Variáveis Aleatórias Contínuas e Distribuição de Probabilidades - parte IV

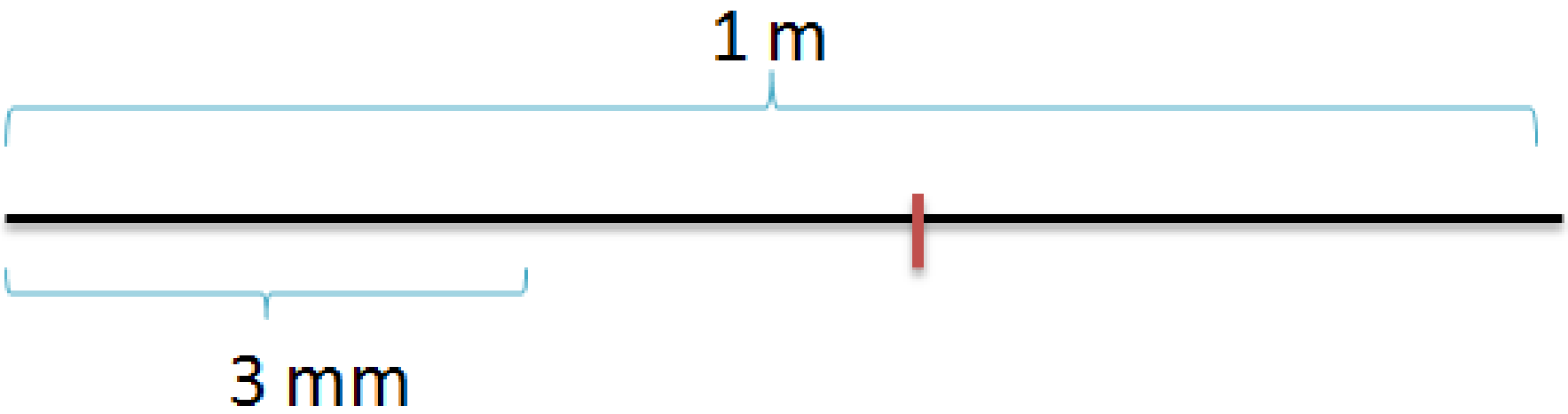
Marcos Oliveira Prates

2012/02

Distribuição Exponencial

- Vamos relembrar a definição de uma variável com Distribuição Poisson.
- Número de falhas ao longo do comprimento de um fio de cobre.
- A distância entre as falhas é uma variável de interesse.
- Seja X o comprimento de um ponto inicial até a primeira falha do fio.

- A distribuição de X pode ser obtida a partir da distribuição das falhas.
- Vejamos como estabelecer essa relação.
- A distância até a primeira falha excederá 3 milímetros somente se:
  - não houver nenhuma falha até um comprimento de 3 milímetros.
- Significa que até o comprimento de 3 milímetros nenhuma falha ainda foi observada.



- Se o número de falhas em um metro tem distribuição Poisson(λ).
- Em 3 mm esperamos observar $\lambda(0,03)$.
- A probabilidade requerida será

$$P(X > 3) = P(\text{zero falhas no intervalo de 0 a 3 mm})$$

$$= \frac{e^{-\lambda(0,03)}(\lambda(0,03))^0}{0!}$$

- Vamos ver agora o caso geral.
- Seja N o número de falhas em x milímetros.
- O número médio de falhas por metro é λ.
- N tem distribuição Poisson com média $x\lambda$.
- Consideramos que o fio tem comprimento maior que x.
- Temos que

$$P(X \geq x) = P(N = 0) = \frac{e^{-\lambda x} x^0}{0!} = e^{-\lambda x} \ .$$

- Logo

$$F(x) = P(X \leq x) = 1 - e^{-\lambda x} \qquad X \geq 0.$$

- A função de distribuição cumulativa de X é

$$F(x) = P(X \leq x) = 1 - e^{-\lambda x}.$$

- Diferenciando $F(X)$ temos que

$$f(x) = \lambda e^{-\lambda x} \qquad x \geq 0\,.$$

- A distribuição de X está ligada ao Processo de Poisson.
- Veremos que o ponto inicial para medir X não importa.
- No processo de Poisson o número de falhas depende apenas do comprimento do intervalo
  - e não da localização do intervalo.

Distribuição Exponencial

- Seja X a distância entre contagens em um Processo de Poisson.
- Considere que a média desse processo é $\lambda > 0$.
- X é então uma variável aleatória exponencial com parâmetro λ.
- A função densidade de probabilidade de X é

$$f(x) = \lambda e^{-\lambda x} \quad \text{para} \quad 0 \leq x < \infty \, .$$

- A distribuição Exponencial tem esse nome por causa da função exponencial que a aparece em sua densidade.
- Veja abaixo gráficos da distribuição exponencial para alguns valores de λ.

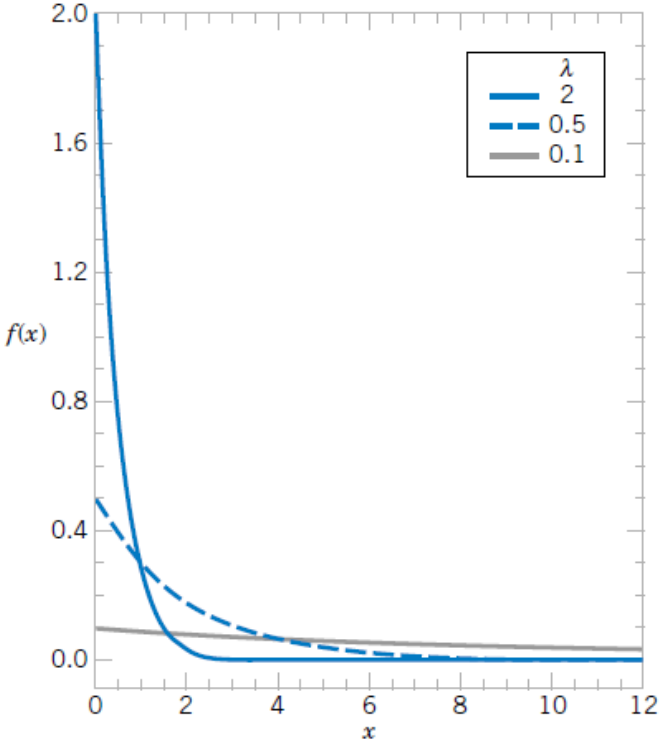


Figure 4-22 Probability density function of exponential random variables for selected values of λ.

- Vamos agora calcular média e variância dessa variável.
- Relembre que

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

- Usando a definição temos que

$$E(X) = \int_{0}^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \ .$$

(fazer no quadro)

- Então o tempo de espera até a falha é:
  - inversamente proporcional à taxa com que as falhas ocorrem.

- Para a variância, usamos

$$V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - (E(X))^2 .$$

- Integrando por partes temos que

$$\int_{0}^{\infty} x^2 f(x)dx = \frac{2}{\lambda^2} .$$

- Portanto

$$Var(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} .$$

Média e Variância

- Seja X uma variável aleatória com distribuição exponencial de parâmetro λ.
- Temos então que

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

Exemplo:

- Considere uma rede de usuários de computadores.
- As conexões dos usuários ao sistema podem ser modeladas por um Processo de Poisson.
- A média desse processo é de 25 conexões por hora.
- Qual a probabilidade de não haver conexões em um intervalo de 6 minutos?

Exemplo: (solução)

- Seja X o tempo em horas até a primeira conexão.
- X tem distribuição exponencial com $\lambda = 25$ conexões por hora.
- Queremos saber a probabilidade de X exceder 6 minutos.
- λ é dado em conexões por hora.
- Precisamos expressar as unidades de tempo em horas

$$6 \text{ minutos} = \ 0{,}1 \text{ horas}\,.$$

- A probabilidade requerida é

$$P(X > 0,1) = \int_{0,1}^{\infty} 25e^{-25x} = e^{-25(0,1)} = 0,082\,.$$

(Fazer no quadro)

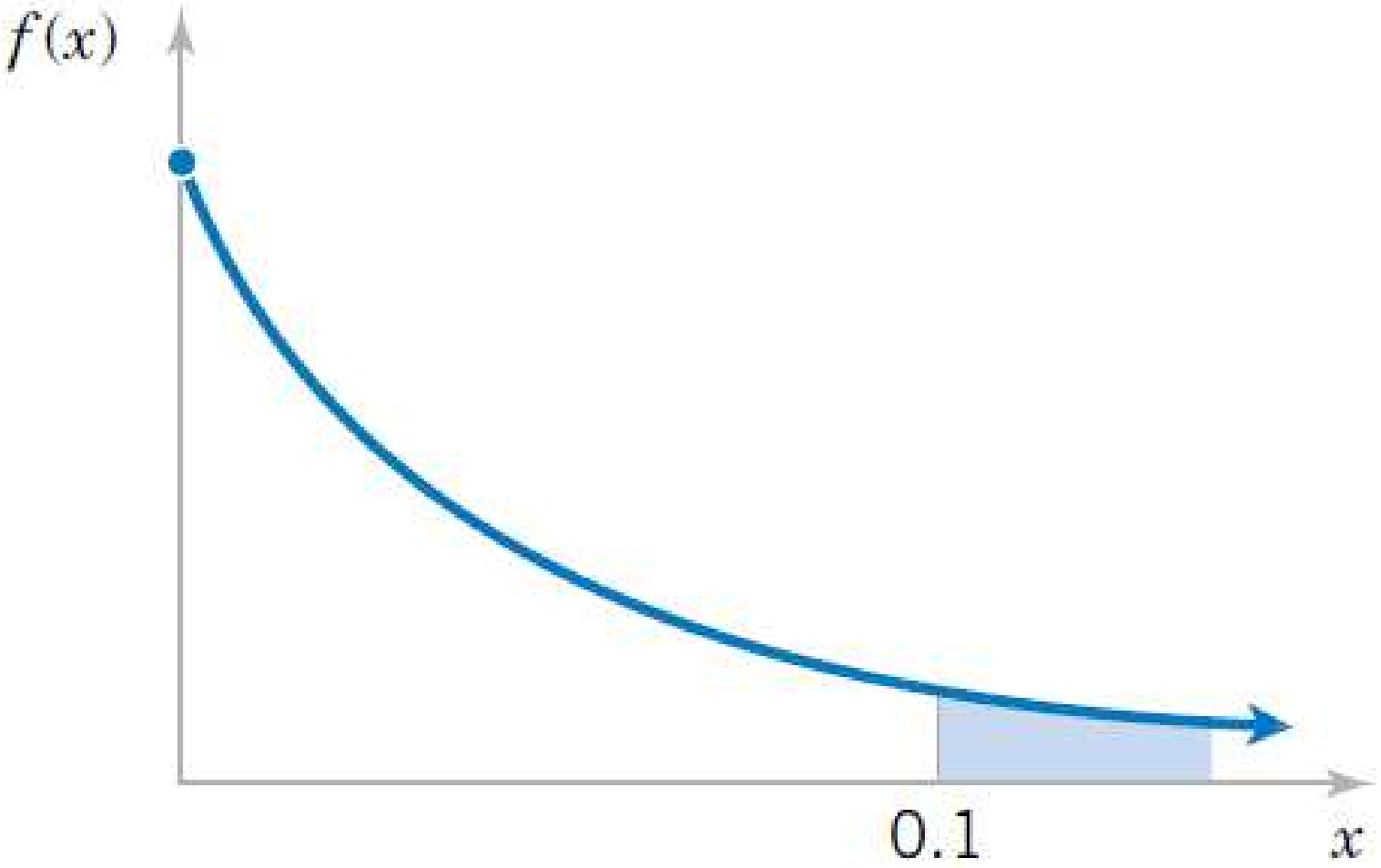


Figure 4-23 Probability for the exponential distribution in Example 4-21.

Exemplo:

- Qual a probabilidade do tempo até a próxima conexão estar entre 2 e 3 minutos?
- Temos que

$$2 \text{ minutos} = 0{,}033 \text{ horas} \qquad 3 \text{ minutos} = 0{,}05 \text{ horas}\,.$$

- A probabilidade requerida é

$$P(0,033 < X < 0,05) = \int_{0{,}033}^{0{,}05} 25e^{-25x}dx = -\,e^{-25x}\Big|_{0{,}033}^{0{,}05}$$

$$= 0,152\,.$$

(Fazer no quadro.)

Exemplo:

- Determine o intervalo tal que a probabilidade de nenhuma conexão ocorrer nesse intervalo é 0,9.
- Queremos saber o comprimento de x tal que

$$P(X > x) = e^{-25x} = 0,9$$

- Tirando logarítmo dos dois lados temos que

$$-25x = ln(0,9) \Rightarrow x = 0,00421 \text{ hora } = 0,25 \text{ minuto } .$$

- O tempo médio até a próxima conexão é

$$E(X) = 1/\lambda = 1/25 = 0,04 \text{ hora } = 2,4 \text{ minutos } .$$

- A variância é

$$Var(X) = 1/\lambda^2 = 1/25^2 .$$

- O desvio padrão é

$$\sqrt{Var(X)} = 1/\lambda = 1/25 = 0,04 \text{ hora } = 2,4 \text{ minutos } .$$

- Vimos no exemplo anterior que
  - a probabilidade de não haver conexões em um intervalo de 6 minutos é 0,082 $\Rightarrow$ não depende do tempo inicial do intervalo.
- O processo de Poisson supõe que os eventos ocorrem uniformemente em um intervalo.
- Não há agrupamento de eventos.
- A probabilidade de que a primeira conexão depois do 12:00 ocorra depois de 12:06 é igual a :
  - probabilidade de que a primeira conexão depois das 15:00 ocorra após 15:06.
- Falta de memória da exponencial.

- O ponto inicial de observação não importa.
- Se houver períodos com picos de conexão o modelo de Poisson não é um bom modelo.
- Pode ser razoável modelar períodos do dia diferentes por processos de Poisson com parâmetros distintos.
- Colocamos um valor de λ alto em alguns períodos e baixo em outros.
- A distribuição exponencial com o λ correspondente pode ser usada para calcular probabilidades de conexão.

Exemplo:

- Considere um contador que detecta partículas raras.
- Seja X o tempo entre detecções de uma partícula.
- X tem distribuição exponencial com

$$E(X) = 1,4 \text{ minuto} .$$

- A probabilidade de detectarmos uma partícula dentro de 30 segundos a partir do começo da contagem é

$$P(X < 0,5 \text{ minuto}) = F(0,5) = 1 - e^{-\frac{0,5}{1,4}} = 0,30 .$$

Exemplo: (continuação)

- Suponha que ligamos o contador e esperamos 3 minutos sem detectar nenhuma partícula.
- Qual a probabilidade de uma partícula ser detectada nos próximos 30 segundos?
- Como já esperamos 3 minutos sentimos que já é tempo suficiente.
- A probabilidade de detecção nos próximos 30 segundos deveria ser menor que 0,3.
- Isso porém não é verdade.

Exemplo: (solução)

- A probabilidade requerida é

$$P(X < 3,5|X > 3) = \frac{P(X < 3,5; X > 3)}{P(X > 3)} = \frac{P(3 < X < 3,5)}{P(X > 3)}.$$

- Temos que

$$P(3 < X < 3,5) = F(3,5) - F(3) = [1 - e^{-\frac{3,5}{1,4}}] - [1 - e^{-\frac{3}{1,4}}]$$

$$= 0,035$$

$$P(X > 3) = 1 - F(3) = e^{-\frac{3}{1,4}} = 0,117\,.$$

- Assim

$$P(X < 3,5|X > 3) = \frac{0,035}{0,117} = 0,3\,.$$

Observações:

- Depois de esperar por 3 minutos
  - a probabilidade de detecção nos próximos 30 segundos
  - é a mesma de detecção nos 30 segundos imediatamente depois de começar a contagem.
- O fato de que você esperou 3 minutos sem detecção não muda a probabilidade de detecção.

Propriedade de Falta de Memória

- Para uma variável aleatória exponencial X

$$P(X < t_1 + t_2 | X > t_1) = P(X < t_2)\,.$$

- Veja a figura abaixo.
- A soma das áreas é um

$$A + B + C + D = 1\,.$$

- A área A é $P(X < t_2)$.
- A área C dividida por $C + D$ é

$$P(X < t_1 + t_2 | X > t_1)\,.$$

- A falta de memória implica que
  - a proporção em A é igual a proporção de C em $C + D$.

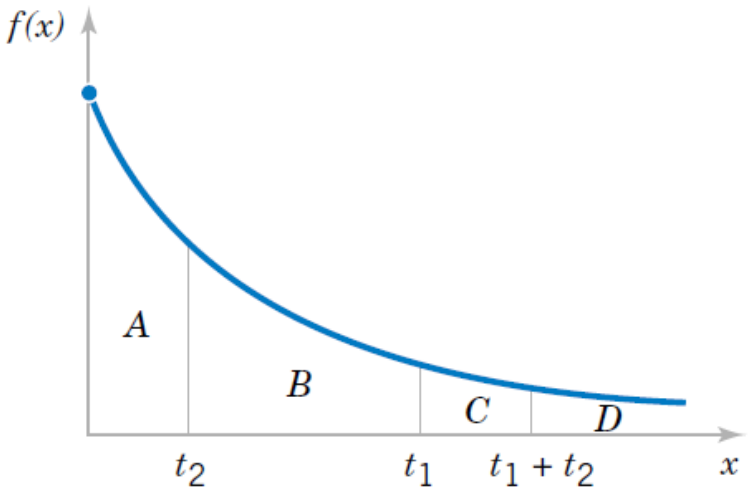


Figure 4-24 Lack of memory property of an exponential distribution.

- A distribuição exponencial é muito usada para modelar tempo de falha de um equipamento.
- Exemplo: o tempo de vida de um *chip*.
- A falta de memória significa que o equipamento não desgasta.
- Independente de quanto tempo o equipamento tenha operado a probabilidade de falha não muda.
- Quando o equipamento sofre desgaste esse não é um modelo bom.
- Uma distribuição usada nesse caso é a Weibull.

Distribuição Gama

- Algumas variáveis aleatórias são sempre não negativas.
- A Distribuição Gama generaliza a Distribuição Exponencial.
- A Distribuição Gama é utilizada na modelagem de problemas que a resposta é estritamente positiva.

- Antes de apresentar a função de densidade de uma variável alaetória, X, que segue uma distribuição Gama. Iremos intorudizir a função Gama:

$$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx \text{ para todo } r > 0.$$

- A função Gama é a generalização da função fatorial. Dessa forma temos:
  - se r inteiro: $\Gamma(r) = (r-1)!$
  - $\Gamma(r) = (r-1)\Gamma(r-1)$
  - $\Gamma(1) = 1$

- Uma variável alaetória X, com função de densidade

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}, \text{ para } x > 0$$

é uma variável alaetória Gama, com parâmetros $\lambda > 0$ e $r > 0$.

- $\int_0^\infty f(x) = 1$ (fazer no quadro)

- Vamos agora calcular média e variância dessa variável.
- Relembre que

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \ .$$

- Usando a definição temos que

$$E(X) = \int_{0}^{\infty} \frac{\lambda^r x^r e^{-\lambda x}}{\Gamma(r)} dx = \frac{r}{\lambda}$$

(fazer no quadro)

- Para a variância, usamos

$$V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - (E(X))^2 .$$

- Logo temos que

$$\int_{0}^{\infty} x^2 f(x)dx = \frac{r(r+1)}{\lambda^2} .$$

- Portanto

$$Var(X) = \frac{r(r+1)}{\lambda^2} - \frac{r^2}{\lambda^2} = \frac{r}{\lambda^2} .$$

(fazer no quadro)

Média e Variância

- Seja X uma variável aleatória com distribuição gama de parâmetros λ e r.
- Temos então que

$$E(X) = \frac{r}{\lambda} \qquad Var(X) = \frac{r}{\lambda^2}\,.$$

- A função de distribuição acumulada de uma variável gama não possui forma fechada.
- A distribuição exponencial é um caso especial da distribuição gama quando $r = 1$.
- Outra distribuição conhecida que também é um caso especial da distribuição gama é a distribuição chi-quadrado (χ^2).

- A distribuição χ^2 é obtida quando $r = \nu/2$ e $\lambda = 1/2$.
- O parâmetro ν é chamado de graus de liberdade.
- A função de densidade de uma variável alaetória χ^2

$$f(x) = \frac{x^{\nu/2-1}e^{-x/2}}{2^{\nu/2}\Gamma(\nu/2)}, \text{ para } x > 0$$

Média e Variância

- Seja X uma variável aleatória com distribuição χ^2 de parâmetro ν.
- Temos então que

$$E(X) = \nu \qquad Var(X) = 2\nu \, .$$

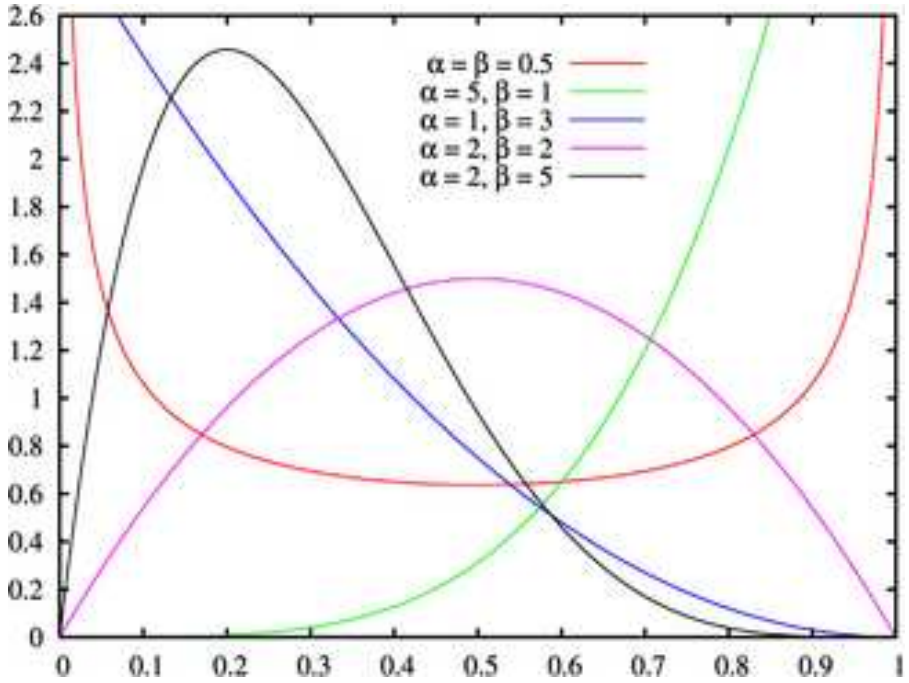
Distribuição Beta

- Uma distribuição que seja contínua, mas limitada a uma faixa finita é útil para modelos de probabilidade.
- Dados de proporções são exemplos de variáveis contínuas ao longo do intervalo $[0, 1]$.

- Uma variável alaetória X, com função de densidade

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \text{ para } 0 \leq x \leq 1$$

é uma variável alaetória beta, com parâmetros $\alpha > 0$ e $\beta > 0$.



Média e Variância

- Seja X uma variável aleatória com distribuição beta de parâmetros α e β.
- Temos então que

$$E(X) = \frac{\alpha}{\alpha + \beta} \qquad Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$