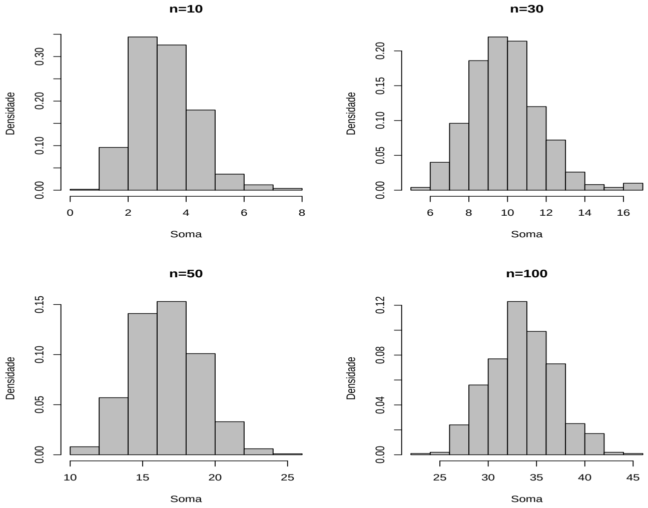
Histogramas Soma de Exponenciais com $\lambda = 1$



Distribuição Exponencial $\lambda = 3$



Histogramas Soma de Exponenciais com $\lambda = 3$



Conclusões

Conclusões

Nota-se pelos gráficos que à medida que n cresce a distribuição de

$$X = X_1 + \cdots + X_n,$$

se aproxima da distribuição de $Y \sim \mathrm{N}(\mu_X, \sigma_X^2)$ em que

- $\mu_X = n\frac{1}{\lambda} = \frac{n}{\lambda}$
- $\sigma_X^2 = n\frac{1}{\lambda^2} = \frac{n}{\lambda^2}$

Sumário

Teorema do Limite Central

Enunciado para a Soma Amostral

Para variáveis aleatórias $X_1, \ldots, X_n$ independentes e com mesma distribuição de média μ e variância σ^2 finitas, a distribuição da soma

$$X = X_1 + \cdots + X_n$$

se aproxima à medida que n cresce da distribuição de $Y \sim \mathrm{N}(\mu_X, \sigma_X^2)$, em que $\mu_X = n\mu$ e $\sigma_X^2 = n\sigma^2$.

Teorema do Limite Central

Aproximação para n Grande

$$
\begin{aligned}
P(a \leq X \leq b) &\cong P(a \leq Y \leq b) \\
&= P\left(\frac{a - n\mu}{\sigma\sqrt{n}} \leq Z \leq \frac{b - n\mu}{\sigma\sqrt{n}}\right),
\end{aligned}
$$

em que $Z \sim \mathrm{N}(0, 1)$.

Teorema do Limite Central

Média Amostral

Para a média amostral $\bar{X} = \frac{X_1 + \cdots + X_n}{n}$ temos que

$$\begin{aligned} E(\bar{X}) &= \frac{\mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n)}{n} \\ &= \frac{n\mu}{n} = \mu \ \text{ e} \\ \mathrm{Var}(\bar{X}) &= \frac{\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)}{n^2} \\ &= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}. \end{aligned}$$

Teorema do Limite Central

Enunciado para a Média Amostral

Para variáveis aleatórias $X_1, \ldots, X_n$ independentes e com mesma distribuição de média μ e variância σ^2 finitas, a distribuição da média amostral

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$$

se aproxima à medida que n cresce da distribuição de $Y \sim \mathrm{N}(\mu_{\bar{X}}, \sigma^2_{\bar{X}})$, em que $\mu_{\bar{X}} = \mu$ e $\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$.

Teorema do Limite Central

Aproximação para n Grande

$$
\begin{aligned}
P(a \leq \bar{X} \leq b) &\cong P(a \leq Y \leq b) \\
&= P\left(\frac{a-\mu}{\sigma/\sqrt{n}} \leq Z \leq \frac{b-\mu}{\sigma/\sqrt{n}}\right),
\end{aligned}
$$

em que $Z \sim \mathrm{N}(0,1)$.

Sumário

Exemplo 1

Exemplo 1

Uma loja recebe em média 16 clientes por dia com desvio padrão de 4 clientes. Calcule aproximadamente a probabilidade de num período de 30 dias a loja receber mais do que 500 clientes. Calcule também a probabilidade aproximada de nesse mesmo período a média de clientes ultrapassar a 18 clientes.

Exemplo 1

Dados do Problema

Seja U:número de clientes que a loja recebe num dia. Temos que

$$\begin{aligned} \mathrm{E}(U) &= \mu = 16 \\ \mathrm{Var}(U) &= \sigma^2 = 4^2 = 16. \end{aligned}$$

Exemplo 1

Dados do Problema

Seja U:número de clientes que a loja recebe num dia. Temos que

$$\begin{aligned} \mathrm{E}(U) &= \mu = 16 \\ \mathrm{Var}(U) &= \sigma^2 = 4^2 = 16. \end{aligned}$$

Soma Amostral

Seja X:número de clientes que a loja recebe em 30 dias. Temos que

$$\begin{aligned} \mu_X &= n \times \mu = 30 \times 16 = 480 \\ \sigma_X^2 &= n \times \sigma^2 = 30 \times 16 = 480 \\ \sigma_X &= \sqrt{480} \cong 21,91. \end{aligned}$$

Exemplo 1

Média Amostral

Seja $\bar{X}$:número médio de clientes que a loja recebe em 30 dias.
Temos que

$$\begin{aligned} \mu_{\bar{X}} &= \mu = 16 \\ \sigma^2_{\bar{X}} &= \frac{\sigma^2}{n} = \frac{16}{30} \cong 0,533 \\ \sigma_{\bar{X}} &= \sqrt{0,533} \cong 0,73. \end{aligned}$$

Exemplo 1

A probabilidade da loja receber mais do que 500 clientes em 30 dias fica dada por

$$
\begin{aligned}
P(X \geq 500) &\cong P\left(Z \geq \frac{500 - \mu_X}{\sigma_X}\right) \\
&= P\left(Z \geq \frac{500 - 480}{21,91}\right) \\
&= P(Z \geq 0,91) \\
&= 1 - P(Z \leq 0,91) \\
&= 1 - A(0,91) \\
&= 1 - 0,8186 \\
&= 0,1814 (18,14\%).
\end{aligned}
$$

Exemplo 1

A probabilidade da média de clientes ultrapassar 18 clientes em 30 fica dada por

$$
\begin{aligned}
P(\bar{X} \geq 18) &\cong P\left(Z \geq \frac{18 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) \\
&= P\left(Z \geq \frac{18 - 16}{0,73}\right) \\
&= P(Z \geq 2,74) \\
&= 1 - P(Z \leq 2,74) \\
&= 1 - A(2,74) \\
&= 1 - 0.9969 \\
&= 0,0031(0,31\%).
\end{aligned}
$$

Exemplo 2

Exemplo 2

Sabe-se que numa corrida de revesamento de 42 km com 8 atletas (cada um correndo 5,25 km) o tempo que cada atleta demora para completar o percurso tem distribuição aproximadamente normal de média 30 minutos e desvio padrão de 8 minutos. Se 8 atletas são escolhidos ao acaso para um prova, qual a probabilidade da equipe completar o percurso em menos de 3 horas? E em mais de 4 horas? Qual é tempo que apenas 5% das equipes farão abaixo dele?

Exemplo 2

Dados do Problema

Seja T:tempo que um atleta demora para completar o percurso.
Temos que

$$\begin{aligned} \mathrm{E}(T) &= \mu = 30 \\ \mathrm{Var}(T) &= \sigma^2 = 8^2 = 64. \end{aligned}$$

Exemplo 2

Dados do Problema

Seja T:tempo que um atleta demora para completar o percurso. Temos que

$$\begin{aligned} \mathrm{E}(T) &= \mu = 30 \\ \mathrm{Var}(T) &= \sigma^2 = 8^2 = 64. \end{aligned}$$

Soma Amostral

Seja X:tempo que a equipe (de 8 atletas) demora para completar o percurso. Temos que

$$\begin{aligned} \mu_X &= n \times \mu = 8 \times 30 = 240 \\ \sigma_X^2 &= n \times \sigma^2 = 8 \times 64 = 512 \\ \sigma_X &= \sqrt{512} \cong 22,63. \end{aligned}$$

Exemplo 2

A probabilidade da equipe completar o percurso em menos de 3 horas (180 minutos) fica dada por

$$\begin{aligned} P(X < 180) &= P\left(Z < \frac{180 - \mu_X}{\sigma_X}\right) \\ &= P\left(Z < \frac{180 - 240}{22,63}\right) \\ &= P(Z < -2,65) \\ &= P(Z > 2,65) \\ &= 1 - P(z \leq 2,65) \\ &= 1 - 0,996 \\ &= 0,004 (0,4\%). \end{aligned}$$

Exemplo 2

A probabilidade da equipe completar o percurso em mais de 4 horas (240 minutos) fica dada por

$$\begin{aligned} P(X > 240) &= P\left(Z > \frac{240 - \mu_X}{\sigma_X}\right) \\ &= P\left(Z > \frac{240 - 240}{22,63}\right) \\ &= P(Z > 0) \\ &= 0,5(50\%). \end{aligned}$$

Exemplo 2

Seja t_0 o tempo superado por 95% das equipes (apenas 5% das equipes fazem abaixo desse tempo). Temos que

$$\begin{aligned} P(X < t_0) &= P\left(Z < \frac{t_0 - \mu_X}{\sigma_X}\right) \\ &= P\left(Z < \frac{t_0 - 240}{22,63}\right) \\ &= P(Z < a) = 0,05, \end{aligned}$$

em que $a = (t_0 - 240)/22,63$. Pela tabela normal $a = -1,64$. Assim, obtemos $t_0 = 240 - 1,64 \times 22,63 \cong 203$ minutos.