Probabilidade e Estatística
Tema: Variáveis Aleatórias Contínuas

Profa. Luciana Graziela de Godoi
Departamento de Estatística
CCE-UFES

Material desenvolvido no projeto Elaboração de Material Didático para o Ensino da Estatística na UFES. Elaborado em parceria com o Prof. Dr. Alessandro José Queiroz Sarnaglia.

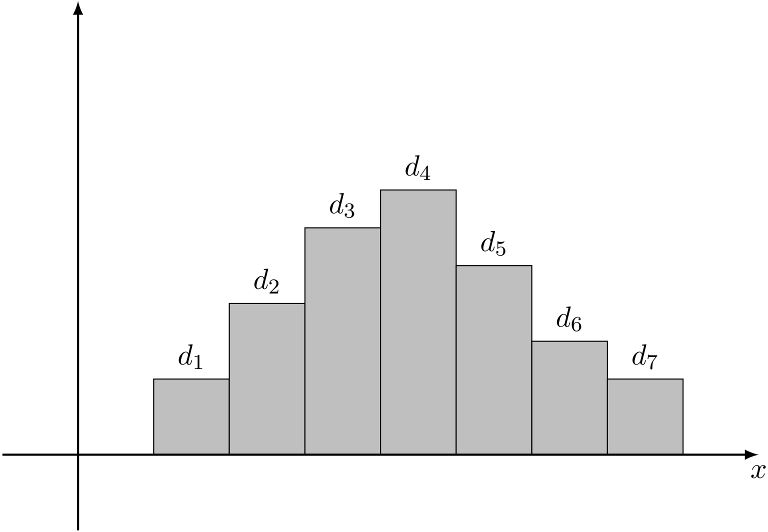
Apoio: Programa de aprimoramento e desenvolvimento do ensino (PRÓ-ENSINO).

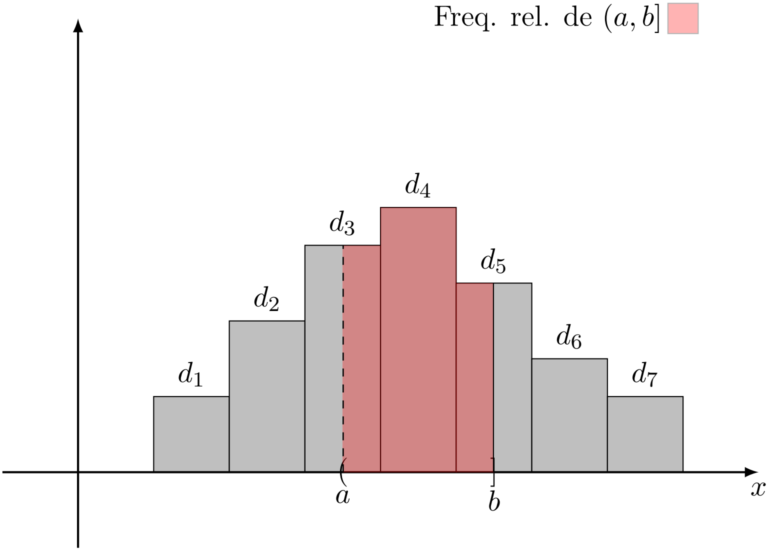
Variáveis aleatórias contínuas

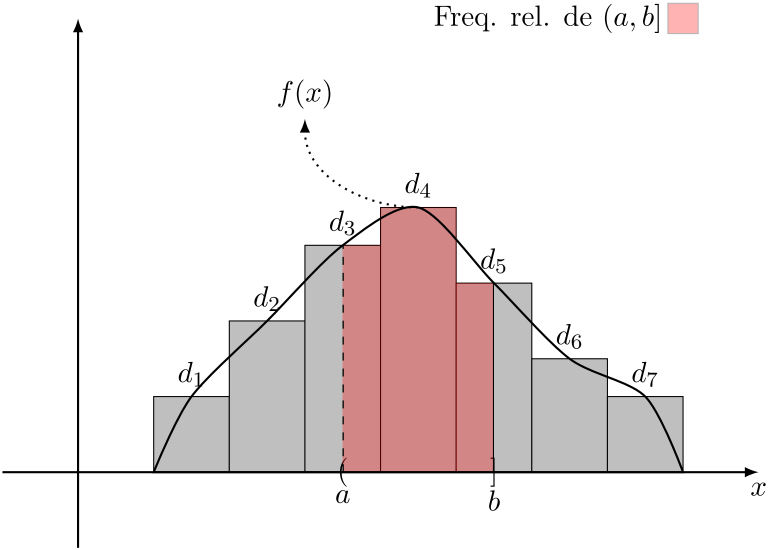
Suponha que X é uma v.a. contínua. Devido a natureza não contável de $Im(X)$ neste caso, não devemos estabelecer probabilidades "ponto-a-ponto" conforme no caso discreto.

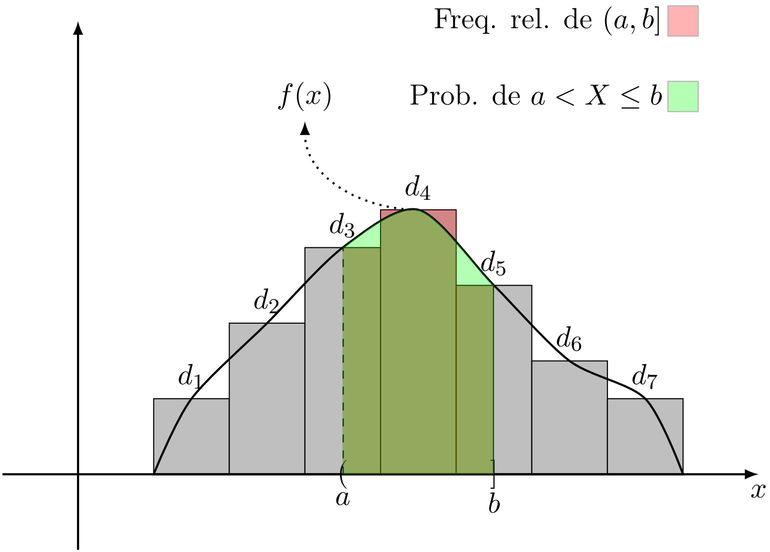
Caso discreto $\dashrightarrow$ Um modelo para as frequências relativas em cada ponto de $Im(X)$

Caso contínuo $\dashrightarrow$ Um modelo para as densidades


d_1
d_2
d_3
d_4
d_5
d_6
d_7
x


Freq. rel. de $(a, b]$
d_1
d_2
d_3
d_4
d_5
d_6
d_7
a
b
x


Freq. rel. de $(a, b]$
$f(x)$
d_1
d_2
d_3
d_4
d_5
d_6
d_7
a
b
x



Freq. rel. de $(a, b]$
Prob. de $a < X \leq b$
$f(x)$
d_4
d_3
d_5
d_2
d_6
d_1
d_7
a
b
x

Função densidade de probabilidade

Definição

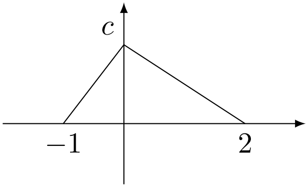
Seja $f(x)$ uma função, tal que:

- $f(x) \geq 0$, *para todo* $x \in \mathbb{R}$;
- *a área entre* $f(x)$ *e o eixo horizontal é igual a* 1.

Dizemos que $f(x)$ é função densidade de probabilidade (f.d.p.) de alguma v.a. contínua.

Exemplo: Seja

$$f(x) = \begin{cases} c(x+1), & -1 < x < 0; \\ c(1 - x/2), & 0 \leq x \leq 2; \\ 0, & \text{caso contrário.} \end{cases}$$



Para qual valor de c, $f(x)$ é densidade?

Função densidade de probabilidade

Definição

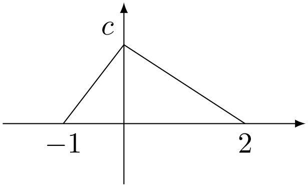
Seja $f(x)$ uma função, tal que:

- $f(x) \geq 0$, *para todo* $x \in \mathbb{R}$;
- *a área entre* $f(x)$ *e o eixo horizontal é igual a* 1.

Dizemos que $f(x)$ é função densidade de probabilidade (f.d.p.) de alguma v.a. contínua.

Exemplo: Seja

$$f(x) = \begin{cases} c(x+1), & -1 < x < 0; \\ c(1 - x/2), & 0 \leq x \leq 2; \\ 0, & \text{caso contrário.} \end{cases}$$



Para qual valor de c, $f(x)$ é densidade? Primeiramente, $c > 0$.

Função densidade de probabilidade

Definição

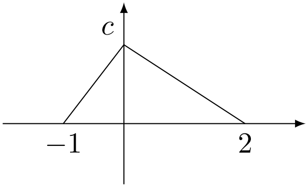
Seja $f(x)$ uma função, tal que:

- $f(x) \geq 0$, *para todo* $x \in \mathbb{R}$;
- *a área entre* $f(x)$ *e o eixo horizontal é igual a* 1.

Dizemos que $f(x)$ *é função densidade de probabilidade (f.d.p.) de alguma v.a. contínua.*

Exemplo: Seja

$$f(x) = \begin{cases} c(x+1), & -1 < x < 0; \\ c(1 - x/2), & 0 \leq x \leq 2; \\ 0, & \text{caso contrário.} \end{cases}$$



Para qual valor de c, $f(x)$ é densidade? Primeiramente, $c > 0$. A área abaixo de $f(x)$ por sua vez é $\frac{c}{2} + c = \frac{3c}{2}$.

Função densidade de probabilidade

Definição

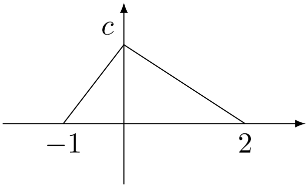
Seja $f(x)$ uma função, tal que:

- $f(x) \geq 0$, *para todo* $x \in \mathbb{R}$;
- *a área entre* $f(x)$ *e o eixo horizontal é igual a* 1.

Dizemos que $f(x)$ *é função densidade de probabilidade (f.d.p.) de alguma v.a. contínua.*

Exemplo: Seja

$$f(x) = \begin{cases} c(x+1), & -1 < x < 0; \\ c(1 - x/2), & 0 \leq x \leq 2; \\ 0, & \text{caso contrário.} \end{cases}$$



Para qual valor de c, $f(x)$ é densidade? Primeiramente, $c > 0$. A área abaixo de $f(x)$ por sua vez é $\frac{c}{2} + c = \frac{3c}{2}$. Logo, devemos ter $c = \frac{2}{3}$.

Definição formal

Como obter as probabilidades

Dada uma v.a. contínua X com função densidade de probabilidade $f(x)$. A probabilidade $P(a < X \leq b)$ é dada pela área entre a f.d.p. $f(x)$ e o eixo horizontal compreendida no intervalo $(a, b]$.
Por este motivo

$$P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) = P(a \leq X \leq b)$$

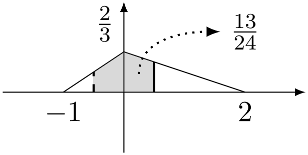
e, além disso,

$$P(X = x) = 0, \text{ para todo } x \in \mathbb{R}.$$

Definição formal

Exercício: Mostre que, se X tem a f.d.p. do exemplo anterior, então

$$P(-0.5 < X < 0.5) = \frac{13}{24}.$$



Função de distribuição cumulativa

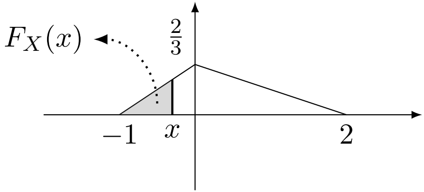
Definição

Seja X v.a. contínua com f.d.p. $f(x)$. Definimos a função de distribuição cumulativa de X por $F_X(x) = P(X \leq x)$.

Exemplo: Retornando ao exemplo anterior, temos que

$$F_X(x) = P(X \leq x) = \frac{(x+1)^2}{3},$$

se $-1 < x < 0$.



Definição

Seja X v.a. contínua com f.d.p. $f(x)$. Definimos a função de distribuição cumulativa de X por $F_X(x) = P(X \leq x)$.

Exemplo: Retornando ao exemplo anterior, temos que

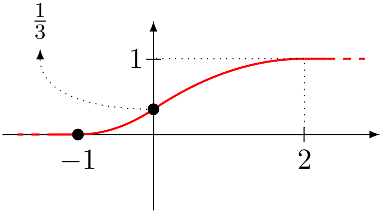
$$F_X(x) = P(X \leq x) = \frac{1}{3} + \frac{2x}{3} - \frac{x^2}{6},$$

se $0 \leq x < 2$.
Assim, a função de distribuição cumulativa (f.d.c.) é dada por

$$F_X(x) = \begin{cases} 0, & x \leq -1; \\ (x+1)^2/3, & -1 < x < 0; \\ \frac{1}{3} + \frac{2x}{3} - \frac{x^2}{6}, & 0 \leq x < 2; \\ 1, & x \geq 2. \end{cases}$$

Definição formal

Gráfico de $F_X(x)$:



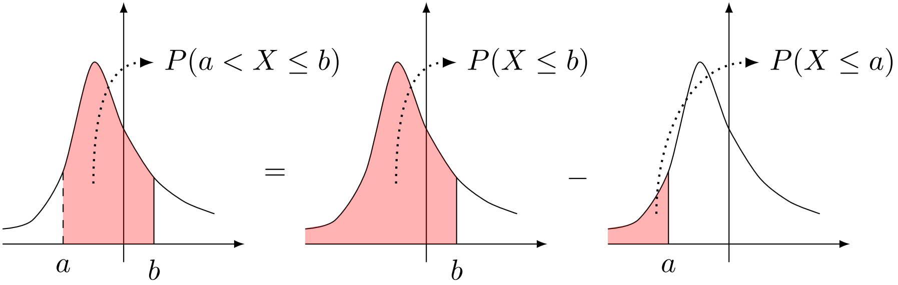
No caso contínuo, a função de distribuição cumulativa satisfaz as mesmas propriedades que no caso discreto. A saber:

- $F_X(x) \to 0$, quando $x \to -\infty$;
- $F_X(x) \to 1$, quando $x \to \infty$;
- $x \leq y \Rightarrow F_X(x) \leq F_X(y)$;
- $P(a < X \leq b) = F_X(b) - F_X(a)$.

Além disso, a f.d.c. de uma v.a. contínua é uma função contínua.
A esperança e a variância são definidas no caso contínuo através do uso de integrais.

Definição formal

A última propriedade enunciada anteriormente pode ser vista através da seguinte figura:



Logo,

$$P(a < X \leq b) = F_X(b) - F_X(a).$$

Valor Esperado e Variância

Definição

Seja X v.a. contínua com f.d.p. $f(x)$. Definimos a esperança e a variância de X por

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

e

$$\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx.$$

Todas as propriedades apresentadas para o caso discreto são válidas para o caso contínuo, em especial:

$$Var(X) = E(X^2) - \mu^2, \text{ em que } E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx.$$

Exemplo: Considere que a variável aleatória T represente o tempo para a execução de determinada tarefa em horas e cuja a função de densidade é dada por:

$$f(t) = \begin{cases} 2e^{-2t}, & t \geq 0; \\ 0, & t < 0. \end{cases}$$

Determine o tempo médio de execução e seu desvio-padrão.

Exemplo: Considere que a variável aleatória T represente o tempo para a execução de determinada tarefa em horas e cuja a função de densidade é dada por:

$$f(t) = \begin{cases} 2e^{-2t}, & t \geq 0; \\ 0, & t < 0. \end{cases}$$

Determine o tempo médio de execução e seu desvio-padrão.

Solução:

$$\begin{aligned} E(T) &= \int_{-\infty}^{\infty} t\ f(t)dt \\ &= \int_{-\infty}^{0} t \times 0\ dt + \int_{0}^{\infty} t \times 2e^{-2t}dt \\ &= \int_{0}^{\infty} 2t\ e^{-2t}dt. \end{aligned}$$

Integrando por partes, temos que

$$u = t \rightarrow u' = 1$$
$$v' = 2\ e^{-2t} \rightarrow v = -e^{-2t}$$

Então,

$$\begin{aligned}
\mu = E(T) &= \int_0^{\infty} 2t\ e^{-2t} dt \\
&= -te^{-2t}\Big|_0^{\infty} + \int_0^{\infty} e^{-2t} dt \\
&= 0 + \left(-\frac{e^{-2t}}{2}\right)\Big|_0^{\infty} \\
&= \frac{1}{2}.
\end{aligned}$$

Ou seja, o tempo médio de execução da tarefa é de meia hora.

Temos também que

$$Var(T) = E(T^2) - \mu^2, \text{ em que}$$

$$E(T^2) = \int_{-\infty}^{\infty} t^2 f(t)dt = \int_{-\infty}^{0} t^2 \times 0\ dt + \int_{0}^{\infty} t^2 \times 2e^{-2t}dt.$$

Integrando por parte duas vezes na última parcela, mostramos que

$$E(T^2) = \frac{1}{2}.$$

Logo, $Var(T) = E(T^2) - \mu^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}$.

Ou seja, $\sigma = DP(T) = \sqrt{\frac{1}{4}} = \frac{1}{2}$ e, sendo assim, o desvio-padrão do tempo de execução é igual a meia hora.

Exercício: Encontre a esperança e a variância da v.a. contínua com f.d.p.

$$f(x) = \begin{cases} \frac{2}{3}(x+1), & -1 < x < 0; \\ \frac{2}{3}(1 - x/2), & 0 \leq x \leq 2; \\ 0, & \text{caso contrário.} \end{cases}$$

Modelo uniforme contínuo

Definição

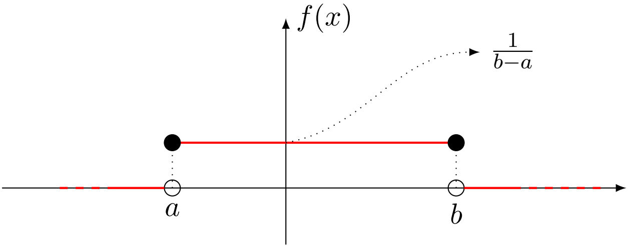
Seja X v.a. contínua com conjunto imagem $Im(X) = [a, b]$ e com f.d.p. dada por

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ \\ 0, & c.\ c. \end{cases}$$

Então, dizemos que X tem distribuição uniforme contínua no intervalo $[a, b]$. Notação: $X \sim U[a, b]$.

Note que não há restrição para os valores de a e b, exceto o fato de que $a < b$.

Gráfico da função de densidade



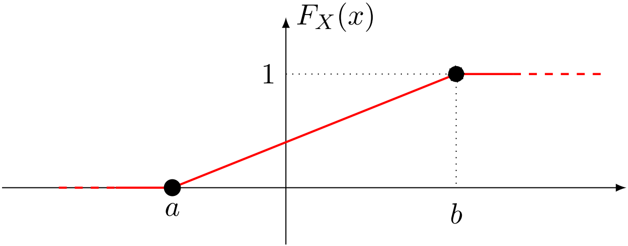
Interpretação: O modelo uniforme pressupõe que os valores possíveis para a variável aleatória tem todos a mesma probabilidade de ocorrência.

Função de distribuição acumulada

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\, dx = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a}, & a \leq x < b; \\ 1, & x \geq b. \end{cases}$$

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\, dx = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a}, & a \leq x < b; \\ 1, & x \geq b. \end{cases}$$

Gráfico da função de distribuição acumulada



Esperança e variância

Suponha que $X \sim U[a, b]$. É possível mostrar que a esperança e a variância de X são dadas, respectivamente, por

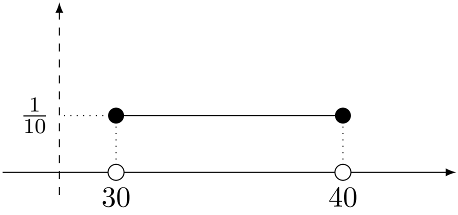
$$E(X) = \frac{a+b}{2} \quad \text{e} \quad Var(X) = \frac{(b-a)^2}{12}.$$

Exercício: Mostre, usando as definições de média e esperança para variáveis aleatórias contínuas, o resultado apresentado anteriormente quando $X \sim U[a, b]$.

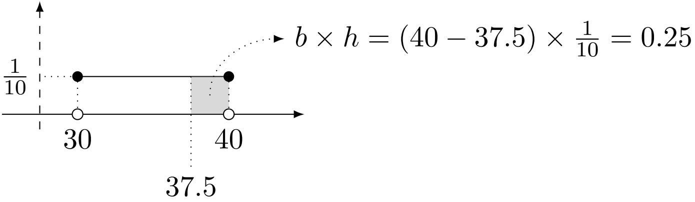
Exemplo: Suponha que o tempo em segundos requerido para completar uma operação de montagem seja $X \sim U[30, 40]$. Determinemos:

1. a proporção de operações que duram mais do que 37.5 segundos;
2. o tempo que é excedido por 90% das montagens;
3. a média e a variância da duração das montagens.

Gráfico da função de densidade:

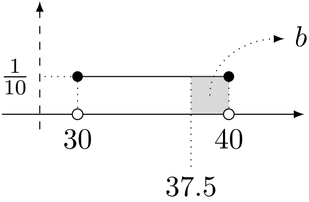


Em (1), temos que $P(X > 37.5)$, assim



Logo, $P(X > 37.5) = 0,25$.

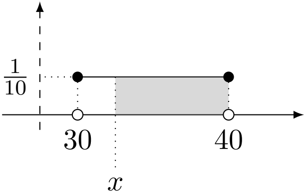
Em (1), temos que $P(X > 37.5)$, assim



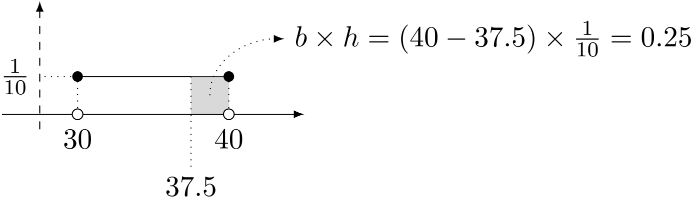
$b \times h = (40 - 37.5) \times \frac{1}{10} = 0.25$

Logo, $P(X > 37.5) = 0,25$.

Em (2) queremos encontrar x, tal que $P(X > x) = 0.9$. Note que

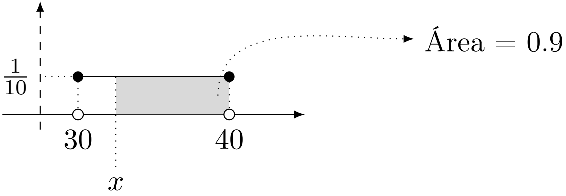


Em (1), temos que $P(X > 37.5)$, assim



Logo, $P(X > 37.5) = 0,25$.

Em (2) queremos encontrar x, tal que $P(X > x) = 0.9$. Note que



Assim,

$$\text{Área} = 0.9 \Leftrightarrow b \times h = 0.9 \Leftrightarrow (40 - x)\frac{1}{10} = 0.9 \Leftrightarrow x = 31.$$

Assim,

$$\text{Área} = 0.9 \Leftrightarrow b \times h = 0.9 \Leftrightarrow (40 - x)\frac{1}{10} = 0.9 \Leftrightarrow x = 31.$$

Uma outra maneira de fazer:

$$P(X > x) = 0.9 \quad \Leftrightarrow$$

Assim,

$$\text{Área} = 0.9 \Leftrightarrow b \times h = 0.9 \Leftrightarrow (40 - x)\frac{1}{10} = 0.9 \Leftrightarrow x = 31.$$

Uma outra maneira de fazer:

$$\begin{aligned} P(X > x) = 0.9 &\quad \Leftrightarrow \\ 1 - P(X \leq x) = 0.9 &\quad \Leftrightarrow \end{aligned}$$

Assim,

$$\text{Área} = 0.9 \Leftrightarrow b \times h = 0.9 \Leftrightarrow (40 - x)\frac{1}{10} = 0.9 \Leftrightarrow x = 31.$$

Uma outra maneira de fazer:

$$\begin{aligned} P(X > x) = 0.9 \quad &\Leftrightarrow \\ 1 - P(X \leq x) = 0.9 \quad &\Leftrightarrow \\ 1 - F_X(x) = 0.9 \quad &\Leftrightarrow \end{aligned}$$

Assim,

$$\text{Área} = 0.9 \Leftrightarrow b \times h = 0.9 \Leftrightarrow (40 - x)\frac{1}{10} = 0.9 \Leftrightarrow x = 31.$$

Uma outra maneira de fazer:

$$\begin{aligned} P(X > x) = 0.9 &\quad \Leftrightarrow \\ 1 - P(X \leq x) = 0.9 &\quad \Leftrightarrow \\ 1 - F_X(x) = 0.9 &\quad \Leftrightarrow \\ F_X(x) = 0.1 &\quad \Leftrightarrow \end{aligned}$$

Assim,

$$\text{Área} = 0.9 \Leftrightarrow b \times h = 0.9 \Leftrightarrow (40 - x)\frac{1}{10} = 0.9 \Leftrightarrow x = 31.$$

Uma outra maneira de fazer:

$$\begin{aligned} P(X > x) = 0.9 &\quad \Leftrightarrow \\ 1 - P(X \leq x) = 0.9 &\quad \Leftrightarrow \\ 1 - F_X(x) = 0.9 &\quad \Leftrightarrow \\ F_X(x) = 0.1 &\quad \Leftrightarrow \\ \frac{x - 30}{40 - 30} = 0.1 &\quad \Leftrightarrow \end{aligned}$$

Assim,

$$\text{Área} = 0.9 \Leftrightarrow b \times h = 0.9 \Leftrightarrow (40 - x)\frac{1}{10} = 0.9 \Leftrightarrow x = 31.$$

Uma outra maneira de fazer:

$$\begin{aligned}
P(X > x) = 0.9 &\quad \Leftrightarrow \\
1 - P(X \leq x) = 0.9 &\quad \Leftrightarrow \\
1 - F_X(x) = 0.9 &\quad \Leftrightarrow \\
F_X(x) = 0.1 &\quad \Leftrightarrow \\
\frac{x - 30}{40 - 30} = 0.1 &\quad \Leftrightarrow \\
x = 31.
\end{aligned}$$

Portanto, 90% das operações de montagem duram mais do que 31 segundos.

Assim,

$$\text{Área} = 0.9 \Leftrightarrow b \times h = 0.9 \Leftrightarrow (40 - x)\frac{1}{10} = 0.9 \Leftrightarrow x = 31.$$

Uma outra maneira de fazer:

$$\begin{aligned}
P(X > x) = 0.9 &\quad \Leftrightarrow \\
1 - P(X \leq x) = 0.9 &\quad \Leftrightarrow \\
1 - F_X(x) = 0.9 &\quad \Leftrightarrow \\
F_X(x) = 0.1 &\quad \Leftrightarrow \\
\frac{x - 30}{40 - 30} = 0.1 &\quad \Leftrightarrow \\
x = 31. &
\end{aligned}$$

Portanto, 90% das operações de montagem duram mais do que 31 segundos.

Em (3), basta utilizarmos as fórmulas de esperança e variância para esse modelo.

Logo, temos que

$$\mu = E(X) = \frac{30 + 40}{2} = 35$$

e

$$\sigma^2 = Var(X) = \frac{(40 - 30)^2}{12} = \frac{100}{12} = 8.33.$$

Modelo exponencial

Uma distribuição muito utilizada para representar o “tempo” (ou a distância) até a ocorrência de determinado evento geralmente é modelada pela variável aleatória exponencial.

Enquanto a distribuição de Poisson pode ser usada para modelar o número de ocorrências em um período contínuo (de tempo ou de comprimento, por exemplo), a distribuição exponencial pode modelar a variável aleatória contínua que representa o intervalo (de tempo ou de comprimento, por exemplo) entre ocorrências.

O modelo exponencial, assim como o modelo de Poisson, pressupõe que haja independência entre as ocorrências e que a taxa média seja constante no intervalo considerado.

Definição

Seja X v.a. contínua com conjunto imagem $Im(X) = [0, \infty)$. Suponha que a f.d.p de X seja

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0; \\ 0, & c.c., \end{cases}$$

onde $\lambda > 0$. Dizemos que X tem distribuição exponencial com parâmetro λ. Notação: $X \sim Exp(\lambda)$.

Aplicações: Tempo de vida útil de equipamentos, tempos de falha, tempos de sobrevivência entre as espécies e intervalos entre solicitação de recursos.

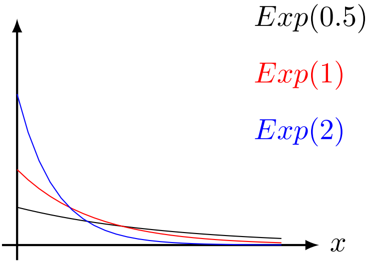
Esperança, variância e função de distribuição acumulada

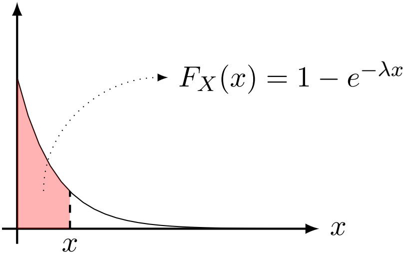
É possível mostrar que se $X \sim Exp(\lambda)$. Então,

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

Além disso, é possível mostrar que a função de distribuição acumulada de X é dada por

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0; \\ 0, & \text{c.c.} \end{cases}$$


$Exp(0.5)$
$Exp(1)$
$Exp(2)$
x


$F_X(x) = 1 - e^{-\lambda x}$
x
x

Exemplo

Seja X = "tempo em anos decorrido até a falha de determinado equipamento mecânico". Suponha que $X \sim Exp(\lambda)$ e que, em **média**, o equipamento demore 2 anos até falhar. Qual a probabilidade de esse equipamento não falhe antes de 3 anos?

Exemplo

Seja X = "tempo em anos decorrido até a falha de determinado equipamento mecânico". Suponha que $X \sim Exp(\lambda)$ e que, em **média**, o equipamento demore 2 anos até falhar. Qual a probabilidade de esse equipamento não falhe antes de 3 anos?

Temos que encontrar primeiramente o valor de λ.

Como $\mu = E(X) = \frac{1}{\lambda} = 2$, então $\lambda = 0.5$. Logo,

$$X \sim Exp(0.5).$$

Agora, temos que

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) = 1 - P(X \leq 3) = 1 - F_X(3) \\ &= 1 - (1 - e^{-0.5 \times 3}) = e^{-1.5} \approx 0.22. \end{aligned}$$

Importância:

- Permite modelar uma infinidade de fenômenos naturais,

Modelo normal (ou gaussiano)

Importância:

- Permite modelar uma infinidade de fenômenos naturais,
- possibilita realizar aproximações para calcular probabilidades de muitas variáveis aleatórias,

Importância:

- Permite modelar uma infinidade de fenômenos naturais,
- possibilita realizar aproximações para calcular probabilidades de muitas variáveis aleatórias,
- inferência estatística,

Modelo normal (ou gaussiano)

Importância:

- Permite modelar uma infinidade de fenômenos naturais,
- possibilita realizar aproximações para calcular probabilidades de muitas variáveis aleatórias,
- inferência estatística,
- grande aplicabilidade devido ao Teorema Central do Limite (TCL).

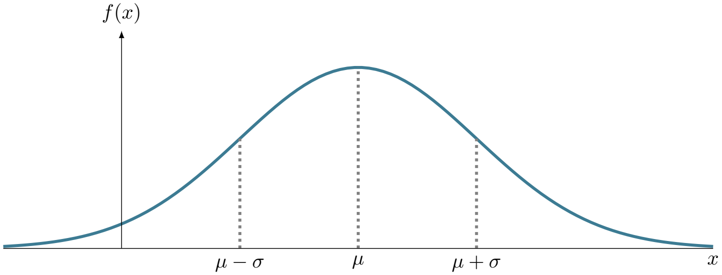
Modelo normal (ou gaussiano)

Definição

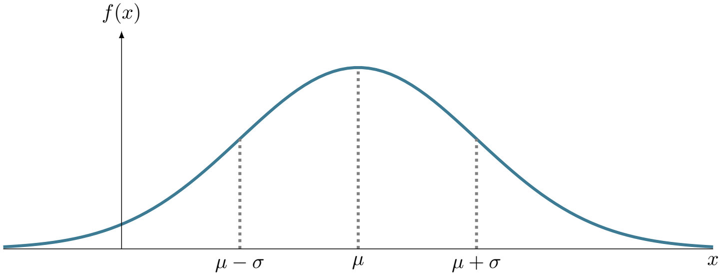
Dizemos que X *tem distribuição normal com parâmetros* μ *e* σ^2, *com* $-\infty < \mu < \infty$ *e* $\sigma^2 > 0$, *se sua f.d.p. é dada por*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

Notação: $X \sim N(\mu, \sigma^2)$.



Algumas observações relevantes:



Algumas observações relevantes:

- $f(x)$ é simétrica com relação a μ, isto é, $f(x-\mu) = f(x+\mu)$,
- $f(x) \to 0$, quando $x \to \overset{+}{-} \infty$,
- o valor máximo de x ocorre para $x = \mu$, isto é, $f(x) = \frac{1}{\sigma\sqrt{2\pi}}$,
- $\mu - \sigma$ e $\mu + \sigma$ são os pontos de inflexão da curva.

Proposição

Se $X \sim N(\mu, \sigma^2)$ *então* $E(X) = \mu$ *e* $Var(X) = \sigma^2$.

Proposição

Se $X \sim N(\mu, \sigma^2)$ *então* $E(X) = \mu$ *e* $Var(X) = \sigma^2$.

Observações: Considere $X_1 \sim N(\mu_1, \sigma_1^2)$ e $X_2 \sim N(\mu_2, \sigma_2^2)$ e obtenha os gráficos das funções de densidade quando

Proposição

Se $X \sim N(\mu, \sigma^2)$ *então* $E(X) = \mu$ *e* $Var(X) = \sigma^2$.

Observações: Considere $X_1 \sim N(\mu_1, \sigma_1^2)$ e $X_2 \sim N(\mu_2, \sigma_2^2)$ e obtenha os gráficos das funções de densidade quando

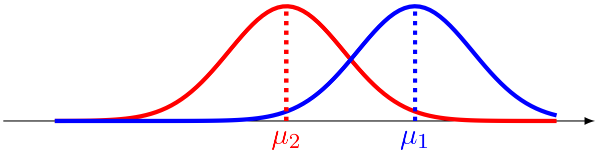
a) $\mu_1 \neq \mu_2$ e $\sigma_1^2 = \sigma_2^2$.

Proposição

Se $X \sim N(\mu, \sigma^2)$ *então* $E(X) = \mu$ *e* $Var(X) = \sigma^2$.

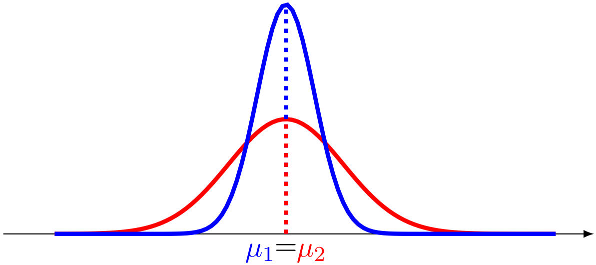
Observações: Considere $X_1 \sim N(\mu_1, \sigma_1^2)$ e $X_2 \sim N(\mu_2, \sigma_2^2)$ e obtenha os gráficos das funções de densidade quando

a) $\mu_1 \neq \mu_2$ e $\sigma_1^2 = \sigma_2^2$.

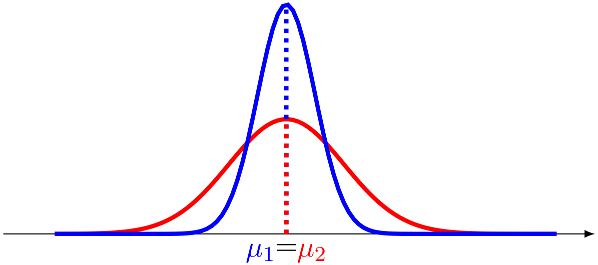


b) $\mu_1 = \mu_2$ e $\sigma_1^2 \neq \sigma_2^2$.

b) $\mu_1 = \mu_2$ e $\sigma_1^2 \neq \sigma_2^2$.



b) $\mu_1 = \mu_2$ e $\sigma_1^2 \neq \sigma_2^2$.



Pergunta: Analisando o último gráfico, qual das duas variáveis aleatórias X_1 e X_2 possui maior variabilidade. Por quê?

Observações

Quando $\mu = 0$ *e* $\sigma^2 = 1$ *temos a distribuição normal padrão, a qual denotamos por* $Z \sim N(0, 1)$.

Observações

Quando $\mu = 0$ *e* $\sigma^2 = 1$ *temos a distribuição normal padrão, a qual denotamos por* $Z \sim N(0, 1)$.

Se $X \sim N(\mu, \sigma^2)$, então uma nova variável aleatória $Z = \frac{X-\mu}{\sigma}$ tem as seguintes propriedades:

- $E(Z) = E\left(\frac{X-\mu}{\sigma}\right) =$

Observações

Quando $\mu = 0$ *e* $\sigma^2 = 1$ *temos a distribuição normal padrão, a qual denotamos por* $Z \sim N(0, 1)$.

Se $X \sim N(\mu, \sigma^2)$, então uma nova variável aleatória $Z = \frac{X-\mu}{\sigma}$ tem as seguintes propriedades:

- $E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E\left(X - \mu\right) = \frac{1}{\sigma}\left(E(X) - \mu\right) = \frac{1}{\sigma}\left(\mu - \mu\right) = 0.$

Observações

Quando $\mu = 0$ e $\sigma^2 = 1$ temos a distribuição normal padrão, a qual denotamos por $Z \sim N(0,1)$.

Se $X \sim N(\mu, \sigma^2)$, então uma nova variável aleatória $Z = \frac{X-\mu}{\sigma}$ tem as seguintes propriedades:

- $E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E\left(X-\mu\right) = \frac{1}{\sigma}\left(E(X)-\mu\right) = \frac{1}{\sigma}\left(\mu-\mu\right) = 0.$

- $Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) =$

Observações

Quando $\mu = 0$ e $\sigma^2 = 1$ temos a distribuição normal padrão, a qual denotamos por $Z \sim N(0,1)$.

Se $X \sim N(\mu, \sigma^2)$, então uma nova variável aleatória $Z = \frac{X-\mu}{\sigma}$ tem as seguintes propriedades:

- $E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E\left(X-\mu\right) = \frac{1}{\sigma}\left(E(X)-\mu\right) = \frac{1}{\sigma}\left(\mu-\mu\right) = 0.$
- $Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}Var\left(X-\mu\right) = \frac{1}{\sigma^2}Var(X) = \frac{1}{\sigma^2}\ \sigma^2 = 1.$

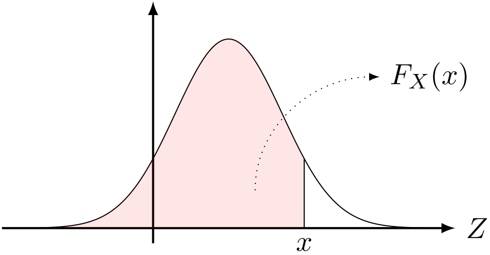
Pode-se mostrar que a v.a. Z também tem distribuição normal. Dessa forma, a transformação

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1).$$

Função de distribuição acumulada

Seja $X \sim N(\mu, \sigma^2)$, então, para $x \in \mathbb{R}$,

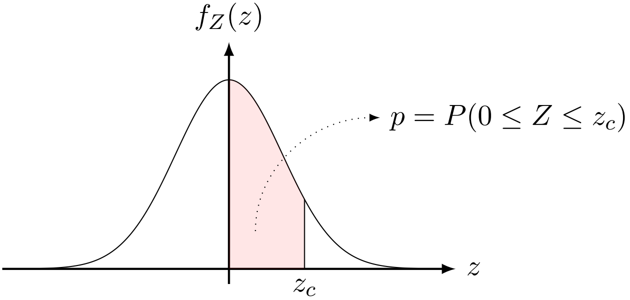
$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt.$$



O integrando não pode ser resolvido analiticamente. **Solução:** Integração numérica, uso da tabela de probabilidade da distribuição normal padrão.

Tabela da distribuição normal padrão

Na tabela são apresentadas os valores para $P(0 \leq Z \leq z_c)$, onde $z_c \geq 0$ e $Z \sim N(0, 1)$.



Note que $P(Z \leq 0) = P(Z \geq 0) = 0.5$.

Exercício: Queremos calcular, para $Z \sim N(0,1)$,

a) $P(0 \leq Z \leq 0.21)$

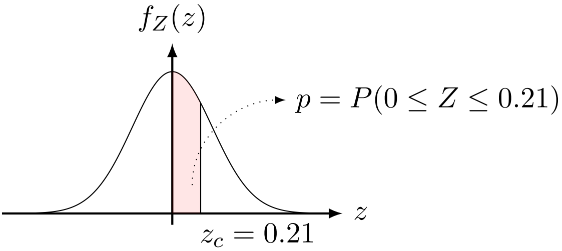
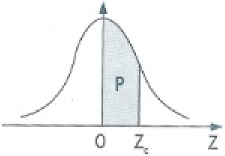


Tabela III — Distribuição Normal Padrão

$Z \sim N(0, 1)$

Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z < Z_c)$



parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c: 0	1	2	3	4	5	6	7	8	9	parte inteira e primeira decimal de Z_c
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

*Tabela retirada do livro Estatística Básica, de Pedro A. Morettin e Wilton de O. Bussab.

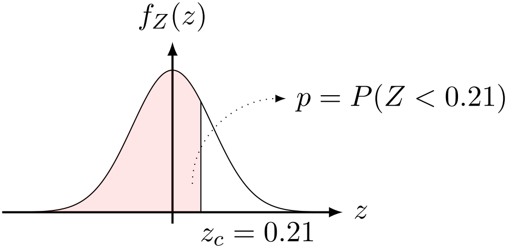
Portanto, $P(0 \leq Z \leq 0.21) = 0.08317$.

Portanto, $P(0 \leq Z \leq 0.21) = 0.08317$.

b) $P(Z < 0.21)$

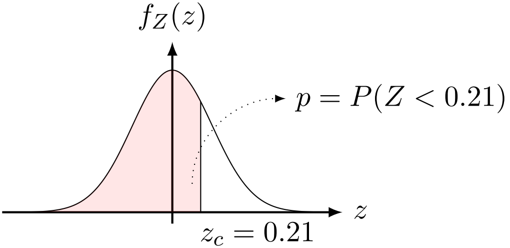
Portanto, $P(0 \leq Z \leq 0.21) = 0.08317$.

b) $P(Z < 0.21)$



Portanto, $P(0 \leq Z \leq 0.21) = 0.08317$.

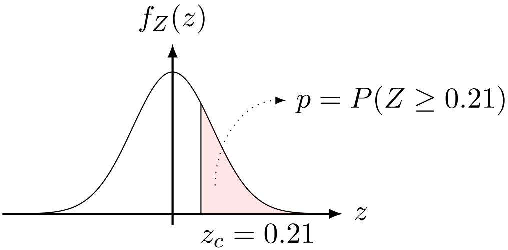
b) $P(Z < 0.21)$



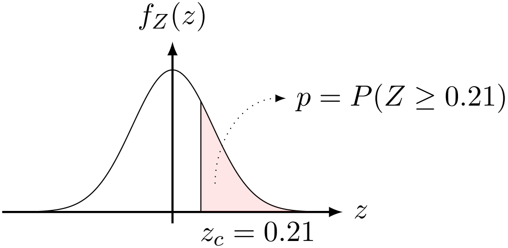
$$P(Z < 0.21) = 0.5 + P(0 \leq Z < 0.21) = 0.5 + 0.08317 = 0.58317.$$

c) $P(Z \geq 0.21)$

c) $P(Z \geq 0.21)$



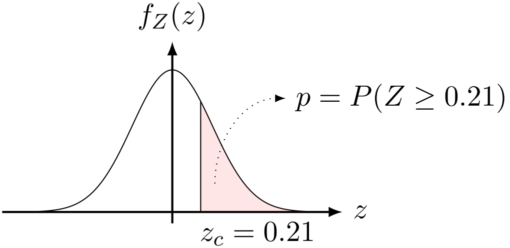
c) $P(Z \geq 0.21)$



Duas maneiras de resolver:

1ª maneira: $P(Z > -0.21) = 1 - P(Z < 0.21) \overset{(b)}{=} 1 - 0.58317 = 0.41683.$

c) $P(Z \geq 0.21)$



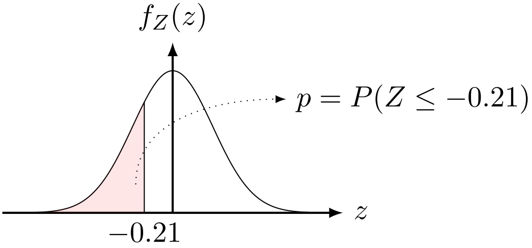
Duas maneiras de resolver:

1ª maneira: $P(Z > -0.21) = 1 - P(Z < 0.21) \overset{(b)}{=} 1 - 0.58317 = 0.41683$.

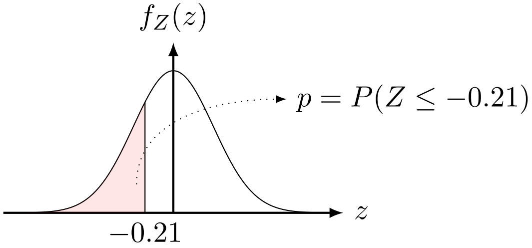
2ª maneira: $P(Z < 0.21) = 0.5 - P(0 \leq Z < 0.21) = 0.5 - 0.08317 = 0.41683$.

d) $P(Z \leq -0.21)$

d) $P(Z \leq -0.21)$



d) $P(Z \leq -0.21)$



Duas maneiras de resolver:

1ª maneira: Note que,

$$P(Z \leq 0.21) = 0.5 - P(-0.21 < Z < 0)$$

Mas, pela simetria da distribuição normal,

$$P(-0.21 < Z < 0) = P(0 < Z < 0.21) = 0.08317.$$

Assim,

$$P(Z \leq 0.21) = 0.5 - P(-0.21 < Z < 0) = 0.5 - 0.08317 = 0.41683.$$

$$P(Z \leq 0.21) = 0.5 - P(-0.21 < Z < 0) = 0.5 - 0.08317 = 0.41683.$$

2ª maneira: $P(Z < -0.21) = P(Z > 0.21) \stackrel{(c)}{=} 0.41683.$

$$P(Z \leq 0.21) = 0.5 - P(-0.21 < Z < 0) = 0.5 - 0.08317 = 0.41683.$$

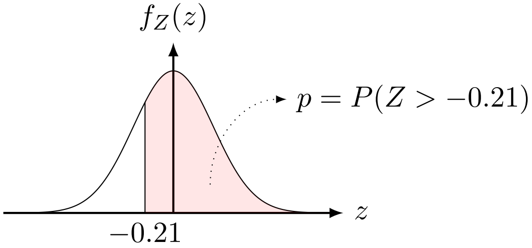
2ª maneira: $P(Z < -0.21) = P(Z > 0.21) \overset{(c)}{=} 0.41683.$

e) $P(Z > -0.21)$

$$P(Z \leq 0.21) = 0.5 - P(-0.21 < Z < 0) = 0.5 - 0.08317 = 0.41683.$$

2ª maneira: $P(Z < -0.21) = P(Z > 0.21) \overset{(c)}{=} 0.41683.$

e) $P(Z > -0.21)$



Duas maneiras de resolver:

1ª maneira: $P(Z > -0.21) = P(Z < 0.21) \stackrel{\text{(b)}}{=} 0.58317.$

Duas maneiras de resolver:

1ª maneira: $P(Z > -0.21) = P(Z < 0.21) \overset{\text{(b)}}{=} 0.58317.$

2ª maneira: Note que,

$$P(Z > -0.21) = 0.5 + P(-0.21 < Z < 0)$$

Mas, pela simetria da distribuição normal,

$$P(-0.21 < Z < 0) = P(0 < Z < 0.21) = 0.08317.$$

Assim,

$$P(Z \leq 0.21) = 0.5 + P(-0.21 < Z < 0) = 0.5 + 0.08317 = 0.58317.$$

Exemplo: Seja X o custo de manutenção de determinada aplicação bancária. Suponha que $X \sim N(10, 4)$. Qual a probabilidade de encontrarmos uma aplicação bancária com custo superior a 13 u.m.?

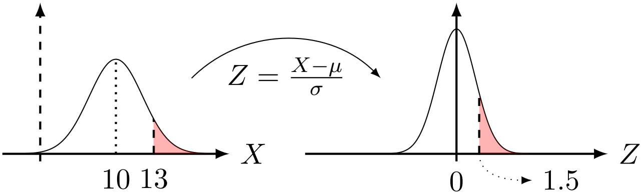
Exemplo: Seja X o custo de manutenção de determinada aplicação bancária. Suponha que $X \sim N(10, 4)$. Qual a probabilidade de encontrarmos uma aplicação bancária com custo superior a 13 u.m.?

Queremos encontrar $P(X > 13)$.

Vimos que se $X \sim N(\mu, \sigma^2)$, então uma nova variável aleatória $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

No nosso problema $\mu = 10$ e $\sigma = 2$, assim $Z = \frac{X-10}{2} \sim N(0, 1)$. Dessa forma,

$$P(X > 13) = P\left(\frac{X - 10}{2} > \frac{13 - 10}{2}\right) = P(Z > 1.5).$$



Assim:

- basta conhecer as probabilidades para uma v.a. normal padrão Z;
- utilizando a padronização $Z = \frac{X-\mu}{\sigma}$, obtemos as probabilidades para normais com quaisquer parâmetros μ e σ^2.

Assim,

$$= P(Z > 1.5) = 0.5 - P(0 < Z \leq 1.5) = 0.5 - 0.43319 = 0.06681.$$

Logo, a probabilidade de encontrarmos uma aplicação bancária com custo superior a 13 u.m. é 0.06681.

Assim:

- basta conhecer as probabilidades para uma v.a. normal padrão Z;
- utilizando a padronização $Z = \frac{X-\mu}{\sigma}$, obtemos as probabilidades para normais com quaisquer parâmetros μ e σ^2.

Assim,

$$= P(Z > 1.5) = 0.5 - P(0 < Z \leq 1.5) = 0.5 - 0.43319 = 0.06681.$$

Logo, a probabilidade de encontrarmos uma aplicação bancária com custo superior a 13 u.m. é 0.06681.

Exemplo: Seja $X \sim N(3, 16)$. Encontre a probabilidade de $P(2 \leq X \leq 5)$.

Combinação linear

Qualquer combinação linear de variáveis aleatórias normais independentes também terá distribuição normal.

Combinação linear

Qualquer combinação linear de variáveis aleatórias normais independentes também terá distribuição normal.

Matematicamente, se $X_1, X_2, ..., X_n$ formam uma sequencia de variáveis aleatórias $N(\mu_i, \sigma_i^2)$ independentes e $a_1, a_2, ..., a_n$ são constantes quaisquer então

$$W = \sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right).$$

Exemplo 1: Sejam $X \sim N(2,3)$ e $Y \sim N(-1,4)$ variáveis aleatórias independentes. Determine a distribuição de

a) $W = 2X + 3Y$

b) $Z = 2X - 3Y$

Exemplo 1: Sejam $X \sim N(2,3)$ e $Y \sim N(-1,4)$ variáveis aleatórias independentes. Determine a distribuição de

a) $W = 2X + 3Y$

b) $Z = 2X - 3Y$

Solução item (a):

$$E(W) = E(2X + 3Y)$$

Exemplo 1: Sejam $X \sim N(2,3)$ e $Y \sim N(-1,4)$ variáveis aleatórias independentes. Determine a distribuição de

a) $W = 2X + 3Y$
b) $Z = 2X - 3Y$

Solução item (a):

$$\begin{aligned} E(W) &= E(2X + 3Y) \\ &= E(2X) + E(3Y) \end{aligned}$$

Exemplo 1: Sejam $X \sim N(2,3)$ e $Y \sim N(-1,4)$ variáveis aleatórias independentes. Determine a distribuição de

a) $W = 2X + 3Y$

b) $Z = 2X - 3Y$

Solução item (a):

$$\begin{aligned} E(W) &= E(2X + 3Y) \\ &= E(2X) + E(3Y) \\ &= 2E(X) + 3E(Y) \end{aligned}$$

Exemplo 1: Sejam $X \sim N(2,3)$ e $Y \sim N(-1,4)$ variáveis aleatórias independentes. Determine a distribuição de

a) $W = 2X + 3Y$

b) $Z = 2X - 3Y$

Solução item (a):

$$\begin{aligned} E(W) &= E(2X+3Y) \\ &= E(2X) + E(3Y) \\ &= 2E(X) + 3E(Y) \\ &= 2 \times 2 + 3 \times (-1) \end{aligned}$$

Exemplo 1: Sejam $X \sim N(2,3)$ e $Y \sim N(-1,4)$ variáveis aleatórias independentes. Determine a distribuição de

a) $W = 2X + 3Y$
b) $Z = 2X - 3Y$

Solução item (a):

$$
\begin{aligned}
E(W) &= E(2X + 3Y) \\
&= E(2X) + E(3Y) \\
&= 2E(X) + 3E(Y) \\
&= 2 \times 2 + 3 \times (-1) \\
&= 1
\end{aligned}
$$

$$Var(W) = Var(2X + 3Y)$$

$$
\begin{aligned}
Var(W) &= Var(2X + 3Y) \\
&= Var(2X) + Var(3Y)
\end{aligned}
$$

$$
\begin{aligned}
Var(W) &= Var(2X + 3Y) \\
&= Var(2X) + Var(3Y) \\
&= 2^2 Var(X) + 3^2 Var(Y)
\end{aligned}
$$

$$
\begin{aligned}
Var(W) &= Var(2X+3Y) \\
&= Var(2X)+Var(3Y) \\
&= 2^2Var(X)+3^2Var(Y) \\
&= 4\times 3+9\times 4
\end{aligned}
$$

$$
\begin{aligned}
Var(W) &= Var(2X+3Y) \\
&= Var(2X)+Var(3Y) \\
&= 2^2Var(X)+3^2Var(Y) \\
&= 4\times 3+9\times 4 \\
&= 48
\end{aligned}
$$

Então, como X e Y tem distribuição normal independente,

$$W = 2X+3Y \sim N(1,48).$$

Solução item (b):

$$E(Z) = E(2X - 3Y)$$

Solução item (b):

$$
\begin{aligned}
E(Z) &= E(2X - 3Y) \\
&= E(2X) + E(-3Y)
\end{aligned}
$$

Solução item (b):

$$\begin{aligned} E(Z) &= E(2X - 3Y) \\ &= E(2X) + E(-3Y) \\ &= 2E(X) - 3E(Y) \end{aligned}$$

Solução item (b):

$$
\begin{aligned}
E(Z) &= E(2X - 3Y) \\
&= E(2X) + E(-3Y) \\
&= 2E(X) - 3E(Y) \\
&= 2 \times 2 - 3 \times (-1)
\end{aligned}
$$

Solução item (b):

$$
\begin{aligned}
E(Z) &= E(2X - 3Y) \\
&= E(2X) + E(-3Y) \\
&= 2E(X) - 3E(Y) \\
&= 2 \times 2 - 3 \times (-1) \\
&= 7
\end{aligned}
$$

Solução item (b):

$$\begin{aligned} E(Z) &= E(2X - 3Y) \\ &= E(2X) + E(-3Y) \\ &= 2E(X) - 3E(Y) \\ &= 2 \times 2 - 3 \times (-1) \\ &= 7 \end{aligned}$$

$$Var(W) = Var(2X - 3Y)$$

Solução item (b):

$$
\begin{aligned}
E(Z) &= E(2X - 3Y) \\
&= E(2X) + E(-3Y) \\
&= 2E(X) - 3E(Y) \\
&= 2 \times 2 - 3 \times (-1) \\
&= 7
\end{aligned}
$$

$$
\begin{aligned}
Var(W) &= Var(2X - 3Y) \\
&= Var(2X) + Var(-3Y)
\end{aligned}
$$

Solução item (b):

$$\begin{aligned} E(Z) &= E(2X - 3Y) \\ &= E(2X) + E(-3Y) \\ &= 2E(X) - 3E(Y) \\ &= 2 \times 2 - 3 \times (-1) \\ &= 7 \end{aligned}$$

$$\begin{aligned} Var(W) &= Var(2X - 3Y) \\ &= Var(2X) + Var(-3Y) \\ &= 2^2 Var(X) + (-3)^2 Var(Y) \end{aligned}$$

Solução item (b):

$$\begin{aligned} E(Z) &= E(2X - 3Y) \\ &= E(2X) + E(-3Y) \\ &= 2E(X) - 3E(Y) \\ &= 2 \times 2 - 3 \times (-1) \\ &= 7 \end{aligned}$$

$$\begin{aligned} Var(W) &= Var(2X - 3Y) \\ &= Var(2X) + Var(-3Y) \\ &= 2^2 Var(X) + (-3)^2 Var(Y) \\ &= 4 \times 3 + 9 \times 4 \end{aligned}$$

Solução item (b):

$$\begin{aligned} E(Z) &= E(2X-3Y) \\ &= E(2X)+E(-3Y) \\ &= 2E(X)-3E(Y) \\ &= 2\times 2-3\times(-1) \\ &= 7 \end{aligned}$$

$$\begin{aligned} Var(W) &= Var(2X-3Y) \\ &= Var(2X)+Var(-3Y) \\ &= 2^2Var(X)+(-3)^2Var(Y) \\ &= 4\times 3+9\times 4 \\ &= 48 \end{aligned}$$

Então, como X e Y tem distribuição normal independente,

$$Z = 2X-3Y \sim N(7,48).$$

Exemplo 2: Uma corretora negocia Títulos da Bolsa de Valores e utiliza um modelo probabilístico para avaliar seus lucros com respeito as áreas de agricultura, indústria e comércio.

$$L = 2\ L_A + 5\ L_I + 3\ L_C,$$

em que L_A, L_I e L_C representam, respectivamente, os lucros diários nos setores de agricultura, indústria e comércio. Considere que $L_A \sim N(3, 4)$, $L_I \sim N(6, 9)$ e $L_C \sim N(4, 16)$. Supondo independência entre os 3 setores, qual a probabilidade do lucro diário ser acima de 50 mil?

Exemplo 2: Uma corretora negocia Títulos da Bolsa de Valores e utiliza um modelo probabilístico para avaliar seus lucros com respeito as áreas de agricultura, indústria e comércio.

$$L = 2\ L_A + 5\ L_I + 3\ L_C,$$

em que L_A, L_I e L_C representam, respectivamente, os lucros diários nos setores de agricultura, indústria e comércio. Considere que $L_A \sim N(3, 4)$, $L_I \sim N(6, 9)$ e $L_C \sim N(4, 16)$. Supondo independência entre os 3 setores, qual a probabilidade do lucro diário ser acima de 50 mil?

Solução:
Queremos calcular a $P(L > 50)$.

L tem distribuição normal pois é uma combinação linear de normais independentes, então.

L tem distribuição normal pois é uma combinação linear de normais independentes, então.

$$E(L) = E(2L_A + 5L_I + 3L_C)$$

L tem distribuição normal pois é uma combinação linear de normais independentes, então.

$$\begin{aligned} E(L) &= E(2L_A + 5L_I + 3L_C) \\ &= 2E(L_A) + 5E(L_I) + 3E(L_C) \end{aligned}$$

L tem distribuição normal pois é uma combinação linear de normais independentes, então.

$$\begin{aligned} E(L) &= E(2L_A + 5L_I + 3L_C) \\ &= 2E(L_A) + 5E(L_I) + 3E(L_C) \\ &= 2 \times 3 + 5 \times 6 + 3 \times 4 \end{aligned}$$

L tem distribuição normal pois é uma combinação linear de normais independentes, então.

$$\begin{aligned} E(L) &= E(2L_A + 5L_I + 3L_C) \\ &= 2E(L_A) + 5E(L_I) + 3E(L_C) \\ &= 2 \times 3 + 5 \times 6 + 3 \times 4 \\ &= 48 \end{aligned}$$

L tem distribuição normal pois é uma combinação linear de normais independentes, então.

$$\begin{aligned} E(L) &= E(2L_A + 5L_I + 3L_C) \\ &= 2E(L_A) + 5E(L_I) + 3E(L_C) \\ &= 2 \times 3 + 5 \times 6 + 3 \times 4 \\ &= 48 \end{aligned}$$

$$Var(L) = Var(2L_A + 5L_I + 3L_C)$$

L tem distribuição normal pois é uma combinação linear de normais independentes, então.

$$\begin{aligned} E(L) &= E(2L_A + 5L_I + 3L_C) \\ &= 2E(L_A) + 5E(L_I) + 3E(L_C) \\ &= 2 \times 3 + 5 \times 6 + 3 \times 4 \\ &= 48 \end{aligned}$$

$$\begin{aligned} Var(L) &= Var(2L_A + 5L_I + 3L_C) \\ &= 2^2 Var(L_A) + 5^2 Var(L_I) + 3^2 Var(L_C) \end{aligned}$$

L tem distribuição normal pois é uma combinação linear de normais independentes, então.

$$\begin{aligned} E(L) &= E(2L_A + 5L_I + 3L_C) \\ &= 2E(L_A) + 5E(L_I) + 3E(L_C) \\ &= 2 \times 3 + 5 \times 6 + 3 \times 4 \\ &= 48 \end{aligned}$$

$$\begin{aligned} Var(L) &= Var(2L_A + 5L_I + 3L_C) \\ &= 2^2 Var(L_A) + 5^2 Var(L_I) + 3^2 Var(L_C) \\ &= 4 \times 4 + 25 \times 9 + 9 \times 16 \end{aligned}$$

L tem distribuição normal pois é uma combinação linear de normais independentes, então.

$$\begin{aligned} E(L) &= E(2L_A + 5L_I + 3L_C) \\ &= 2E(L_A) + 5E(L_I) + 3E(L_C) \\ &= 2 \times 3 + 5 \times 6 + 3 \times 4 \\ &= 48 \end{aligned}$$

$$\begin{aligned} Var(L) &= Var(2L_A + 5L_I + 3L_C) \\ &= 2^2 Var(L_A) + 5^2 Var(L_I) + 3^2 Var(L_C) \\ &= 4 \times 4 + 25 \times 9 + 9 \times 16 \\ &= 385 \end{aligned}$$

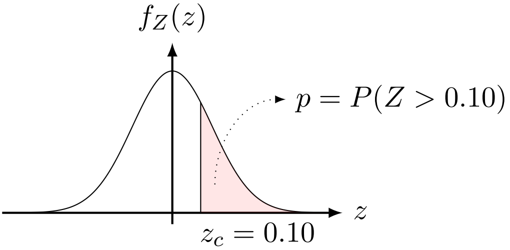
L tem distribuição normal pois é uma combinação linear de normais independentes, então.

$$\begin{aligned} E(L) &= E(2L_A + 5L_I + 3L_C) \\ &= 2E(L_A) + 5E(L_I) + 3E(L_C) \\ &= 2 \times 3 + 5 \times 6 + 3 \times 4 \\ &= 48 \end{aligned}$$

$$\begin{aligned} Var(L) &= Var(2L_A + 5L_I + 3L_C) \\ &= 2^2Var(L_A) + 5^2Var(L_I) + 3^2Var(L_C) \\ &= 4 \times 4 + 25 \times 9 + 9 \times 16 \\ &= 385 \end{aligned}$$

Assim, $L \sim N(48, 385)$ e

$$P(L > 50) = P\left(\frac{L-48}{\sqrt{385}} > \frac{50-48}{\sqrt{385}}\right) = P(Z > 0.10)$$



$$\begin{aligned} P(L > 50) &= P(Z > 0,10) \\ &= 0.5 - P(0 < Z \leq 0.10) \\ &= 0.5 - 0.0398 \\ &= 0.4602. \end{aligned}$$

Aproximação da binomial pela normal

Exemplo: Considere que uma linha de fabricação produz peças defeituosas com probabilidade de 0.05. Seja $Y =$ "número de peças defeituosas num total de 5000 observadas". Percebemos que $Y \sim B(n, p)$, onde $n = 5000$ e $p = 0.05$. Se fosse perguntado a probabilidade dessa amostra conter menos do que 230 peças defeituosas, como calcular essa probabilidade?

Aproximação da binomial pela normal

Exemplo: Considere que uma linha de fabricação produz peças defeituosas com probabilidade de 0.05. Seja Y = "número de peças defeituosas num total de 5000 observadas". Percebemos que $Y \sim B(n, p)$, onde $n = 5000$ e $p = 0.05$. Se fosse perguntado a probabilidade dessa amostra conter menos do que 230 peças defeituosas, como calcular essa probabilidade?

A rigor, deveríamos calcular

$$P(Y < 230) = \sum_{x=0}^{229} \binom{5000}{x} 0.05^x (0.95)^{5000-x}.$$

Aproximação da binomial pela normal

Exemplo: Considere que uma linha de fabricação produz peças defeituosas com probabilidade de 0.05. Seja $Y =$ "número de peças defeituosas num total de 5000 observadas". Percebemos que $Y \sim B(n, p)$, onde $n = 5000$ e $p = 0.05$. Se fosse perguntado a probabilidade dessa amostra conter menos do que 230 peças defeituosas, como calcular essa probabilidade?
A rigor, deveríamos calcular

$$P(Y < 230) = \sum_{x=0}^{229} \binom{5000}{x} 0.05^x (0.95)^{5000-x}.$$

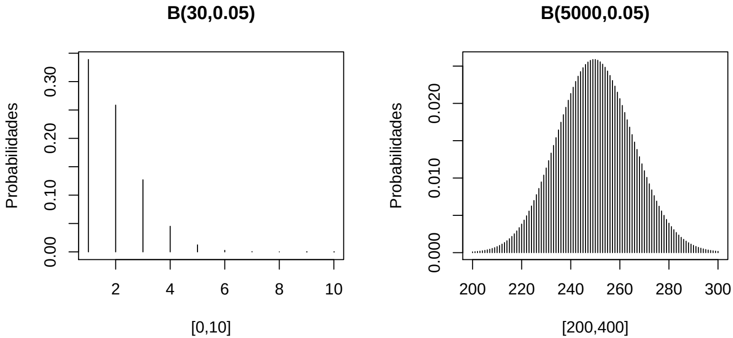
Essa tarefa pode ser facilitada percebendo que, para n suficientemente grande, o gráfico da função de probabilidades de uma v.a. $B(n, p)$ se assemelha muito com a f.d.p. de uma v.a. $N(\mu, \sigma^2)$!

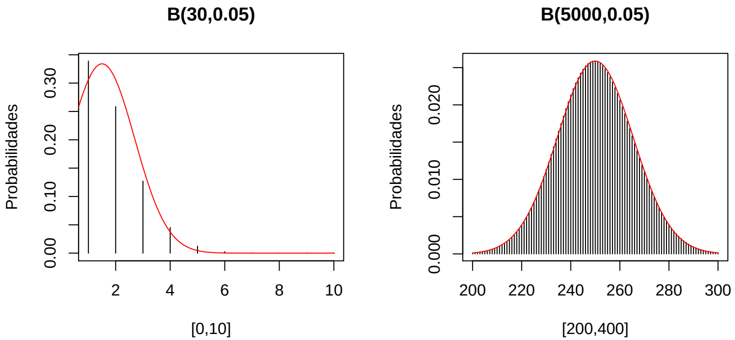
Aproximação da binomial pela normal

Exemplo: Considere que uma linha de fabricação produz peças defeituosas com probabilidade de 0.05. Seja $Y =$ "número de peças defeituosas num total de 5000 observadas". Percebemos que $Y \sim B(n, p)$, onde $n = 5000$ e $p = 0.05$. Se fosse perguntado a probabilidade dessa amostra conter menos do que 230 peças defeituosas, como calcular essa probabilidade?
A rigor, deveríamos calcular

$$P(Y < 230) = \sum_{x=0}^{229} \binom{5000}{x} 0.05^x (0.95)^{5000-x}.$$

Essa tarefa pode ser facilitada percebendo que, para n suficientemente grande, o gráfico da função de probabilidades de uma v.a. $B(n, p)$ se assemelha muito com a f.d.p. de uma v.a. $N(\mu, \sigma^2)$!
Mas quais valores utilizar para μ e σ^2?


B(30,0.05)
Probabilidades
0.00
0.10
0.20
0.30
2
4
6
8
10
[0,10]
B(5000,0.05)
Probabilidades
0.000
0.010
0.020
200
220
240
260
280
300
[200,400]



$$N(np, np(1-p)).$$

Critério para aproximação na prática: $np \geq 5$ e $n(1-p) \geq 5$
Como no exercício $n = 5000$ e $p = 0.05$, então $np = 5000 \times 0.05 = 250 \geq 5$ e $n(1-p) = 5000 \times 0.95 = 4750 \geq 5$, logo podemos aplicar a aproximação.

Portanto, a probabilidade do exemplo anterior pode ser aproximada da seguinte maneira:

$$P(Y < 230) \approx P(X < 230),$$

onde $X \sim N(\mu, \sigma^2)$, com

$\mu = np = 5000 \cdot 0.05 = 250$ e $\sigma^2 = np(1-p) = 5000 \cdot 0.05 \cdot 0.95 = 237.5.$

Logo,

$$P(Y < 230) \approx P(X < 230) = P\left(\frac{X-\mu}{\sigma} < \frac{230-250}{\sqrt{237.5}}\right) = P(Z < -1.29),$$

onde Z é a v.a. normal padrão. Olhando a tabela temos que

$$P(Y < 230) \approx 0.0985$$

Portanto, a probabilidade do exemplo anterior pode ser aproximada da seguinte maneira:

$$P(Y < 230) \approx P(X < 230),$$

onde $X \sim N(\mu, \sigma^2)$, com

$\mu = np = 5000 \cdot 0.05 = 250$ e $\sigma^2 = np(1-p) = 5000 \cdot 0.05 \cdot 0.95 = 237.5.$

Logo,

$$P(Y < 230) \approx P(X < 230) = P\left(\frac{X-\mu}{\sigma} < \frac{230-250}{\sqrt{237.5}}\right) = P(Z < -1.29),$$

onde Z é a v.a. normal padrão. Olhando a tabela temos que

$$P(Y < 230) \approx 0.0985$$

O valor verdadeiro é 0.0904.

Para melhorar a aproximação, alguns autores sugerem utilizar a chamada correção por continuidade no cálculo com a normal.

Y: binomial

X: normal

- $P(Y = k) \approx P(k - 0.5 \leq X \leq k + 0.5)$
- $P(Y < k) \approx P(X < k - 0.5)$
- $P(Y > k) \approx P(X > k + 0.5)$
- $P(Y \leq k) \approx P(X < k + 0.5)$
- $P(Y \geq k) \approx P(X > k - 0.5)$

No exercício,

$P(Y < 230) \approx P(X < 230 - 0.5) = P(X < 229.5)$. Assim,

$$P(Y < 230) \approx P\left(\frac{X-\mu}{\sigma} < \frac{229.5 - 250}{\sqrt{237.5}}\right) = P(Z < -1.33)$$

onde Z é a v.a. normal padrão. Olhando a tabela temos que

$$P(Y < 230) \approx 0.5 - 0.40824 = 0.09176.$$

No exercício,

$P(Y < 230) \approx P(X < 230 - 0.5) = P(X < 229.5)$. Assim,

$$P(Y < 230) \approx P\left(\frac{X-\mu}{\sigma} < \frac{229.5-250}{\sqrt{237.5}}\right) = P(Z < -1.33)$$

onde Z é a v.a. normal padrão. Olhando a tabela temos que

$$P(Y < 230) \approx 0.5 - 0.40824 = 0.09176.$$

O valor verdadeiro é 0.0904.

Aproximação normal à Poisson

A distribuição de Poisson se aproxima da normal quando λ é grande. Quando $Y \sim Po(\lambda)$, vimos que $E(Y) = \lambda$ e $Var(Y) = \lambda$.

Aproximação normal à Poisson

A distribuição de Poisson se aproxima da normal quando λ é grande. Quando $Y \sim Po(\lambda)$, vimos que $E(Y) = \lambda$ e $Var(Y) = \lambda$.

$$Y \sim Po(\lambda) \xrightarrow{\lambda \text{ grande}} X \sim N(\lambda, \lambda).$$

A aproximação é razoável para $\lambda \geq 5$. Aqui também se faz necessária a correção por continuidade para melhorar as aproximações.