Distribuições de probabilidade

Visão geral, principais modelos discretos

Prof. Me. Lineu Alberto Cavazani de Freitas

Departamento de Estatística
Laboratório de Estatística e Geoinformação

UNIVERSIDADE FEDERAL DO PARANÁ

Introdução

Introdução

- Vimos anteriormente conceitos a respeito de **variáveis aleatórias** e funções que atribuem probabilidades aos possíveis valores das variáveis (fp e fdp).
- Na prática temos a **evidência empírica**, isto é, o que o dado mostra.
- Com base na evidência empírica precisamos chegar a **funções** que atribuam probabilidades aos possíveis resultados das variáveis aleatórias.
- O processo para obtenção destas funções pode ser complexo.

Modelos de probabilidade

- Existe um conjunto de **distribuições de probabilidade** que podem ser utilizadas para descrever fenômenos: **os modelos**.
- De forma geral, os modelos são **comportamentos teóricos** que vão servir como instrumento para estudar fenômenos aleatórios com características comuns.
- A ideia é que em vez de construir a função de probabiliade ou densidade de probabilidade para o problema possamos usar uma expressão genérica.
- Tentamos obter a melhor combinação entre dado e modelo.

Modelos de probabilidade

- Um modelo possui **parâmetros**: quantidades desconhecidas que assumem valores dentro de um intervalo (espaço paramétrico) que definem características da distribuição.

- Estes parâmetros são estimados por meio dos dados.

- Se o modelo se adequar bem aos dados, utilizamos o modelo para determinar probabilidades, estimar parâmetros, testar hipóteses, avaliar efeito de outras variáveis, fazer predições, etc.

- Em alguns casos sabemos a priori o modelo que descreve bem o fenômeno.

- Em outros casos precisamos encontrar este modelo.

Modelos de probabilidade

- Existem diversos modelos disponíveis.
- Muitos destes modelos aplicáveis a problemas similares.
- E diferentes modelos podem apresentar vantagens e desvantagens.
- Veremos alguns dos principais modelos **discretos** e **contínuos** com foco na **definição** de cada um deles, suposições, fp ou fdp (expressão e comportamento), média, variância e também exemplos.

Modelos de probabilidade

Alguns dos modelos que serão discutidos:

- Principais modelos discretos:
  - Uniforme discreta.
  - Bernoulli/Binomial.
  - Geométrica/Binomial Negativa.
  - Poisson.
  - Hipergeométrico.
- Principais modelos contínuos:
  - Uniforme contínua.
  - Normal.
  - Exponencial.

Modelo Uniforme Discreto

Modelo Uniforme Discreto

Definição

Uma variável aleatória Y segue o modelo Uniforme Discreto se todos os m valores do suporte ocorrem com mesma probabilidade.

Notação

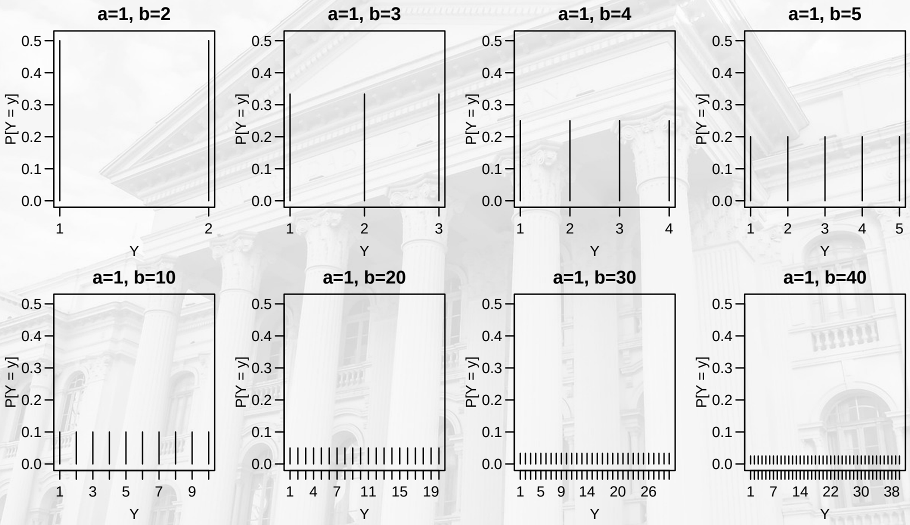
- $Y \sim \text{UD}(m)$

Função de probabilidade

$$P[Y = y_i] = \frac{1}{m}, \qquad i = 1, \cdots, m.$$

- m representa o número de possíveis desfechos da variável aleatória.
- Se Y tem suporte definido no conjunto de números inteiros consecutivos a, $a+1$, $a+2$, $\ldots$, $b-1$, b, para $a < b$. Dessa forma, o número de valores é $m = b - a + 1$, cada um com probabilidade $p = 1/m$
  - $E(Y) = \frac{b+a}{2}$
  - $V(Y) = \frac{(b-a+1)^2-1}{12}$

Modelo Uniforme Discreto



Exemplo

Considere que o experimento aleatório de interesse é o lançamento de um dado honesto e será avaliada a face voltada para cima.

$$Y : \text{face do dado.}$$

$$Y \sim \text{UD}(m = 6)$$

- Temos que

Y	$P[Y = y]$
1	$1/m = 1/6$
2	$1/m = 1/6$
3	$1/m = 1/6$
4	$1/m = 1/6$
5	$1/m = 1/6$
6	$1/m = 1/6$

Modelo Bernoulli

Modelo Bernoulli

Definição
Uma variável aleatória Y segue o modelo Bernoulli se assume apenas os valores 0 ("fracasso") ou 1 ("sucesso").

Notação

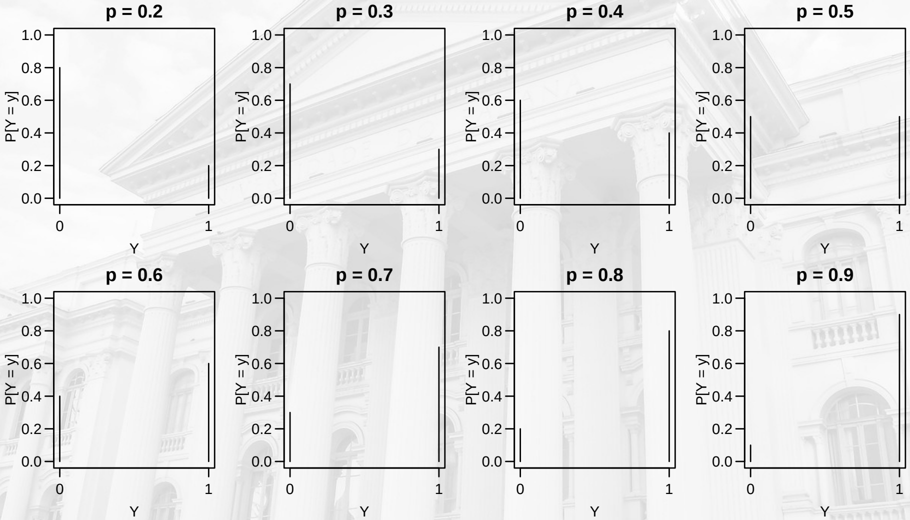
- $Y \sim \text{Ber}(p)$

Função de probabilidade

$$P[Y = y] = p^y(1-p)^{1-y}, \qquad y \in \{0, 1\}$$

- p representa a probabilidade de sucesso: $0 \leq p \leq 1$.
- $E(Y) = p$
- $V(Y) = p(1-p)$

Modelo Bernoulli



Exemplo

Considere o lançamento de uma moeda em que a probabilidade de cara é 0,7 e a probabilidade de coroa é 0,3.

$$Y: \text{observar cara.}$$

$$Y \sim \text{Ber}(p = 0{,}7)$$

$$Y = \begin{cases} 1, & \text{cara (sucesso)} \\ 0, & \text{coroa (fracasso)} \end{cases}$$

Y	$P[Y = y]$
0	$1 - p = 0{,}3$
1	$p = 0{,}7$

Modelo Binomial

Modelo Binomial

Definição

- A variável aleatória Y representa o número de sucessos em n ensaios de Bernoulli, Y pode assumir os valores $0, 1, \ldots, n$.
- Cada tentativa é o desfecho de uma variável dicotômica.
- As tentativas devem ser independentes.
- A probabilidade de sucesso em cada tentativa é constante.
- Os parâmetros são o número de tentativas (n) e a probabilidade de sucesso (p).

Modelo Binomial

Notação

- $Y \sim B(n,p)$

Função de probabilidade:

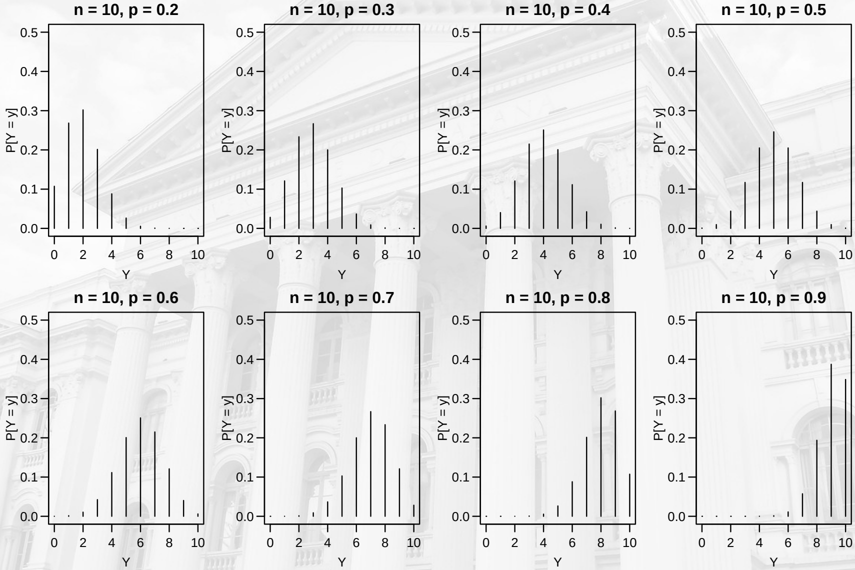
$$P[Y = y] = \binom{n}{y} p^y (1-p)^{n-y}, \qquad y \in \{0, 1, \ldots, n\}$$

em que

$$\binom{n}{y} = \frac{n!}{y!(n-y)!}$$

- $E(Y) = np$.
- $V(Y) = np(1-p)$.

Modelo Binomial



Exemplo

Suponha o experimento de lançar a moeda viciada do exemplo anterior 10 vezes.

1. Qual a probabilidade de obter 10 caras em 10?
2. Qual a probabilidade de obter 3 caras em 10?
3. Qual a probabilidade de obter entre 6 e 8 caras em 10?

Exemplo

Y: Número de caras obtidos em 10 lançamentos.

$$Y \sim B(n = 10, p = 0{,}7)$$

1. Qual a probabilidade de obter uma cara em 10?
   - $P(Y = 10) = 0{,}0282$
2. Qual a probabilidade de obter 3 caras em 10?
   - $P(Y = 3) = 0{,}0090$
3. Qual a probabilidade de obter entre 6 e 8 caras em 10?
   - $P(6 \leq Y \leq 8) = 0{,}7004$

Modelo Geométrico

Modelo Geométrico

Definição

- A variável aleatória Y representa o **número de fracassos** até o primeiro sucesso, Y pode assumir os valores $0, 1, 2, \ldots$.
- Cada tentativa é o desfecho de uma variável dicotômica.
- As tentativas devem ser independentes.
- A probabilidade de sucesso em cada tentativa é constante.
- O parâmetro é a probabilidade de sucesso (p).

Modelo Geométrico

Notação

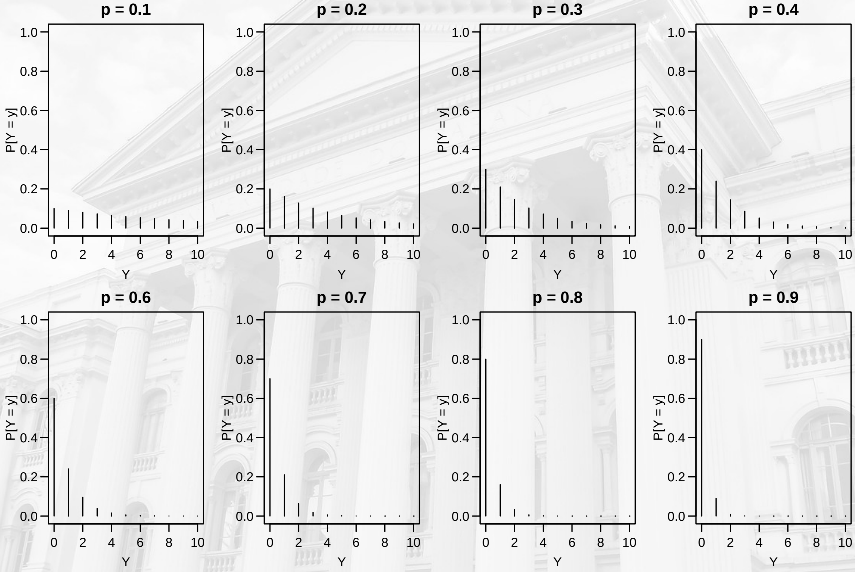
- $Y \sim G(p)$

Função de probabilidade:

$$P[Y = y] = (1 - p)^y p, \qquad y \in \{0, 1, 2, \ldots\}$$

- $E(Y) = \frac{1-p}{p}$ (número médio de fracassos).
- $V(Y) = \frac{1-p}{p^2}$.

Modelo Geométrico



Exemplo

Suponha o experimento de lançar a moeda viciada do exemplo anterior repetidamente até obter a primeira cara.

1. Qual a probabilidade de não ter nenhum fracasso até a primeira cara?
2. Qual a probabilidade de ter dois fracassos até a primeira cara?
3. Qual a probabilidade de ter mais que dois fracassos até a primeira cara?

Exemplo

Y: Número de fracassos até a primeira cara, $y \in \{0, 1, 2, \ldots\}$

$$Y \sim G(p = 0{,}7).$$

1. Qual a probabilidade de não ter nenhum fracasso até a primeira cara?
   - $P(Y = 0) = (1 - 0.7)^0 \times 0{,}7 = 0{,}7$
2. Qual a probabilidade de ter dois fracassos até a primeira cara?
   - $P(Y = 2) = (1 - 0{,}7)^2 \times 0{,}7 = 0{,}063$
3. Qual a probabilidade de ter mais que dois fracassos até a primeira cara?
   - $P(Y > 2) = 1 - P(Y \leq 2) = 1 - \{P(Y = 0) + P(Y = 1) + P(Y = 2)\} = 1 - 0{,}7 - 0{,}21 - 0{,}06 = 0{,}03$

Modelo Geométrico (definição alternativa)

Definição alternativa

- A variável aleatória Y representa o **número de tentativas** até o primeiro sucesso.

Neste caso, a função de probabilidade é dada por

$$P[Y = y] = (1-p)^{y-1}p, \ y \in \{1, 2, \dots\}.$$

- $E(Y) = \frac{1}{p}$ (número médio de tentativas).
- $V(Y) = \frac{1-p}{p^2}$.

Modelo Binomial Negativo

Modelo Binomial Negativo

Definição

- A variável aleatória Y representa o **número de fracassos** até o k-ésimo sucesso, Y pode assumir os valores $0, 1, 2, \ldots$.
- Cada tentativa é o desfecho de uma variável dicotômica.
- As tentativas devem ser independentes.
- A probabilidade de sucesso em cada tentativa é constante.
- O parâmetro é a probabilidade de sucesso (p).

Modelo Binomial Negativo

Notação

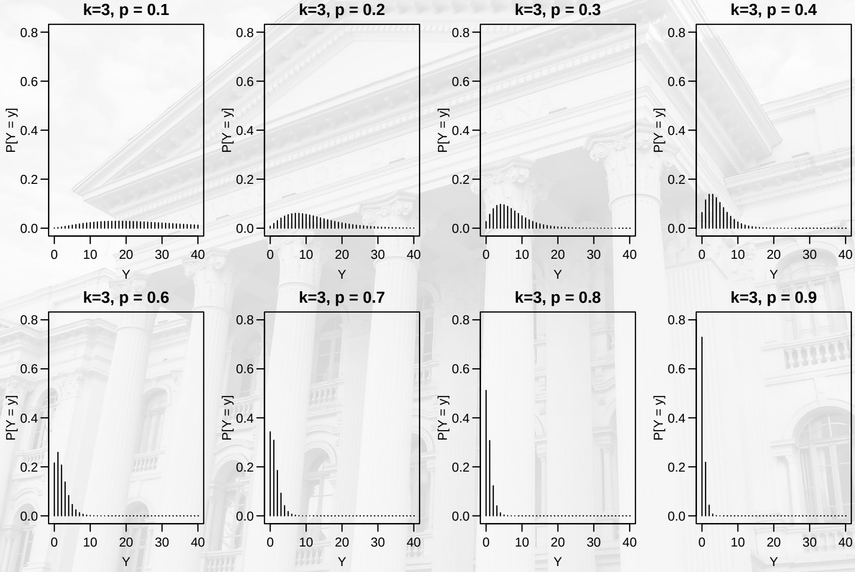
- $Y \sim BN(k,p)$

Função de probabilidade:

$$P(Y = y) = \binom{y+k-1}{y} p^k (1-p)^y, \qquad y \in \{0, 1, 2, \ldots\}$$

- $E(Y) = \frac{k(1-p)}{p}$ (número médio de fracassos).
- $V(Y) = \frac{k(1-p)}{p^2}$.
- Modelo Geométrico é um caso particular quando $k = 1$.

Modelo Binomial Negativo



Exemplo

Suponha o experimento de lançar a moeda viciada do exemplo anterior repetidamente até obter a terceira cara.

1. Qual a probabilidade de não ter nenhum fracasso até a terceira cara?
2. Qual a probabilidade de ter 5 fracassos até a terceira cara?
3. Qual a probabilidade de ter mais que 2 fracassos até a terceira cara?

Exemplo

$$Y\text{: Número de fracassos até a terceira cara, } y \in \{0, 1, 2, \ldots\}$$

$$Y \sim BN(k = 3, p = 0{,}7).$$

1. Qual a probabilidade de não ter nenhum fracasso até a terceira cara?
   - $P(Y = 0) = \binom{0+3-1}{0} \times 0{,}7^3 \times (1 - 0{,}7)^0 = 0{,}343$
2. Qual a probabilidade de ter 5 fracassos até a terceira cara?
   - $P(Y = 5) = \binom{5+3-1}{5} \times 0{,}7^3 \times (1 - 0{,}7)^5 = 0{,}0595$
3. Qual a probabilidade de ter mais que 2 fracassos até a terceira cara?
   - $P(Y > 2) = 1 - P(Y \leq 2) = 1 - \{P(Y = 0) + P(Y = 1) + P(Y = 2)\} = 1 - 0{,}343 - 0{,}3087 - 0{,}1323 = 0{,}2169$

Modelo Binomial Negativo (definição alternativa)

Definição alternativa

- A variável aleatória Y representa o **número de tentativas** até o k-ésimo sucesso.

Neste caso, a função de probabilidade é dada por

$$P[Y = y] = \binom{y-1}{k-1}(1-p)^{y-k}p^k,\ y \in \{k, 2, \ldots\}.$$

- $E(Y) = \frac{k}{p}$ (número médio de tentativas).
- $V(Y) = \frac{k(1-p)}{p^2}$.

Modelo hipergeométrico

Modelo hipergeométrico

Definição

- Suponha o problema de amostrar sem reposição um número de elementos de um conjunto em que dois resultados são possíveis (sucesso ou fracasso).

- Conjunto de $m + n$ objetos.
- $m > 0$ são considerados como sucesso.
- $n > 0$ são considerados como fracasso.
- Sorteia-se de r objetos $r < m + n$, ao acaso e sem reposição.
- A variável aleatória Y é o número de objetos do tipo sucesso selecionados.

Modelo hipergeométrico

Notação

- $Y \sim \text{HG}(m, n, r)$.

Função de probabilidade:

$$P[Y = y] = \frac{\binom{m}{y}\binom{n}{r-y}}{\binom{m+n}{r}},$$

em que $y \in \{max(0, r-n), \cdots, min(r, m)\}$.

- Seja $p = m/(m+n)$
- $\mu = E(Y) = rp$.
- $\sigma^2 = Var(Y) = rp(1-p)\left(\frac{(m+n)-r}{(m+n)-1}\right)$.

Modelo hipergeométrico: domínio

O domínio definido por $\max(0, r-n) \leq y \leq \min(r, m)$, garante que a amostra seja fisicamente possível dentro das limitações da população.

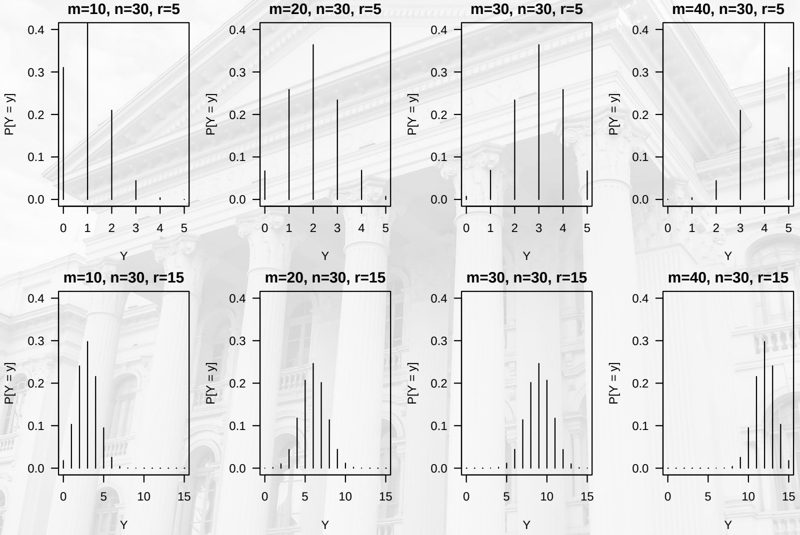
1. **O Limite Superior:** $\min(r, m)$
   O número de sucessos na sua amostra (y) não pode ultrapassar dois valores fundamentais:
   - r (**Tamanho da amostra**): Não podemos ter mais sucessos do que o número total de itens que retiramos. Se sorteamos 5 bolas, é impossível ter 6 sucessos.
   - m (**Sucessos disponíveis**): Não podemos retirar mais sucessos do que o total existente na população. Se só existem 3 bolas brancas na urna, nunca teremos 4 sucessos na amostra.
   - **Conclusão**: y deve ser menor ou igual ao **menor** desses dois valores.

Modelo hipergeométrico: domínio

O domínio definido por $\max(0, r - n) \leq y \leq \min(r, m)$, garante que a amostra seja fisicamente possível dentro das limitações da população.

2. **O Limite Inferior:** $\max(0, r - n)$
   Este limite é menos intuitivo, mas segue a mesma lógica de restrição física:
   - 0: Obviamente, não podemos ter um número negativo de sucessos.
   - $r - n$ (**Obrigação de sucesso**): Este caso ocorre quando a amostra (r) é maior do que o número de fracassos (n) disponíveis.
     - **Exemplo:** Uma urna com 10 bolas: 8 são sucessos (m) e apenas 2 são fracassos (n). Retiramos $r = 5$ bolas. Mesmo que retiremos todos os fracassos possíveis (2), teremos pelo menos 3 sucessos para completar a amostra de 5.
     - Nesse caso, $r - n = 5 - 2 = 3$. Logo, o valor mínimo de y é 3.

Modelo hipergeométrico



Exemplo

Suponha que em um parque estime-se que hajam 200 macacos de determinada espécie. Destes macacos, 50 foram capturados, marcados e soltos no parque. Se forem amostrados 10 macacos, qual a probabilidade de encontrar pelo menos um macaco marcado?

Exemplo

Y: número de macacos marcados em 10 reamostrados.

$$Y \sim \text{HG}(m = 50, n = 150, r = 10)$$

- Objetos do tipo sucesso (m): macacos marcados, $m = 50$.
- Objetos do tipo fracasso (n): macacos não marcados, $n = 150$.
- Tamanho da amostra (r): 10.

$$P(Y \geq 1) = 1 - (Y < 1) = 1 - 0{,}0521 = 0{,}9479$$

Modelo Poisson

Modelo Poisson

Definição

- Distribuição usada para modelar problemas de contagens.

- Algumas suposições devem ser atendidas:
  1. Número de eventos em um domínio (como tempo e espaço).
  2. Taxa de ocorrência constante (probabilidade de um evento é a mesma para qualquer unidade de mesma dimensão).
  3. Independência entre domínios disjuntos.
  4. Taxa proporcional ao tamanho do domínio.

- A variável aleatória Y representa o número de ocorrências em um intervalo.

- Y pode assumir os valores $0, 1, \ldots$.

Modelo Poisson

Notação

- $Y \sim P(\lambda)$.

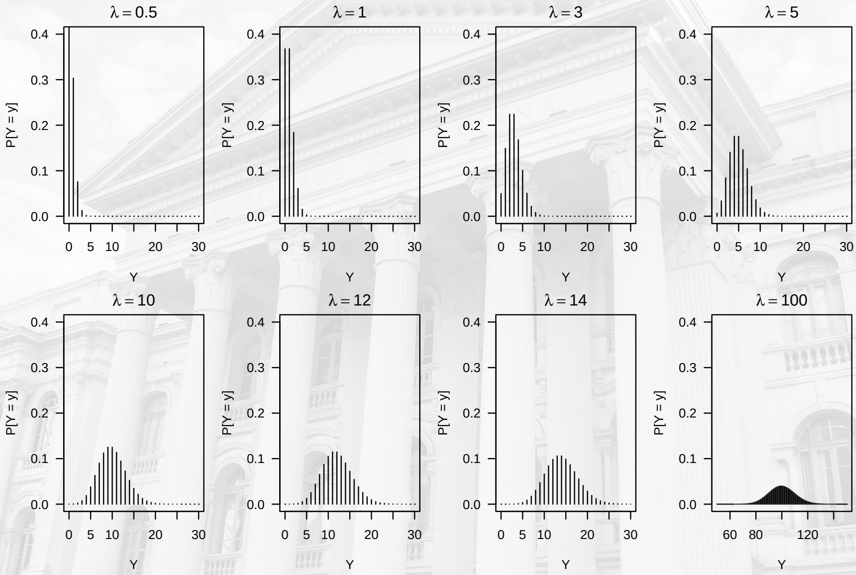
Função de probabilidade:

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

em que $\lambda > 0$ representa a taxa média de ocorrências em um intervalo.

- $\mu = E(Y) = \lambda$.
- $\sigma^2 = Var(Y) = \lambda$.

Modelo Poisson



Exemplo

Suponha que o número de requisições feitas a determinado site do governo se comporta segundo uma distribuição de Poisson com taxa de 5 requisições por minuto.

1. Qual a probabilidade de não haver requisições em um minuto?
2. Qual a probabilidade de haver 10 requisições em um minuto?
3. Qual a probabilidade de haver entre 3 e 6 requisições em um minuto?
4. Qual é o valor esperado de requisições por minuto?

Exemplo

Y: Número de requisições por minuto.

$$Y \sim P(\lambda = 5)$$

1. Qual a probabilidade de não haver requisições em um minuto?
   - $P(Y = 0) = 0{,}0067$
2. Qual a probabilidade de haver 10 requisições em um minuto?
   - $P(Y = 10) = 0{,}0181$
3. Qual a probabilidade de haver entre 3 e 6 requisições em um minuto?
   - $P(3 \leq Y \leq 6) = 0{,}6375$
4. Qual é o valor esperado de requisições por minuto?
   - $\mu = E(Y) = \lambda = 5$

Considerações finais

Considerações

- Existem muitos outros modelos na literatura.
  - Generalizações de modelos clássicos.
  - Modelos para outros fins.
- Devemos estar atentos aos pressupostos e parametrizações.
- Outros modelos discretos:
  - Geométrica.
  - Binomial Negativa.
- Outros modelos contínuos:
  - Lognormal.
  - Gama.
  - Weibull.
  - Beta.
- Modelos multivariados:
  - Distribuição multinomial.
  - Normal multivariada.
  - Distribuição de Dirichlet.

O que foi visto:

- Modelos de probabilidade.
- Alguns modelos discretos.

Próximos assuntos:

- Modelos de probabilidade.
- Alguns modelos contínuos.