Introdução à teoria das probabilidades

Modelos probabilísticos contínuos

Francisco A. Rodrigues

Universidade de São Paulo

Distribuição uniforme

Uma variável aleatória contínua X segue uma distribuição uniforme se sua função densidade de probabilidade é dada por:

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & x \notin [a, b]. \end{cases} \tag{1}$$

Distribuição uniforme

Seja X uma variável aleatória uniformemente distribuída em $[a, b]$. Então,

$$E[X] = \frac{a+b}{2}, \tag{2}$$

$$V(X) = \frac{(b-a)^2}{12}. \tag{3}$$

Distribuição uniforme

Para demonstrarmos esse teorema, temos:

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_a^b \frac{x}{b-a}dx = \left.\frac{x^2}{2(b-a)}\right|_a^b = \frac{a+b}{2}.$$

No caso da variância:

$$\begin{aligned}
V(X) &= E[X^2] - E[X]^2 \\
&= \int_a^b x^2 \left(\frac{1}{b-a}\right) dx - \left(\frac{a+b}{2}\right)^2 \\
&= \left.\frac{x^3}{3(b-a)}\right|_a^b - \left(\frac{a+b}{2}\right)^2 \\
&= \frac{b^3-a^3}{3(b-a)} - \frac{a^2+2ab+b^2}{4} \\
&= \frac{4(b-a)(a^2+ab+b^2)}{12(b-a)} - \frac{3(a+b)^2}{12} \\
&= \frac{4a^2+4ab+4b^2-3a^2-6ab-3b^2}{12} \\
&= \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}.
\end{aligned}$$

Distribuição uniforme

Exemplo: O tempo que uma pessoa espera por um ônibus em um terminal, medido em minutos, é uniformemente distribuído em $[0, 15]$.
a) Calcule a probabilidade de que uma pessoa espere menos de 12,5 minutos pelo ônibus.
b) Em média, quanto tempo uma pessoa espera no terminal? Calcule também o desvio padrão.

Distribuição uniforme

Solução: a) Calcule a probabilidade de que uma pessoa espere menos de 12,5 minutos pelo ônibus.
Seja X a variável aleatória que representa o tempo de espera. Então,

$$f(x) = \begin{cases} \frac{1}{15}, & 0 \leq x \leq 15 \\ 0, & \text{caso contrário.} \end{cases}$$

Assim,

$$P(X \leq 12,5) = \int_0^{12,5} \frac{1}{15} dx = 0,833.$$

Logo, dentre 100 pessoas, aproximadamente 83 delas aguardarão pelo ônibus por no máximo 12,5 minutos.
b) Em média, quanto tempo uma pessoa espera no terminal? Calcule também o desvio padrão.
Temos:

$$E[X] = \frac{15+0}{2} = 7,5 \text{ minutos.}$$

$$\sigma = \sqrt{V(X)} = \sqrt{\frac{(15-0)^2}{12}} = 4,3 \text{ minutos.}$$

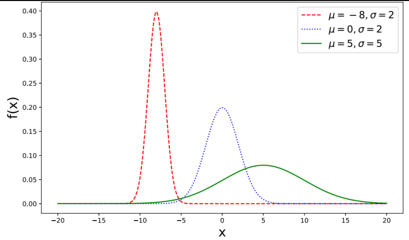
Portanto, se anotarmos o tempo de espera de 100 pessoas, esse tempo médio será próximo de 7,5 minutos.

Distribuição normal

Uma variável aleatória contínua X que tome todos os valores na reta real segue a distribuição normal (ou Gaussiana) se sua função densidade de probabilidade é definida por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad -\infty < x < \infty, \qquad (4)$$

onde $\mu = E[X]$ e $\sigma^2 = V(X) > 0$.



Distribuição normal

A distribuição normal apresenta as seguintes propriedades:

- $f(x)$ é simétrica em relação à μ.
- $f(x) \to 0$ quando $x \to \pm\infty$.
- O valor máximo de $f(x)$ ocorre em $x = \mu$.

Distribuição normal

Se X é uma variável aleatória contínua com distribuição normal, $X \sim \mathcal{N}(\mu, \sigma^2)$, e se $Y = aX + b$, com a e b constantes, então $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

Distribuição normal

Se X é uma variável aleatória contínua com distribuição normal, $X \sim \mathcal{N}(\mu, \sigma^2)$, e se $Y = aX + b$, com a e b constantes, então $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.
Para demonstrar esse teorema, vamos calcular a esperança de Y:

$$E[Y] = E[aX + b] = aE[X] + b = a\mu + b.$$

E a variância:

$$V(Y) = V(aX + b) = a^2 V(X) = a^2\sigma^2.$$

Distribuição normal

Seja $X \sim \mathcal{N}(\mu, \sigma^2)$. Se

$$Z = \frac{X - \mu}{\sigma}. \tag{5}$$

Então $Z \sim \mathcal{N}(0, 1)$.

Distribuição normal

Seja $X \sim \mathcal{N}(\mu, \sigma^2)$. Se

$$Z = \frac{X - \mu}{\sigma}. \tag{6}$$

Então $Z \sim \mathcal{N}(0, 1)$.
Para demonstrar o teorema, temos:

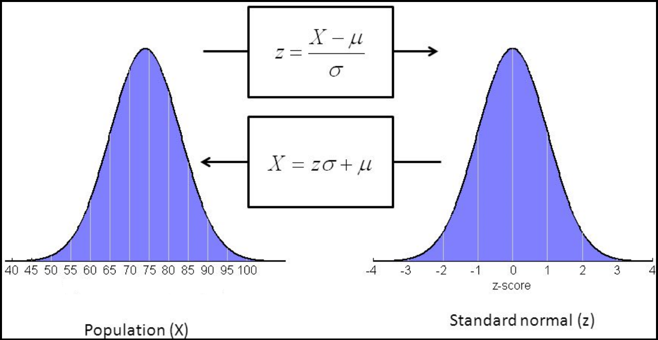
$$E[Z] = E\left[\frac{X - \mu}{\sigma}\right] = \frac{E[X] - \mu}{\sigma} = 0.$$

$$V(Z) = V\left[\frac{X - \mu}{\sigma}\right] = \frac{V(X)}{\sigma^2} = 1.$$

Distribuição normal

Tabela ou Python:

$$P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right)$$
$$= P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right).$$



Distribuição normal

Exemplo: Se $X \sim \mathcal{N}(\mu = 165, \sigma^2 = 9)$, calcule $P(X < 162)$.

Distribuição normal

Exemplo: Se $X \sim \mathcal{N}(\mu = 165, \sigma^2 = 9)$, calcule $P(X < 162)$. Vamos transformar X em Z no cálculo da probabilidade:

$$
\begin{aligned}
P(X < 162) &= P\left(\frac{X-\mu}{\sigma} < \frac{162-\mu}{\sigma}\right) \\
&= P\left(Z < \frac{162-165}{3}\right) \\
&= P(Z < -1) \\
&= 0,158.
\end{aligned}
$$

Esse valor pode ser verificado na tabela normal padronizada ou usando o código a seguir.

```
import scipy.stats as st

media = 165
dp = 3 # desvio padrão
z = (162-media)/dp
print('Z:', z)
print('P(X < 162) = ', st.norm.cdf(z))
```

Distribuição normal

Exemplo: Se $X \sim \mathcal{N}(\mu = 10, \, \sigma^2 = 4)$, calcule $P(X > 13)$.

Distribuição normal

Exemplo: Se $X \sim \mathcal{N}(\mu = 10, \, \sigma^2 = 4)$, calcule $P(X > 13)$.

$$\begin{aligned} P(X > 13) &= P\left(\frac{X-\mu}{\sigma} > \frac{13-\mu}{\sigma}\right) = \\ &= P\left(Z > \frac{13-10}{2}\right) \\ &= P(Z > 1,5) \\ &= 1 - P(Z \leq 1,5) \\ &= 1 - 0,93 \\ &= 0,07. \end{aligned}$$

Esse valor pode ser obtido usando-se o código a seguir.

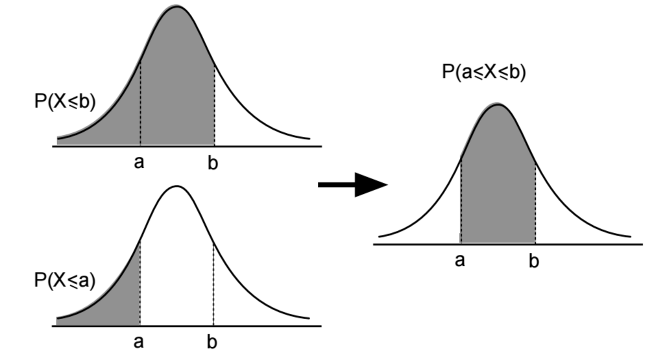
```
import scipy.stats as st

media = 10
dp = 2
z = (13-media)/dp
print('Z:', z)
print('P(X > 13)=', 1-st.norm.cdf(z))
```

Distribuição normal

Seja $X \sim \mathcal{N}(\mu, \sigma^2)$. Então:

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a). \tag{7}$$



Distribuição normal

Exemplo: Se $X \sim \mathcal{N}(\mu = 5, \sigma^2 = 4)$, calcule $P(4 \leq X \leq 6)$.

Distribuição normal

Exemplo: Se $X \sim \mathcal{N}(\mu = 5, \sigma^2 = 4)$, calcule $P(4 \leq X \leq 6)$.

$$
\begin{aligned}
P(4 \leq X \leq 6) &= P\left(\frac{4-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{6-\mu}{\sigma}\right) \\
&= P\left(\frac{4-5}{2} \leq Z \leq \frac{6-5}{2}\right) \\
&= P(-0,5 \leq Z \leq 0,5) \\
&= P(Z \leq 0,5) - P(Z \leq -0,5) \\
&= 0,38.
\end{aligned}
$$

Esse resultado pode ser obtido com o código a seguir.

```python
import scipy.stats as st
media = 5
dp = 2
z1 = (4-media)/dp
z2 = (6-media)/dp
print('z1=',z1,'z2=',z2)
print('P(4 < X < 6)=', st.norm.cdf(z2)- st.norm.cdf(z1))
```

Distribuição normal

Exemplo: O peso médio de 500 estudantes do sexo masculino em uma universidade é igual 70 Kg, sendo o desvio padrão igual a 5 Kg. Admitindo-se que os pesos são normalmente distribuídos, de forma aproximada, determine a percentagem de estudantes que pesam entre 65 Kg e 75 Kg.

Solução:

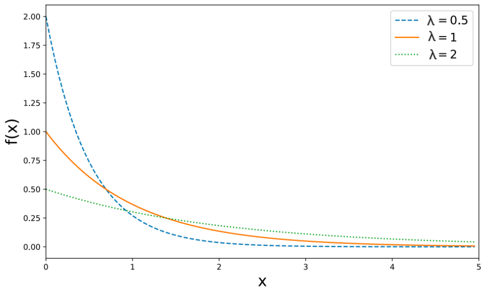
$$\begin{aligned}
P(65 \leq X \leq 75) &= P\left(\frac{65-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{75-\mu}{\sigma}\right) \\
&= P\left(\frac{65-70}{5} \leq Z \leq \frac{75-70}{5}\right) \\
&= P(-1 \leq Z \leq 1) \\
&= P(Z \leq 1) - P(Z \leq -1) \\
&= 0,68.
\end{aligned}$$

Distribuição exponencial

Uma variável aleatória contínua X segue o modelo exponencial se sua função densidade de probabilidade é dada por:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0, \end{cases} \tag{8}$$

onde $\lambda > 0$ e $-\infty < x < \infty$.



Distribuição exponencial

Exemplo: O intervalo de tempo entre emissões consecutivas de uma fonte radioativa é uma variável aleatória com distribuição exponencial com parâmetro $\lambda = 0,2$ emissões por minuto. Qual é a probabilidade de que ocorra uma emissão em um intervalo inferior a 2 minutos?

Distribuição exponencial

Solução: Seja T a variável aleatória que representa o tempo entre as emissões. Então,

$$
\begin{aligned}
P(T \leq 2) &= \int_0^2 \lambda e^{-\lambda t} dt = \int_0^2 0,2e^{-0,2t} dt \\
&= 0,2 \frac{e^{-0,2t}}{(-0,2)} \bigg|_0^2 = 1 - e^{-0,4} = 0,33.
\end{aligned}
$$

Logo, a probabilidade de que ocorra uma emissão em um intervalo inferior a 2 minutos é próxima de 33%.

Distribuição exponencial

Seja X uma variável aleatória com distribuição exponencial com parâmetro λ. Então:

$$E[X] = \frac{1}{\lambda}, \tag{9}$$

$$V(X) = \frac{1}{\lambda^2}. \tag{10}$$

Distribuição exponencial

Seja X uma variável aleatória com distribuição exponencial com parâmetro λ. Então:

$$E[X] = \frac{1}{\lambda}, \tag{11}$$

$$V(X) = \frac{1}{\lambda^2}. \tag{12}$$

Prova:

$$\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x(\lambda e^{-\lambda x}) dx \\
&= \left(-x e^{-\lambda x}\right)\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx \quad \text{(integrando por partes)} \\
&= 0 - 0 + \left(-\frac{1}{\lambda} e^{-\lambda x}\right)\Big|_0^{\infty} = \frac{1}{\lambda}.
\end{aligned}$$

$V(X)$: Exercício.

Distribuição exponencial

Exemplo: Estudantes chegam a uma festa a uma taxa de 30 estudantes por hora, passando pela portaria do evento. Qual é a probabilidade de que a portaria esperará pela chegada de um estudante por um tempo maior do que três minutos?

Distribuição exponencial

Solução: Seja X a variável aleatória que representa o tempo de espera por um novo estudante. Como estamos interessados em calcular a probabilidade de espera considerando o tempo em minutos, devemos converter a taxa de horas para minutos. A chegada de 30 estudantes por hora equivale a $\lambda = 0,5$ estudantes por minuto. Assim,

$$P(X > 3) = \int_3^\infty \frac{1}{2} e^{-0,5x} dx = \left. \frac{e^{-x/2}}{2(-1/2)} \right|_{x=3}^{\infty} = 0,223.$$

Portanto, a probabilidade de que a portaria esperará mais de três minutos por um novo estudante é próxima de 22%.

Distribuição exponencial

```python
import numpy as np
np.random.seed(100)

lbd = 0.5 # taxa
x = 3 # tempo de espera
ns = 100 # número de simulações
m = 0 # número de sucessos
for s in range(0,ns):
    t = np.random.exponential(1/lbd)
    if(t > x):
        m = m + 1
print('P(X > %d) = %s' % (x, m/ns))
```

Distribuição exponencial

Se X segue uma distribuição exponencial com parâmetro λ, então,

$$F(x) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases} \tag{13}$$

Distribuição exponencial

Exemplo: Seja X a variável aleatória que representa o tempo que clientes esperam para usar um caixa eletrônico. Esse tempo tem distribuição exponencial com média igual a quatro minutos.

a) Calcule a probabilidade de que um cliente gaste entre quatro e cinco minutos na fila.

b) Qual a probabilidade de que um cliente gaste mais do que seis minutos na fila?

Distribuição exponencial

Solução: a) Calcule a probabilidade de que um cliente gaste entre quatro e cinco minutos na fila. Temos que:

$$E[X] = 4 = \frac{1}{\lambda} \implies \lambda = \frac{1}{4}.$$

Assim, a função densidade de probabilidade associada a X é dada por:

$$f(x) = \begin{cases} \frac{1}{4}e^{-x/4}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

$$\begin{aligned} P(4 \leq X \leq 5) &= P(X < 5) - P(X < 4) \\ &= F(5) - F(4) \\ &= (1 - e^{-5/4}) - (1 - e^{-4/4}) \\ &= 0,7135 - 0,6321 \\ &= 0,0814. \end{aligned}$$

b) Qual a probabilidade de que um cliente gaste mais do que seis minutos na fila?

$$P(X > 6) = e^{-6/4} = 0,22.$$

Distribuição exponencial

A distribuição exponencial apresenta a propriedade de "ausência de memória", isto é:

$$P(X \geq t + s | X \geq s) = P(X \geq t). \tag{14}$$

Distribuição exponencial

A distribuição exponencial apresenta a propriedade de "ausência de memória", isto é:

$$P(X \geq t + s | X \geq s) = P(X \geq t). \tag{15}$$

Para demonstrar essa propriedade, usamos a definição de probabilidade condicional:

$$P(X \geq t + s | X \geq s) = \frac{P[(X \geq t + s) \cap (X \geq s)]}{P(X \geq s)}.$$

A interseção do intervalo resulta em:
$(X \geq t + s) \cap (X \geq s) \equiv (X \geq t + s)$. Assim,

$$P(X \geq t+s | X \geq s) = \frac{P(X \geq t + s)}{P(X \geq s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t).$$

Distribuição exponencial

Exemplo: Suponha que um cliente está em um banco aguardando para ser atendido. Se o cliente está na agência há 10 minutos, qual é a probabilidade de ele ainda aguarde 7 minutos para ser chamado? Assuma que a cada hora ocorrem, em média, 6 atendimentos.

Distribuição exponencial

Solução: Seja T a variável aleatória que representa o tempo de espera para ser atendido. Então, queremos calcular:

$$P(T > 17 | T > 10) = \frac{P((T > 17) \cap (T > 10))}{P(T > 10)} = P(T > 7).$$

Como ocorrem 6 atendimentos por hora, em média, temos que o tempo médio de espera para cada atendimento é de $E[T] = 60/6 = 10$ minutos. Logo, $\lambda = 1/10$. Portanto,

$$P(T > 7) = e^{-7/10} = 0,496.$$

Logo, a probabilidade é próxima de 50%.

Distribuição gama

Uma variável aleatória contínua X tem distribuição gama com parâmetros $\lambda > 0$ e $\alpha > 0$, se sua função densidade de probabilidade é dada por:

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0, \end{cases} \tag{16}$$

onde Γ é a função gama:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt. \tag{17}$$

Se α for um número inteiro positivo, a distribuição representará uma distribuição Erlang, ou seja, a soma de α variáveis aleatórias independentes distribuídas exponencialmente, cada uma delas com uma média $\theta = 1/\lambda$.

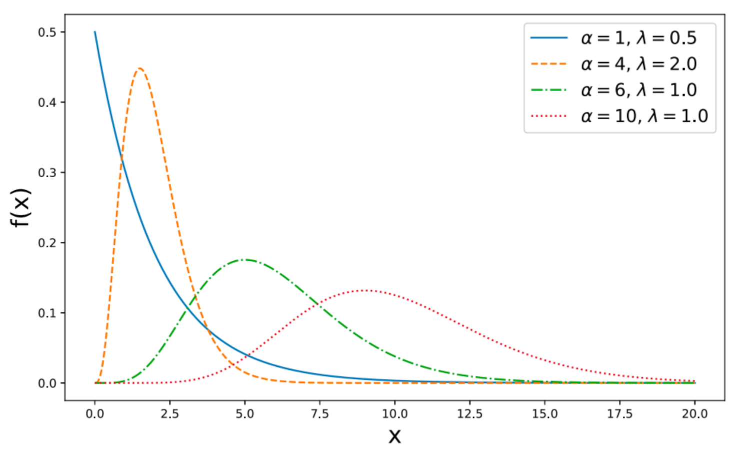
Distribuição gama

Seja X uma variável aleatória com distribuição gama. Então,

$$E[X] = \frac{\alpha}{\lambda}. \tag{18}$$

$$V(X) = \frac{\alpha}{\lambda^2}. \tag{19}$$

Distribuição gama



Distribuição gama

Exemplo: No desenvolvimento de um novo foguete, engenheiros inserem dois tanques de combustível, sendo um deles ativo e o outro usado como reserva, que é acionado caso ocorra uma falha no primeiro tanque. Suponha que em uma missão espacial necessita de 50 horas de voo para ser finalizada. De acordo com o fabricante do tanque, o tempo médio antes de ocorrer uma falha é de 100 horas. Calcule a probabilidade de que a missão será bem-sucedida. Assuma que o tempo de funcionamento dos tanques apresenta distribuição exponencial.

Distribuição gama

Solução: Como temos dois tanques, para que o sistema todo falhe devem ocorrer dois eventos (duas falhas), sendo que o tempo associado a cada um deles segue o modelo exponencial. Logo, o tempo para o sistema falhar segue uma distribuição gama com $\alpha = 2$ (notem que é a soma de duas distribuições exponenciais). Como o tempo médio para ocorrer uma falha em cada tanque é igual a 100 horas, temos que $\lambda = 1/100$ (ver equação (11)). Assim, se a variável aleatória X representa o tempo de funcionamento dos tanques de combustível, vamos calcular a probabilidade usando a equação (16):

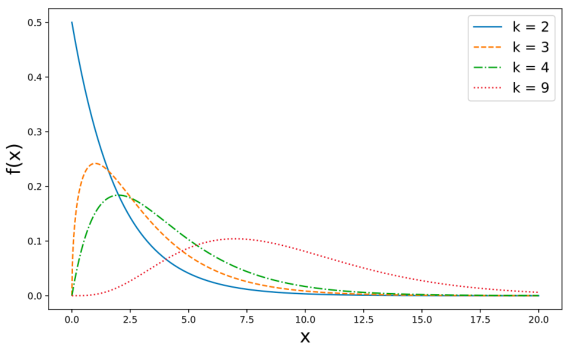
$$
\begin{aligned}
P(X > 50) &= 1 - P(X < 50) \\
&= 1 - \int_0^{50} \left(\frac{1}{100^2}\right) \frac{x^{2-1} e^{-x/100}}{\Gamma(2)} dx \\
&= 1 - \int_0^{50} \frac{x e^{-x/100}}{10000(1!)} dx \\
&= 1 - 0,09 \\
&= 0,91.
\end{aligned}
$$

Portanto, a probabilidade da missão ser bem sucedida é igual a 91%.

Distribuição qui-quadrado

A variável aleatória contínua X segue a distribuição qui-quadrado (denominada χ^2) se sua função densidade de probabilidade é dada por:

$$f(x) = \begin{cases} \frac{x^{k/2-1}e^{-x/2}}{2^{k/2}\Gamma(k/2)}, & x > 0 \\ 0, & x \leq 0, \end{cases} \tag{20}$$



Distribuição qui-quadrado

A distribuição qui-quadrado é definida pela soma de k distribuições normais padronizadas e independentes. Ou seja, X tem distribuição qui-quadrado com k graus de liberdade se

$$X = \sum_{i=1}^{k} Z_i^2, \tag{21}$$

onde $Z_1, Z_2, \ldots, Z_k$ são variáveis aleatórias com distribuição normal padronizada,

$$Z_i \sim \mathcal{N}(\mu = 0, \sigma^2 = 1), \quad i = 1, \ldots, k.$$

Para denominar que X segue uma distribuição qui-quadrado, usamos $X \sim \chi^2(k)$ ou $X \sim \chi^2_k$.

Distribuição beta

Seja X uma variável aleatória contínua limitada em $[0, 1]$. Dizemos que X segue uma distribuição beta se sua função densidade de probabilidade é dada por:

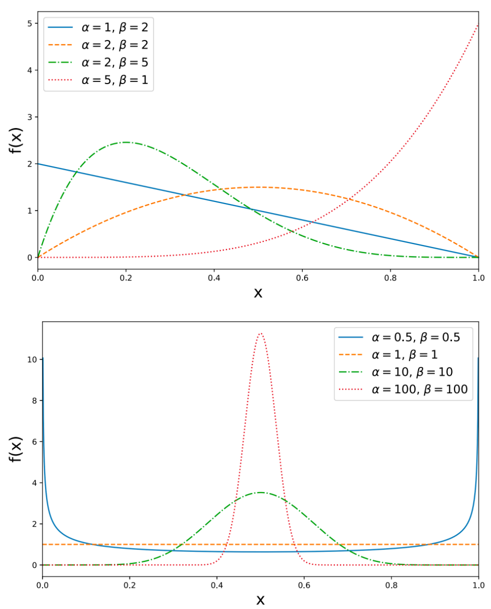
$$f(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & 0 < x < 1 \\ 0 \quad \text{caso contrário,} \end{cases} \tag{22}$$

onde $\alpha, \beta > 0$ e

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-u)^{\beta-1} du, \tag{23}$$

é a função beta, que atua como uma constante de normalização para que a área da função densidade de probabilidade seja igual a um.

Distribuição beta



Distribuição t de Student

A variável aleatória X tem distribuição t de Student com ν graus de liberdade se sua função densidade de probabilidade é dada por:

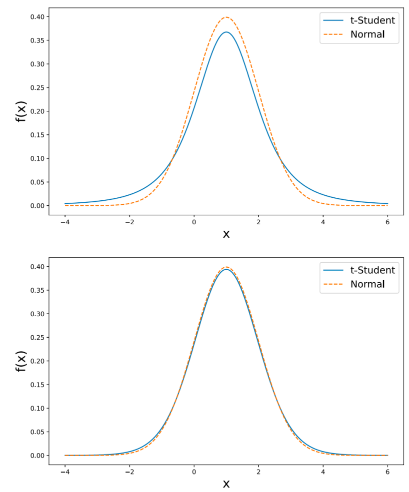
$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-(\frac{\nu+1}{2})}, \quad -\infty < x < \infty, \tag{24}$$

onde Γ é a função gama,

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1}e^{-t}dt. \tag{25}$$

Distribuição t de Student

Distribuição t de Student para $\nu = 3$ e $\nu = 20$ graus de liberdade. Quando aumentamos ν, a distribuição se aproxima da distribuição normal.



Sumário

- Distribuição Uniforme.
- Distribuição Normal.
- Distribuição exponencial.
- Distribuição qui-quadrado.
- Distribuição gama.
- Distribuição beta.
- Distribuição t de Student.