CE085 - Estatística Inferencial
Revisão: Distribuições de Probabilidade

Prof. Wagner Hugo Bonat

Laboratório de Estatística e Geoinformação - LEG
Curso de Bacharelado em Estatatística
Universidade Federal do Paraná - UFPR

20 de agosto de 2018

Conteúdo

Conteúdo

1. Principais modelos discretos
   1.1 Uniforme discreto;
   1.2 Bernoulli;
   1.3 Binomial;
   1.4 Geométrica;
   1.5 Poisson.
2. Principais modelos contínuos
   2.1 Uniforme contínua;
   2.2 Exponencial;
   2.3 Gaussiana.
3. Principais modelos multivariados
   3.1 Multinomial;
   3.2 Normal multivariada.

Principais modelos discretos

Introdução

- Uma v.a. é completamente caracterizada pela sua função de probabilidade.
- Existem muitos tipos de fenômenos aleatórios cuja distribuição de probabilidade foi estudada e especificada com precisão.
- No entanto, para muitos fenômenos os modelos de probabilidade servem como uma aproximação para a verdadeira distribuição de probabilidade.
- Existem uma infinidade de distribuições de probabilidade com aplicações em praticamente todas as áreas da ciência.
- Vamos descrever os principais modelos discretos, contínuos e multivariados.

Modelo Uniforme discreto

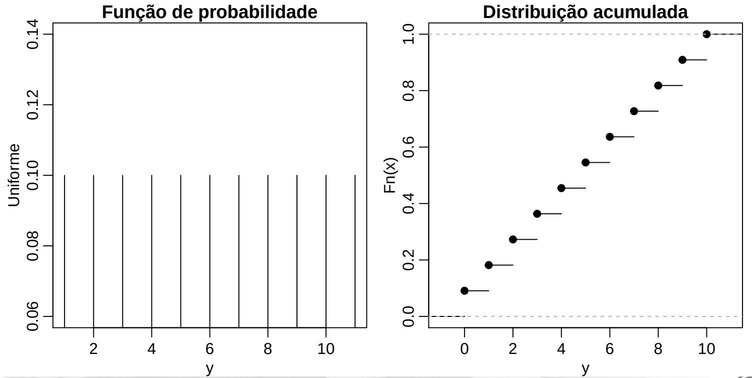
- Y segue o modelo **Uniforme discreto** se sua fp é dada por

$$P(Y = y) = \frac{1}{b - a} \quad \text{para} \quad i = 1, \ldots, k.$$

- Notação $Y \sim U_d(a, b)$.
- Suporte $k \in \{a, a + 1, \ldots, b - 1, b\}$.
- $E(Y) = \frac{a+b}{2}$ e $V(Y) = \frac{(b-a+1)^2-1}{12}$.
- Modela situações onde todos os possíveis valores da v.a são equiprováveis.

Modelo Uniforme discreto

- Graficamente



Exemplo: Modelo Uniforme discreto

- Uma rifa tem 100 bilhetes numerados de 0 a 100. Tenho 5 bilhetes consecutivos numerados de 21 a 25 e meu colega tem outros 5 bilhetes, com os números 1, 11, 29, 68 e 93. Quem tem maior possibilidade de ser sorteado?
- Solução: $Y \sim U_d(0, 100)$.

```
# Probabilidade de eu ganhar
sum(dunif(c(21,22,23,24,25), min = 0, max = 100))

## [1] 0.05

# Probabilidade meu amigo ganhar
sum(dunif(c(1,11,29,68,93), min = 0, max = 100))

## [1] 0.05
```

Modelo Bernoulli

- Y segue o modelo Bernoulli se sua fp é dada por

$$P(Y = y) = p^y(1-p)^{1-y}.$$

- Suporte: $y \in \{0, 1\}$.
- Notação $Y \sim B(n = 1, p)$.
- $E(Y) = p$ e $V(Y) = p(1-p)$.
- Situações de sucesso (p) e fracasso (1-p).
- Exemplo clássico: Lançamento de uma moeda.

Modelo Binomial

- Seja Y o número de sucessos em n ensaios Bernoulli independentes. Assim, Y segue o modelo binomial com parâmetros n e p e sua fp é

$$P(Y = y) = \binom{n}{y} p^y (1 - p)^{n-y}.$$

- Suporte: $y \in \{0, 1, \ldots, n\}$.
- Notação $Y \sim B(n, p)$.
- $E(Y) = np$ e $V(Y) = np(1 - p)$.
- Exemplo: A taxa de imunização de uma vacina é de 80%. Se um grupo de 20 pessoas foram vacinadas, qual a probabilidade
  a) Exatamente quinze pessoas serem imunizadas?

```
dbinom(15, size = 20, prob = 0.8)

## [1] 0.1745595
```

  b) Pelo menos quinze pessoas serem imunizadas?

```
sum(dbinom(15:20, size = 20, prob = 0.8))

## [1] 0.8042078
```

Modelo Geométrico

- Defina Y como sendo o número de fracassos em ensaios Bernoulli anteriores ao primeiro sucesso. Assim, Y tem distribuição Geométrica de parâmetro p e sua fp é dada por

$$P(Y = y) = p(1 - p)^{y-1}.$$

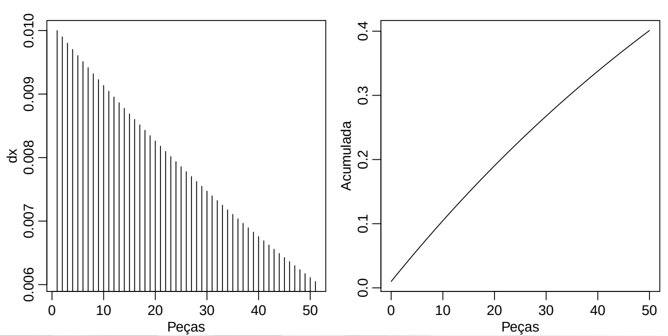
- Suporte: $y \in \{1, 2, \ldots\}$.
- Notação $Y \sim Geo(p)$.
- $E(Y) = 1/p$ e $V(Y) = (1 - p)/p^2$.
- Exemplo: Uma linha de produção está sendo analisada para efeito de controle da qualidade das peças produzidas. Tendo em vista o alto padrão requerido, a produção é interrompida para regulagem toda vez que uma peça defeituosa é observada. Se 0.01 é a probabilidade da peça ser defeituosa, estude o comportamento da variável Y, quantidade de peças boas produzidas antes da primeira defeituosa.

Exemplo: Modelo Geométrico

- $P(Y = k) = 0.01 * 0.99^k$
- Usando o R temos

```
dgeom(x = 1, prob = 0.01)

## [1] 0.0099
```

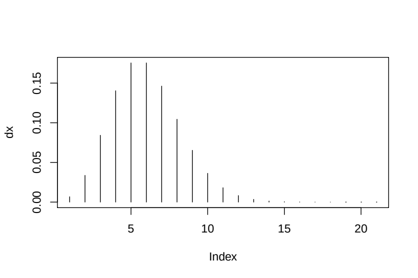


Distribuição de Poisson

- Y tem distribuição Poisson com parâmetro $\lambda > 0$, se sua função de probabilidade é dada por

$$P(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \quad y = 0, 1, 2 \ldots .$$

- Notação $Y \sim Po(\lambda)$.
- $E(Y) = \lambda$ e $V(Y) = \lambda$.



Exemplo: Distribuição de Poisson

- A emissão de partículas radioativas têm sido modelada através de uma distribuição de Poisson com parâmetro $\lambda = 5$, isto é, a taxa média de ocorrência é de 5 emissões a cada minuto. Calcule a probabilidade de haver mais de 2 emissões em um minuto.

$$\begin{aligned} P(Y > 2) &= \sum_{y=3}^{\infty} P(Y = y) = 1 - \sum_{y=0}^{2} P(Y = y) \\ &= 1 - \sum_{y=0}^{2} \frac{\exp\{-5\}5^y}{y!} = 0.875 \end{aligned}$$

- Em R temos

```
1 - ppois(2, lambda = 5)

## [1] 0.875348
```

Exemplo: Distribuição de Poisson

- Engenheiros da companhia telefônica estudam se o modelo de Poisson com taxa de ocorrência de 4.5 chamadas por hora pode ser ajustado ao número N de chamadas interestaduais que chegam por hora, a uma central telefônica, durante o período noturno. Os dados coletados referentes a 650 períodos de uma hora, estão apresentados abaixo. Analise se esta suposição é razoável.

```
obs <- c(9,38,71,115,125,106,79,50,57)
esp <- c(dpois(0:7, lambda = 4.5), 1 - ppois(7, lambda = 4.5))
cbind("Obs" = obs, "Esp" = 650*esp)

##       Obs        Esp
##  [1,]   9   7.220848
##  [2,]  38  32.493815
##  [3,]  71  73.111083
##  [4,] 115 109.666625
##  [5,] 125 123.374953
##  [6,] 106 111.037458
##  [7,]  79  83.278094
##  [8,]  50  53.535917
##  [9,]  57  56.281207
```

Principais modelos contínuos

Distribuição Uniforme contínua

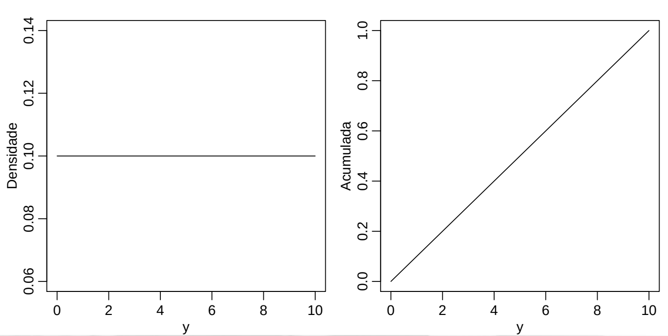
- Uma v.a. Y tem distribuição Uniforme contínua no intervalo $[a, b], a < b$, se sua função densidade de probabilidade é dada por

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{se} \quad a \leq y \leq b; \\ 0 & \text{caso contrário.} \end{cases}$$

- Notação $Y \sim U(a, b)$.
- Neste caso $E(Y) = \frac{a+b}{2}$ e $Var(Y) = \frac{(b-a)^2}{12}$.

Distribuição Uniforme

- Função densidade e acumulada.



Distribuição Exponencial

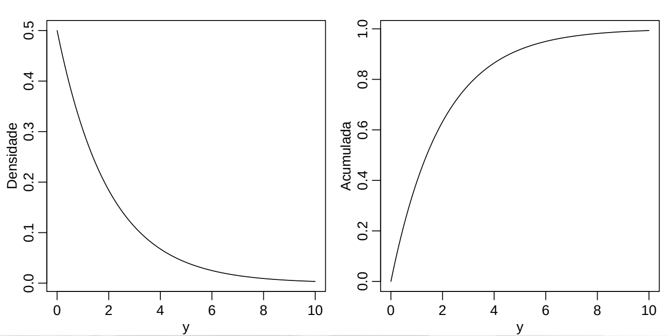
- Uma v.a. contínua Y assumindo valores não negativos, segue o modelo exponencial com parâmetro $\alpha > 0$ se sua densidade é dada por

$$f(x) = \begin{cases} \alpha e^{-\alpha y} & \text{se} \quad y > 0; \\ 0 & \text{caso contrário.} \end{cases}$$

- Neste caso $E(Y) = \mu = \frac{1}{\alpha}$ e $Var(Y) = \frac{1}{\alpha^2}$.
- Notação $Y \sim Exp(\alpha)$.
- $P(a < Y < b) = \int_a^b \alpha e^{-\alpha y} dy = e^{-\alpha a} - e^{-\alpha b}$.

Distribuição Exponencial

- Função densidade e acumulada.



Exemplo: Distribuição Exponencial

- Uma indústria fabrica lâmpadas especiais que ficam em operação continuamente. A empresa oferece a seus clientes a garantia de reposição, caso a lâmpada dure menos de 50 horas. A vida útil dessas lâmpadas é modelada através da distribuição Exponencial com parâmetro $1/8000$. Determine a proporção de troca por defeito de fabricação.

Modelo Normal

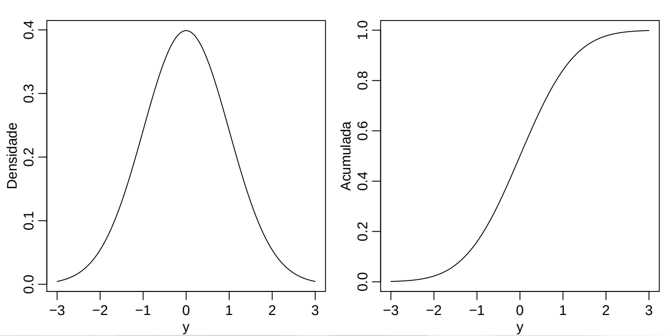
- Uma v.a. contínua Y tem distribuição Normal com parâmetros μ e σ^2, se sua função de densidade é dada por:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad \text{para} \quad -\infty, y < \infty.$$

- Notação $Y \sim N(\mu, \sigma^2)$. item $E(Y) = \mu$ e $Var(Y) = \sigma^2$.
- $P(a \leq Y \leq b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$.

Distribuição Gaussiana

- Função densidade e acumulada.

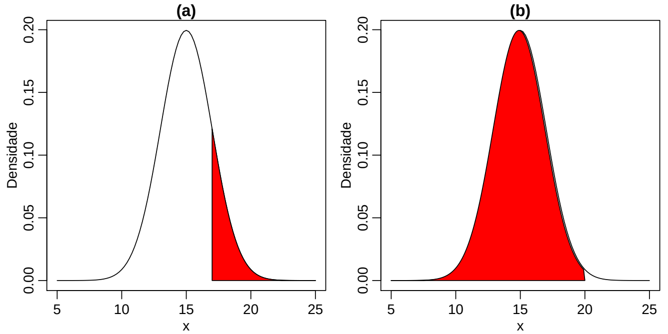


Exemplo: Distribuição Normal

- Doentes sofrendo de certa moléstia são submetidos a um tratamento intensivo cujo tempo de cura foi modelado por uma densidade Normal de média 15 e desvio padrão 2 (em dias). Seja Y o tempo de cura e, portanto temos $Y \sim N(15, 4)$. Calcule a proporção de pacientes que demorão mais de 17 dias para se recuperar. Calcular a probabilidade um paciente ao acaso demorar menos de 20 dias para se recuperar. Qual o número esperado de dias para recuperação?

Exemplo: Distribuição Normal

- Graficamente,



- $P(Y > 17)$

```r
1 - pnorm(17, mean = 15, sd = 2)

## [1] 0.1586553
```

- $P(Y < 20)$

```r
pnorm(20, mean = 15, sd = 2)

## [1] 0.9937903
```

Exemplo: Distribuição Gaussiana

- Um serviço de fiscalização é criado para averiguar se garrafas de um certo refrigerante contém, de fato, o volume especificado pelo fabricante. Para tanto, 10 garrafas do produto são compradas no varejo, em várias regiões da cidade. Cada uma dessas garrafas é esvaziada e o volume de seu conteúdo, que denotaremos por V é aferido. Uma vez obtidos os 10 valores, a média aritmética M é calculada e, se $M < 290$ mililitros (ml), a companhia é multada. Estudos na linha de produção do fabricante mostraram que variações sempre ocorrem, mesmo se as especificações forem seguidas. Por essa razão, considera-se o volume do conteúdo das garrafas como seguindo o modelo Normal, com média $\mu = 300$ ml e desvio-padrão $\sigma = 25$ ml. Gostaríamos de calcular qual é a probabilidade de que o fabricante seja multado injustamente.

Exemplo: Distribuição Normal

- Seja $\bar{V} = \sum_{i=1}^{10} \frac{V_i}{10}$.
- Sabemos que $V \sim N(300, 25^2)$ e quanto a $\bar{V}$?
- Esperança e variância de $\bar{V}$?

$$E(\bar{V}) = E\left(\sum_{i=1}^{10} \frac{V_i}{10}\right) = \sum_{i=1}^{10} \frac{E(V_i)}{10} = \frac{10 * 300}{10} = 300.$$

$$V(\bar{V}) = V\left(\sum_{i=1}^{10} \frac{V_i}{10}\right) = \sum_{i=1}^{10} \frac{V(V_i)}{10} = \frac{10 * 25^2}{10^2} = \frac{25^2}{10}.$$

- Portanto, $\bar{V} \sim N(300, \frac{25^2}{10})$.
- $P(\bar{V} < 290) =$

```r
pnorm(290, mean = 300, sd = 25/sqrt(10))

## [1] 0.1029516
```

Exemplo: Distribuição Gaussiana

- Uma corretora negocia títulos na Bolsa de Valores e utiliza um modelo probabilístico para avaliar seus lucros. Suas aplicações financeiras de compra e venda atingem três áreas: agricultura, indústria e comércio. Admita que o seguinte modelo representa o comportamento do lucro diário da corretora (em milhares):

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

  onde L_A, L_I e L_C representam, os lucros diários nos setores de agricultura, indústria e comércio. As distribuições de probabilidades dessas variáveis aleatórias são $L_A \sim N(3, 4)$, $L_I \sim N(6, 9)$ e $L_C \sim N(4, 16)$. Supondo independência entre os três setores, qual será a probabilidade de um lucro diário acima de 50 mil.

Exemplo: Distribuição Gaussiana

- Esperança de L

$$E(L) = 2E(L_A) + 5E(L_I) + 3E(L_C) = 2 * 3 + 5 * 6 + 3 * 4 = 48.$$

- Variância de L

$$V(L) = 2^2V(L_A) + 5^2V(L_I) + 3^2V(L_C) = 4*4 + 25*9 + 9*16 = 385.$$

- $P(L > 50)$

```
1 - pnorm(50, mean = 48, sd = sqrt(385))

## [1] 0.4594063
```

Principais modelos multivariados

Distribuição Multinomial

- Suponha um experimento que tem k possíveis resultados $\{O_1, O_2, \ldots, O_k\}$ repetidos independentemente n vezes. Seja $p_1, p_2, \ldots, p_k$ a probabilidade de $\{O_1, O_2, \ldots, O_k\}$, respectivamente. Seja Y_i o número de vezes que o resultado O_i ocorre em n ensaios. A distribuição conjunta das v.a. $Y_1, \ldots, Y_k$ é

$$p(y_1, \ldots, y_n) = \frac{n!}{y_1! \ldots y_n!} p_1^{y_1} \ldots p_k^{y_k}.$$

Exemplo: Distribuição Multinomial

- Suponha que o anúncio de rendimento de uma empresa tenha três possibilidades:
  1. Reação positiva dos preços das ações (30% chances).
  2. Sem reação dos preços das ações (50% chances).
  3. Reação negativa (20% chances).
- Suponha que 4 empresas anunciam seus rendimentos ($n = 4$). Seja X o número de reações positivas, Y o número de não reação e Z reações negativas. Encontre a distribuição conjunta de X, Y e Z e calcule $P(X + Y > Z)$.

Exemplo: Distribuição Multinomial

- Distribuição conjunta

$$p(x,y,z) = \frac{4!}{x!y!z!}(0.30)^x(0.50)^y(0.20)^z, \quad x+y+z=4.$$

x	y	z				
		0	1	2	3	4
0	0	0	0	0	0	0.0016
0	1	0	0	0	0.0160	0
0	2	0	0	0.0600	0	0
0	3	0	0.1000	0	0	0
0	4	0.0625	0	0	0	0
1	0	0	0	0	0.0096	0
1	1	0	0	0.0720	0	0
1	2	0	0.1800	0	0	0
1	3	0.1500	0	0	0	0
1	4	0	0	0	0	0
2	0	0	0	0.0216	0	0
2	1	0	0.1080	0	0	0
2	2	0.1350	0	0	0	0
2	3	0	0	0	0	0
2	4	0	0	0	0	0
3	0	0	0.0216	0	0	0
3	1	0.0540	0	0	0	0
3	2	0	0	0	0	0
3	3	0	0	0	0	0
3	4	0	0	0	0	0
4	0	0.0081	0	0	0	0
4	1	0	0	0	0	0
4	2	0	0	0	0	0
4	3	0	0	0	0	0
4	4	0	0	0	0	0

Exemplo: Distribuição Multinomial

- $P(X + Y \geq Z) = 0.9728.$

x	y	z = 0	1	2	3	4
0	0	0	0	0	0	0.0016
0	1	0	0	0	0.0160	0
0	2	0	0	0.0600	0	0
0	3	0	0.1000	0	0	0
0	4	0.0625	0	0	0	0
1	0	0	0	0	0.0096	0
1	1	0	0	0.0720	0	0
1	2	0	0.1800	0	0	0
1	3	0.1500	0	0	0	0
1	4	0	0	0	0	0
2	0	0	0	0.0216	0	0
2	1	0	0.1080	0	0	0
2	2	0.1350	0	0	0	0
2	3	0	0	0	0	0
2	4	0	0	0	0	0
3	0	0	0.0216	0	0	0
3	1	0.0540	0	0	0	0
3	2	0	0	0	0	0
3	3	0	0	0	0	0
3	4	0	0	0	0	0
4	0	0.0081	0	0	0	0
4	1	0	0	0	0	0
4	2	0	0	0	0	0
4	3	0	0	0	0	0
4	4	0	0	0	0	0

Distribuição Normal multivarida

- Um vetor aleatório contínuo $\mathbf{Y}$ ($p \times 1$) tem distribuição Normal multivariada se sua densidade é dada por

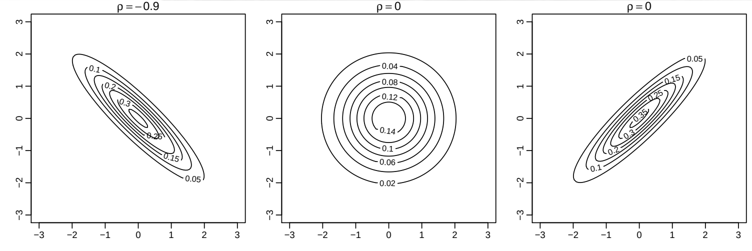
$$f(\mathbf{y}) = \left(\frac{1}{2\pi}\right)^{p/2} |\Sigma|^{-1/2} \exp\{-\frac{1}{2}(\mathbf{y} - \mu)^\top \Sigma^{-1}(\mathbf{y} - \mu)\}.$$

- Notação $\mathbf{Y} \sim N(\mu, \Sigma)$. $E(\mathbf{Y}) = \mu$ e $V(\mathbf{Y}) = \Sigma$.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix} \quad \text{e} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \dots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \sigma_p^2 \end{bmatrix}$$

Normal bivariada

- Normal bivariada diferentes correlações.



Distribuições marginais e condicionais

- Distribuição marginal: A distribuição marginal de um vetor aleatório Gaussiano é também Gaussiano. Sendo,

$$Y_i \sim N(\mu_i, \sigma_i^2).$$

- Distribuições condicionais: A distribuição condicional de um vetor aleatório Gaussiano é também Gaussiana. Sendo,

$$Y_2|Y_1 \sim N(\mu_{2|1}, \Sigma_{2|1}),$$

onde

$$\mu_{2|1} = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(y_1 - \mu_1) \quad \text{e} \quad \Sigma_{2|1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.$$