MÉTODOS EXATOS®
CURSOS AVANÇADOS PARA ANÁLISE DE DADOS

Distribuições de Probabilidades

Variáveis aleatórias

- Cálculos probabilísticos envolvendo X são equivalentes aos cálculos com proposições.

Exemplo:

- Um engenheiro florestal está interessado na abundância de ninhos de uma espécie de ave em florestas da sua região. Ele seleciona aleatoriamente uma entre várias áreas de $400\ m^2$ previamente delimitadas e registra o número de ninhos que encontra $(0, 1, 2, 3, \dots)$. Alternativamente, ele poderia expressar seus dados em termos de densidade de ninhos em número por m^2 $(0;\ 0{,}0025;\ 0{,}005;\ 0{,}0075;\ \dots)$.

$A \equiv$ "$na\ área\ florestal\ de\ 400m^2\ selecionada\ existem\ exatamente\ 4\ ninhos$"

$X \equiv$ "$número\ que\ expressa\ a\ quantidade\ de\ ninhos\ encontrados\ em\ uma\ área\ pesquisada\ de\ 400\ m^2$"

$$P(A|H) = P(X = x_j|H)$$

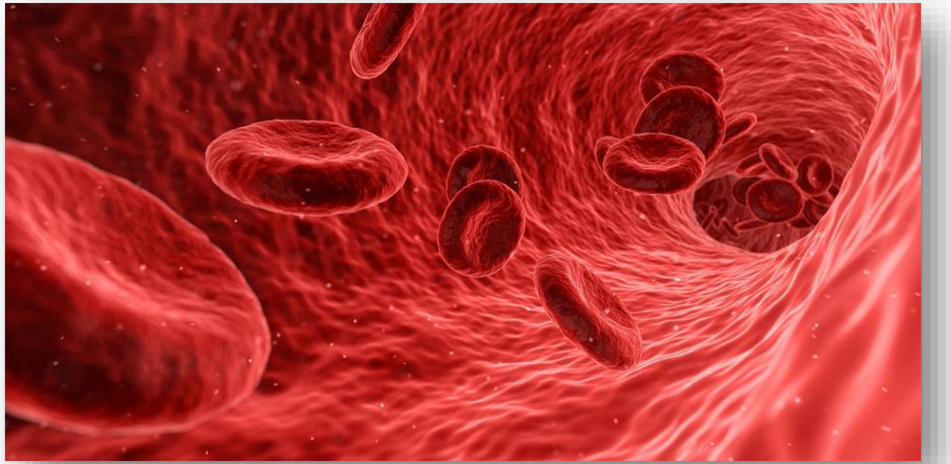
$$P(X = 4|H)$$

Variáveis aleatórias discretas	Variáveis aleatórias contínuas
$Número\ de\ ninhos\ encontrados: (0, 1, 2, 3 \dots)$	$Densidade\ de\ ninhos\ por\ m^2: (0; 0{,}0025; 0{,}005; 0{,}0075 \dots)$

Variáveis aleatórias discretas

- Uma variável aleatória X é uma função que associa as proposições de interesse a um conjunto enumerável (não necessariamente finito) de valores.

Exemplo:



- X caracteriza o número de amostras de sangue avaliadas antes do aparecimento de uma amostra que contém um vírus de interesse.

$X = \{0, 1, 2, \dots\}$

- X representa o número de amostras que contém o vírus em um lote de 10 amostras.

$X = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

- De modo geral:
  - $P(X = x_j | H)$ denota a ***massa de probabilidade*** no ponto x_j.

$$E(X) = \sum_j x_j * P(X = x_j | H) \qquad V(X) = \sum_j [x_j - E(X)]^2 * P(X = x_j | H)$$

Variáveis aleatórias contínuas

- Quando os valores de X formam um subconjunto que compreende pelo menos um intervalo de escala de números reais, a variável aleatória é denominada contínua.
- Variáveis aleatórias contínuas são caracterizadas pela sua **função de distribuição cumulativa** $F(x)$ que é definida para todo número real $X \in (-\infty, +\infty)$.

$$F(x) = P(X \leq x|H)$$

- Define-se $F(-\infty) = 0 \ e \ F(+\infty) = 1.$

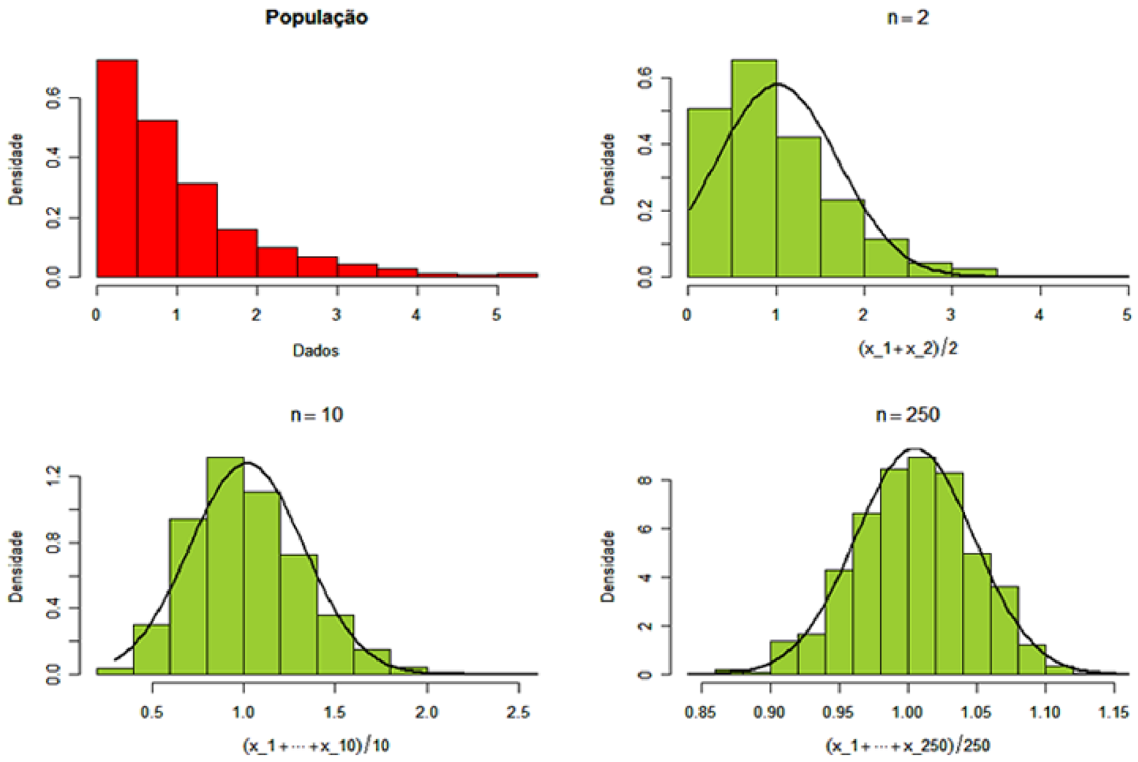
$$E(X) = \int x * f(x) * dx \qquad V(X) = \int [x - E(x)]^2 * f(x) * dx$$

- Para uma variável aleatória contínua a probabilidade para determinado ponto da variável X é zero.

$$P(X|H = x) = 0$$

Distribuições de probabilidade

- Denomina-se de distribuição de probabilidade de alguma variável aleatória, a regra geral que define a função de massa de probabilidade (variável discreta) ou de densidade de probabilidade (variável contínua) para a variável de interesse.



Fonte: portal action

Modelos probabilísticos discretos

MÉTODOS EXATOS
CURSOS AVANÇADOS PARA ANÁLISE DE DADOS

Distribuição Uniforme Discreta

$UD(a,b)$

- Uma v.a. X tem distribuição de probabilidade uniforme discreta entre os números inteiros a e b (inclusive) se sua função de massa de probabilidade (f.m.p.) é definida por:

$$p(x) = \frac{1}{b-a+1} I_{\{x=a,\, a+1,\dots,b-1,\, b\}} \qquad E(X) = \frac{a+b}{2}; \;\; V(X) = \frac{b(b-1)(2b+1) - a(a-1)(2a-1)}{6(b-a+1)} - \frac{(\mathrm{a}+\mathrm{b})^2}{4}$$

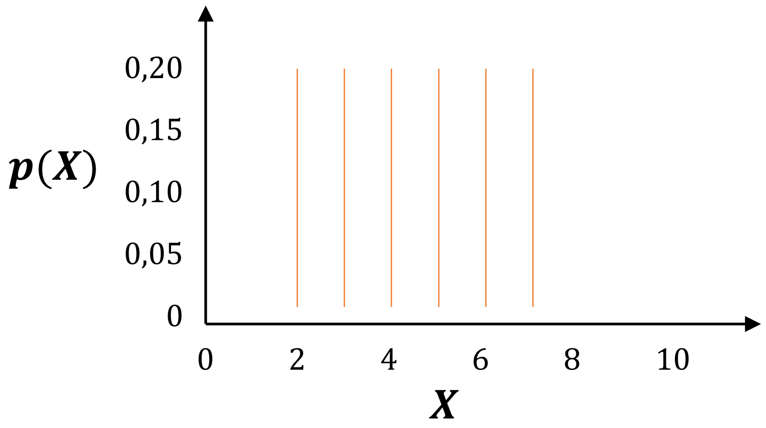
- Este modelo de probabilidade tem dois parâmetros, a e b, os quais são números inteiros tal que $a < b$.
- Script no R para retirada de uma amostra de tamanho $m = 20$ com reposição de uma distribuição com $a \; = \; 2$ e $b \; = \; 7$.

$> \; a <- \; 2$

$> \; b <- \; 7$

$> \; m <- 20$

$> \; sample(a{:}\, b, size = m, replace = T)$



Modelos probabilísticos discretos

Distribuição Binomial

$Bin(n, \theta)$

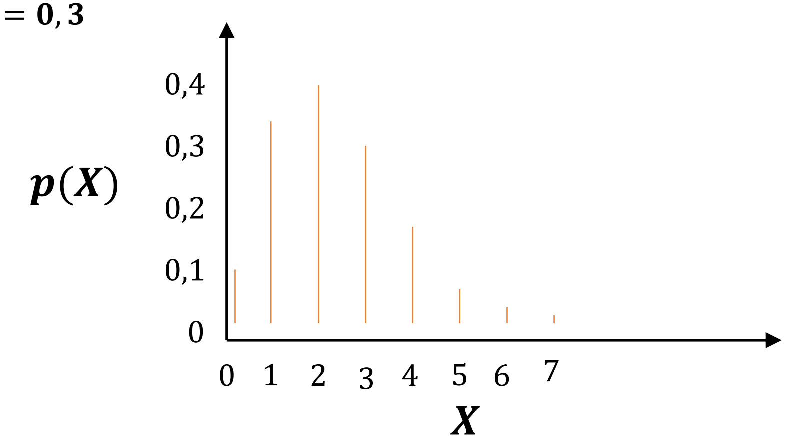
- Uma v.a. X tem distribuição de probabilidade binomial $Bin(n, \theta)$ para números inteiros de 0 a n (inclusive) com $0 < \theta < 1$, se sua f.m.p. é definida por:

$$p(x) = \binom{n}{x} * \theta^x * (1-\theta)^{n-x} * I_{\{x=0,1,2,\ldots,n\}}; \;\; E(x) = n\theta; \;\; V(x) = n\theta(1-\theta); \;\; moda(X) = \lfloor (n+1)\theta \rfloor$$

Exemplo:

- Uma população de animais em que cada um tem igual probabilidade de ser amostrado. Se X é o número de fêmeas entre os n animais amostrados ao acaso e com reposição de uma população com tamanho N em que há M fêmeas, então têm-se uma distribuição binomial com parâmetros n e $\theta = \frac{M}{N}$.

- **A densidade de probabilidade para x = 2 em uma binomial com n =7 e probabilidade $\boldsymbol{\theta = 0,3}$**
  - $> \;\; dbinom(x = 2, size = 7, prob = 0{,}3)$
- **A probabilidade acumulada $\boldsymbol{P(X \leq 2)}$**
  - $> \;\; pbinom(q = 2, size = 7, prob = 0{,}3)$
- **O quantil $\boldsymbol{x_{100q}}$ que satisfaz $\boldsymbol{P(X \leq x_{100q}) = 0,4}$**
  - $> \;\; qbinom(p = 0{,}4, size = 7, prob = 0{,}3)$
- **Uma amostra aleatória de 12 números aleatórios**
  - $> \;\; rbinom(12, size = 7, prob = 0{,}3)$



Modelos probabilísticos discretos

Distribuição Hipergeométrica

$Hip(M, N, n)$

- Uma v.a. X tem distribuição de probabilidade hipergeométrica entre os números inteiros $X_I = \max(0, n - N + M)$ e $x_S = \min(n, M)$ (inclusive) se sua f.m.p. é definida por:

$$p(x) = \frac{\binom{m}{x}\binom{N-M}{n-x}}{\binom{N}{n}} * I_{\{x_I, x_I+1, \dots, x_S-1, x_S\}};\ \ E(x) = n\frac{M}{N};\ \ V(x) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\left(1 - \frac{n-1}{N-1}\right);\ \ moda(X) = \left\lfloor \frac{(n+1)(M+1)}{(N+1)} \right\rfloor$$

Exemplo:

- A densidade de probabilidade para a retirada de uma amostra de tamanho $n = 10$ em que há $x = 2$ casos com a característica de interesse, de uma população que tem $M = 15$ casos com a característica de interesse e $N - M = 30$ casos sem a característica de interesse.

$> \ dhyper(2, 15, 30, 10)$

- **A probabilidade acumulada $P(X \leq 3)$**

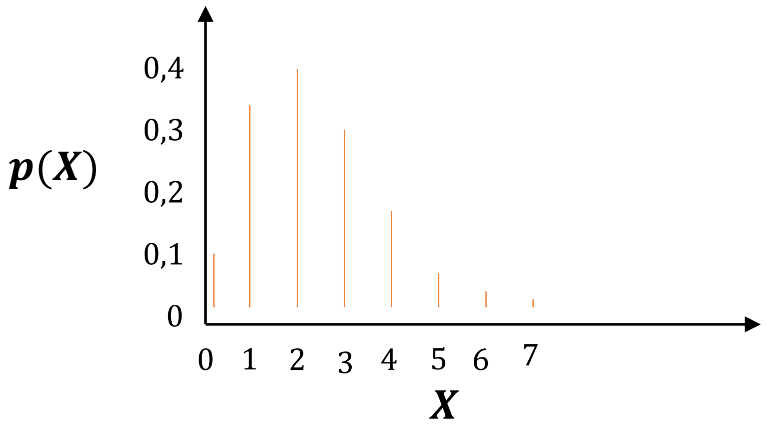
$> \ phyper(3, 15, 30, 10)$

- **O quantil x_{100q} que satisfaz $P\left(X \leq x_{100q}\right) = 0,1$**

$> \ qhyper(0.1, 15, 30, 10)$

- **Uma amostra aleatória de 12 números aleatórios**

$> \ rhyper(12, 15, 30, 10)$



Modelos probabilísticos discretos

Distribuição de Poisson

$Poi(\mu)$

- Uma v.a. X tem distribuição de probabilidade de Poisson sobre os números naturais incluindo zero, se sua f.m.p. é definida por:

$$p(x) = \frac{e^{-\mu}}{x!} * I_{\{x=0,1,2,\dots\}}; E(X) = V(X) = \mu; \ \ moda(X) = \lfloor \mu \rfloor$$

Exemplo:

- O modelo de $Poi(\mu)$ surge quando X é uma contagem por unidade de intervalo de tempo, espaço ou volume e, o parâmetro μ denota o número médio de ocorrências neste intervalo. Quando Y denota uma contagem em k unidades de intervalos então o parâmetro apropriado da distribuição de poisson de Y será $\mu y = k\mu$.

- **A densidade de probabilidade $x = 2$ em uma distribuição de Poisson com média $\mu = 1,8$**

  $> \ \ dpois(2, 1.8)$

- **A probabilidade acumulada $P(X \leq 3)$**

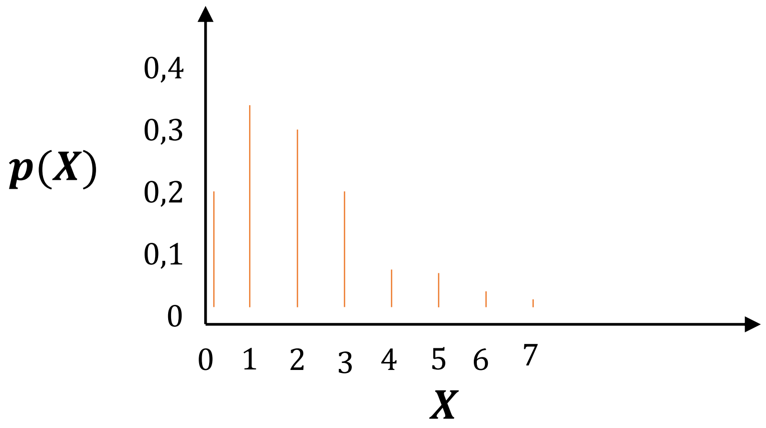
  $> \ \ \text{ppois}(3, 1.8)$

- **O quantil x_{100q} que satisfaz $P(X \leq x_{100q}) = 0,1$**

  $> \ \ \text{q}pois(0.1, 1.8)$

- **Uma amostra aleatória de 8 números aleatórios**

  $> \ \ rpois(8, 1.8)$



Modelos probabilísticos discretos

Distribuição Binomial Negativa

$BinN(a, \theta)$

- Uma v.a. X tem distribuição de probabilidade binomial negativa com parâmetros a e θ no conjunto dos números naturais se:

$$p(x) = \frac{(a+x-1)!}{x!\,(a-1)!}\theta^a(1-\theta) * I_{\{x=0,1,2,\dots\}};\, E(X) = \frac{a(1-\theta)}{\theta};\, V(X) = \frac{a(1-\theta)}{\theta^2};\;\; moda(X) = \begin{cases} 0 & x < 0 \\ \left\lfloor \frac{(a-1)(1-0)}{\theta} \right\rfloor & x \geq 0 \end{cases}$$

Exemplo:

- **A densidade de probabilidade $x = 10$ fracassos antes de $a = 5$ sucessos com probabilidade de sucesso em cada ponto da amostra dada por $\theta = 0,3$**

  $>$ $dnbinom(10, 5, 0.3)$

- **A probabilidade acumulada $P(X \leq 15)$**

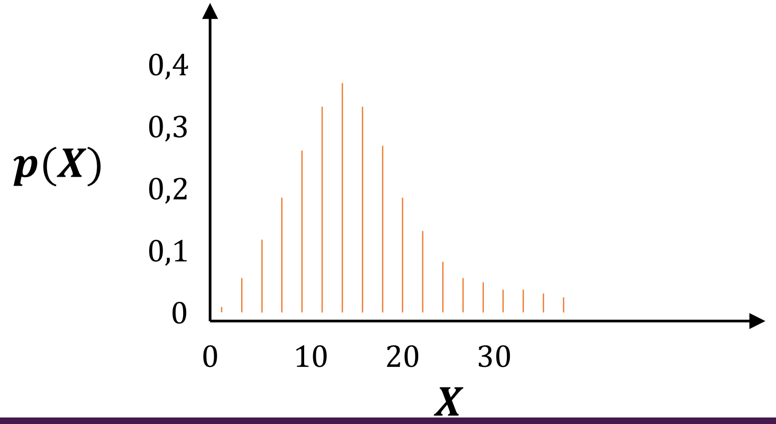
  $>$ $\mathrm{pnbinom}(15, 5, 0.3)$

- **O quantil x_{100q} que satisfaz $P(X \leq x_{100q}) = 0,1$**

  $>$ $\mathrm{qnbinom}(0.1, 5, 0.3)$

- **Uma amostra aleatória de 8 números aleatórios**

  $>$ $rnbinom(8, 5, 0.3)$



Distribuições Contínuas

Distribuição Uniforme

$U(c,d)$

- Uma v.a. X tem distribuição de probabilidade uniforme no intervalo de números reais [c,d] (c<d) se sua função densidade de probabilidade (f.d.p.) é definida por:

$$p(x) = \frac{1}{d-c} * I_{\{c \le x \le d\}}; E(X) = \frac{c+d}{2}; V(X) = \frac{(d-c)^2}{12}$$

Sintaxe no R para:

- **A densidade de probabilidade para $x = 3$ em uma uniforme $U(2,6)$**

  $> \ dunif(3,2,6)$

- **A probabilidade acumulada $P(X \le 3.5)$**

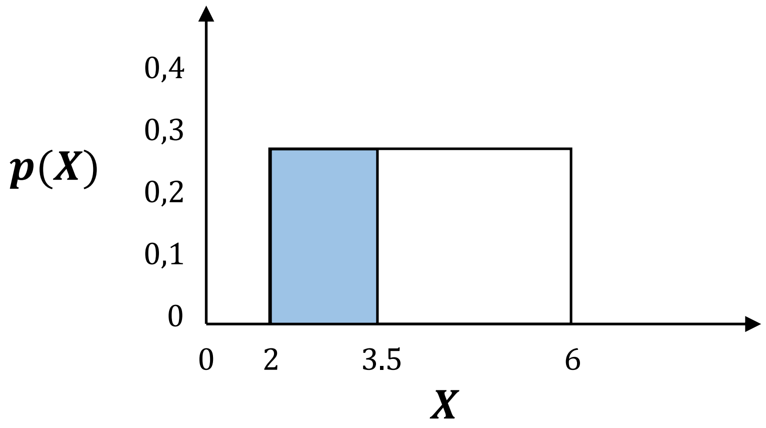
  $> \ \mathrm{punif}(3.5, 2, 6)$

- **O quantil x_{100q} que satisfaz $P(X \le x_{100q}) = 0,1$**

  $> \ \mathrm{qunif}(0.1, 2, 6)$

- **Uma amostra aleatória de 9 números aleatórios**

  $> \ runif(9, 2, 6)$



Distribuições Contínuas

Distribuição Beta

$Beta(\alpha, \beta)$

- Uma v.a. $X \in [0,1]$ tem distribuição de probabilidade beta com parâmetros $\alpha > 0 \; e \; \beta > 0$ se sua f.d.p. é definida por:

$$p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} * I_{\{0 \le x \le 1\}}; E(X) = \frac{\alpha}{\alpha+\beta}; \; V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}; moda(X) = \frac{\alpha-1}{\alpha+\beta-2} (se \; \alpha > 1 \; e \; \beta > 1)$$

Sintaxe no R para:

- **A densidade de probabilidade para $x = 0,2$ em uma $Beta(\alpha = 3, \beta = 10)$**

  $> \; dbeta(0.2, 3, 10)$

- **A probabilidade acumulada $P(X \le 0.2)$**

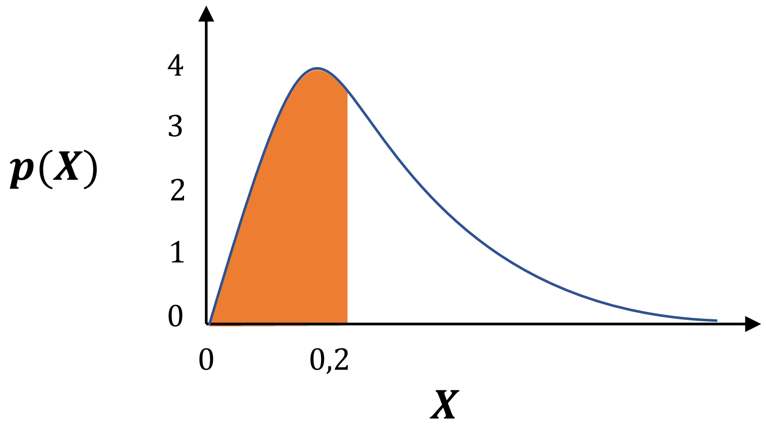
  $> \; \text{pbeta}(0.2, 3, 10)$

- **O quantil x_{100q} que satisfaz $P(X \le x_{100q}) = 0,15$**

  $> \; \text{qbeta}(0.15, 3, 10)$

- **Uma amostra aleatória de 9 números aleatórios**

  $> \; rbeta(9, 3, 10)$

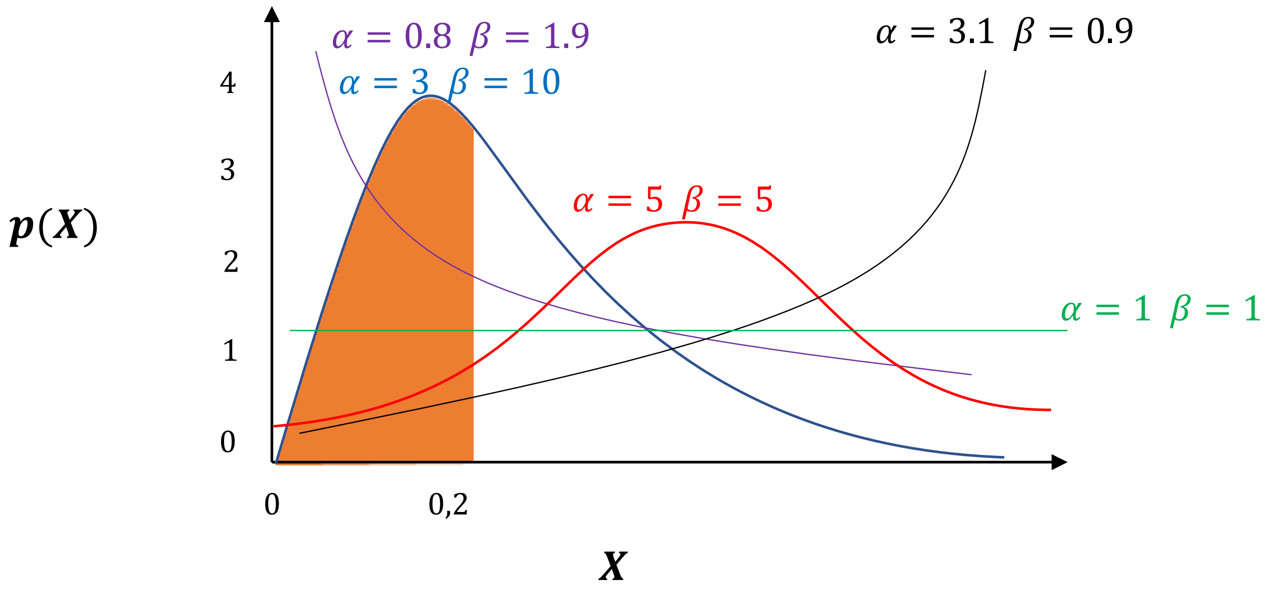


Distribuições Contínuas

Distribuição Beta Generalizada

$BetaG(\alpha, \beta, c, d)$

- A distribuição beta pode ser generalizada de um intervalo unitário [0,1] para um intervalo arbitrário [c,d].



Distribuições Contínuas

Distribuição Exponencial

$Exp(\beta)$

- Uma variável contínua assumindo valores reais não-negativos segue modelo exponencial com parâmetro $\beta > 0$ se sua f.d.p. é:

$$p(x) = \beta e^{\beta x} * I_{\{x \geq 0\}}; E(X) = \frac{1}{\beta};\ V(X) = \frac{1}{\beta^2}$$

Sintaxe no R para:

- **A densidade de probabilidade para $x = 2$ em uma $Exp(\beta = 0,5)$**

  > $dexp(2, 0.5)$

- **A probabilidade acumulada $P(X \leq 3)$**

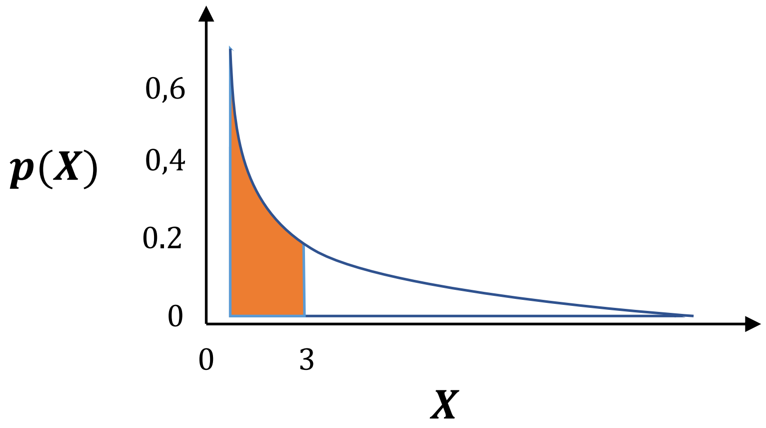
  > $pexp(3, 0.5)$

- **O quantil x_{100q} que satisfaz $P(X \leq x_{100q}) = 0,35$**

  > $qexp(0.35, 0.5)$

- **Uma amostra aleatória de 7 números aleatórios**

  > $rexp(7, 0.5)$



Distribuições Contínuas

Distribuição Gama

$Gama(\alpha, \beta)$

- A variável aleatória X segue um modelo gama com parâmetros $\alpha > 0$ e $\beta > 0$ no intervalo de números reais positivos se sua f.d.p. tem a forma:

$$p(x) = \frac{\beta^x}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} * I_{\{x>0\}}; E(X) = \frac{\alpha}{\beta}; \ V(X) = \frac{\alpha}{\beta^2}; moda(X) = \begin{cases} \frac{\alpha-1}{\beta} & se\ \alpha > 1 \\ 0 & se\ \alpha \leq 1 \end{cases}$$

Sintaxe no R:

- **A densidade de probabilidade para $x = 5$ em uma $Gama(\alpha = 2,\ \beta = 1)$**

  $>\ dgamma(5, 2, 1)$

- **A probabilidade acumulada $P(X \leq 4)$**

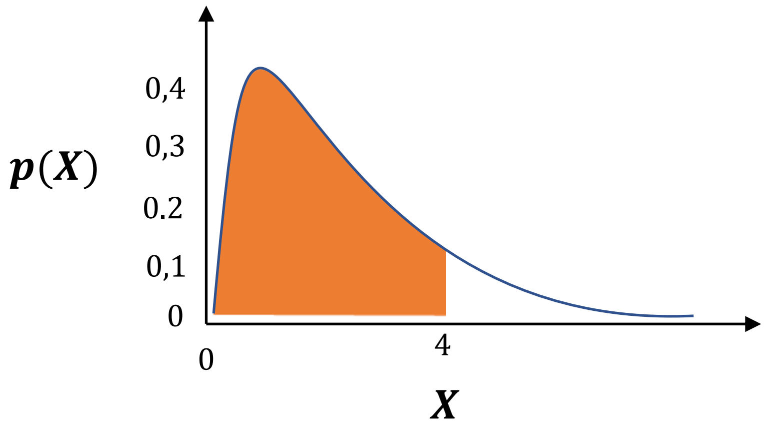
  $>$ p$gamma(4, 2, 1)$

- **O quantil x_{100q} que satisfaz $P(X \leq x_{100q}) = 0,15$**

  $>$ q$gamma(0.15, 2 , 1)$

- **Uma amostra aleatória de 8 números aleatórios**

  $>\ rgamma(8, 2, 1)$



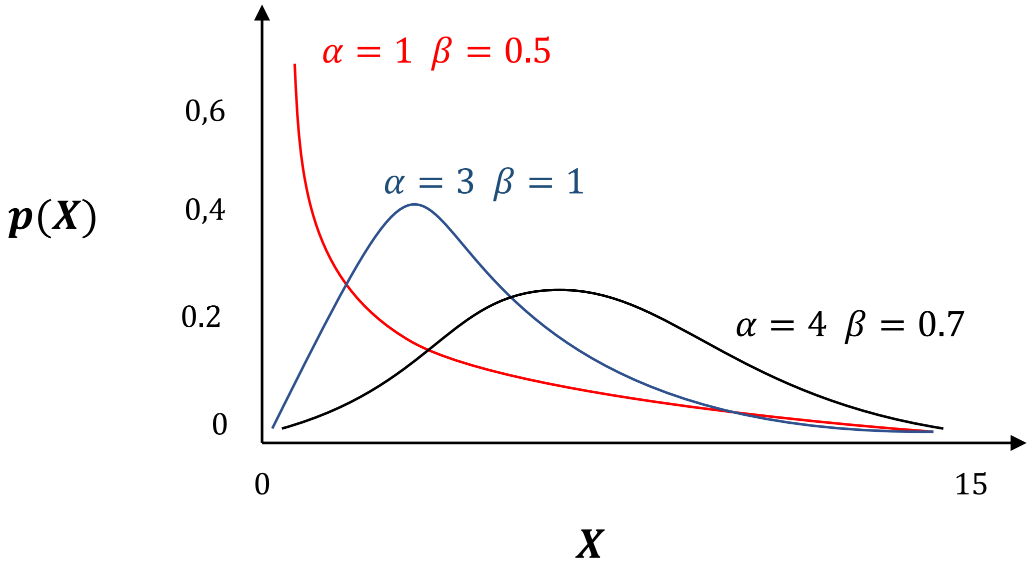
Distribuições Contínuas

Distribuição Gama-Inversa

$GInv(\alpha, \beta)$

- Se X^{-1} tem distribuição ama com parâmetros $\alpha \; e \; \beta$, então X tem a distribuição gama-inversa com esses parâmetros e sua f.d.p. é:

$$p(x) = \frac{\beta^x}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\beta/x} * I_{\{x>0\}}; E(X) = \frac{\beta}{\alpha - 1} \; para \; \alpha > 1; \; V(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)} \; para \; \alpha > 2; moda(X) = \frac{\beta}{\alpha+1}$$



Distribuições Contínuas

Distribuição Normal

$N(\mu, \sigma)$

- Uma variável aleatória X definida sobre o conjunto dos números reais segue o modelo normal com parâmetros $\mu \epsilon (-\infty, +\infty)$ e $\sigma > 0$ se sua f.d.p. é:

$$p(x) = \frac{1}{\sqrt{2\pi * \sigma}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] * I_{\{-\infty<x<+\infty\}}; E(X) = \mu;\ V(X) = \sigma^2; moda(X) = \mu$$

Sintaxe no R:

- **A densidade de probabilidade para $x = 5$ em uma $N(\mu = 3,\ \sigma = 1)$**

  $> \quad dnorm(2, 3, 1)$

- **A probabilidade acumulada $P(X \leq 2)$**

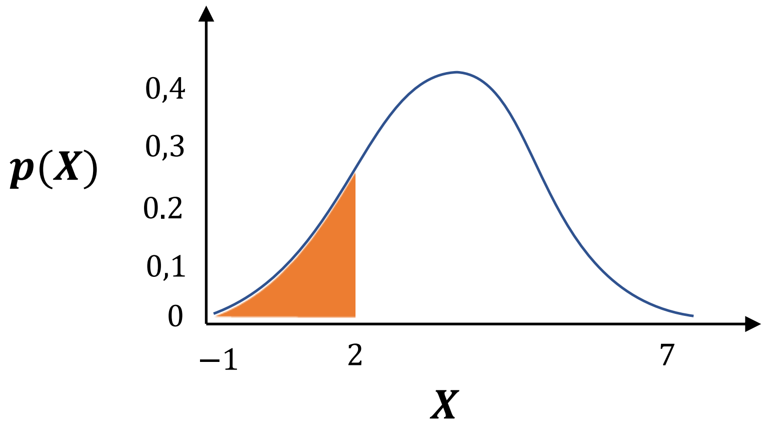
  $>$ pnorm$(2, 3, 1)$

- **O quantil x_{100q} que satisfaz $P(X \leq x_{100q}) = 0,15$**

  $>$ q$norm(0.15, 3, 1)$

- **Uma amostra aleatória de 9 números aleatórios**

  $> \quad rnorm(9, 3, 1)$



Distribuições Contínuas

Distribuição de Student

$St(g, \mu, \sigma)$

- Uma variável aleatória X definida sobre o conjunto dos números reais tem distribuição de Student com g graus de liberdade, média $\mu \in (-\infty, +\infty)$ e parâmetro de escala $\sigma > 0$ se sua f.d.p. é:

$$p(x) = \frac{\Gamma(\frac{g+1}{2})}{\Gamma(\frac{g}{2})\sqrt{g\pi\sigma}}\left[1 + \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]^{-(g+1)/2} * I_{\{-\infty < x < +\infty\}}; E(X) = \mu \; se \; g > 1; \; V(X) = \frac{g}{g-2}; moda(X) = \mu$$

Sintaxe no R:

- **A densidade de probabilidade para $x = 1$ em uma Student padronizada T com g=18 graus de liberdade**

> $dt(1, 18)$

- **A probabilidade acumulada $P(T \leq 1)$**

> pt(1, 18)

- **O quantil x_{100q} que satisfaz $P(X \leq x_{100q}) = 0,15$**

> qt(0.15, 18)

- **Uma amostra aleatória de 9 números aleatórios**

> $rt(9, 18)$