Modelos Probabilísticos

Modelos Lineares Generalizados

Luan Fiorentin

12 de maio de 2021

Sumário

Introdução

Introdução

- **Probabilidade** forma a base para entender os modelos de regressão.
- **Modelos probabilísticos** descrevem a variável aleatória em termos de probabilidade.
- Reconhecer o **suporte** da variável permite escolher de forma adequada um modelo.
- Há modelos **adequados** para variáveis aleatórias discretas e contínuas.

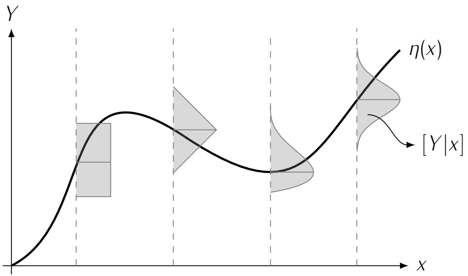

Figura 1. Modelo de probabilidade.

Distribuição Bernoulli

Distribuição Bernoulli

- Variável aleatória Y tem distribuição de Bernoulli se apresenta apenas **dois resultados possíveis**, representados por 0 (fracasso ou negativo) e 1 (sucesso ou positivo).
- O parâmetro $0 < p < 1$ é a **probabilidade de sucesso**.
- A **função de probabilidade** é

$$P(Y = y) = \begin{cases} 1 - p, & \text{se "fracasso" ou } y = 0 \\ p, & \text{se "sucesso" ou } y = 1, \end{cases}$$
$$= p^y \cdot (1 - p)^{1-y}, \quad y \in \{0, 1\}.$$

- Denotamos por $Y \sim \text{Ber}(p)$.
- Y apresenta **média** e **variância** dadas por
  - $\mu = \text{E}(Y) = p$.
  - $\sigma^2 = \text{V}(Y) = p \cdot (1 - p)$.

Exemplos de v.a. com distribuição Bernoulli

1. Ataque de macaco-prego em uma árvore: se árvore está atacada $\rightarrow y = 1$.
2. Espécie de interesse econômico: se é de interesse comercial $\rightarrow y = 1$.
3. Sobrevivência de árvore em floresta: se árvore está viva $\rightarrow y = 1$.
4. Germinação de semente: se semente germinar $\rightarrow y = 1$.
5. Indivíduo dominante: se árvore for dominante $\rightarrow y = 1$.

Nem sempre o que é considerado como "sucesso" ($y = 1$) é algo necessariamente bom.

Gráfico da distribuição Bernoulli

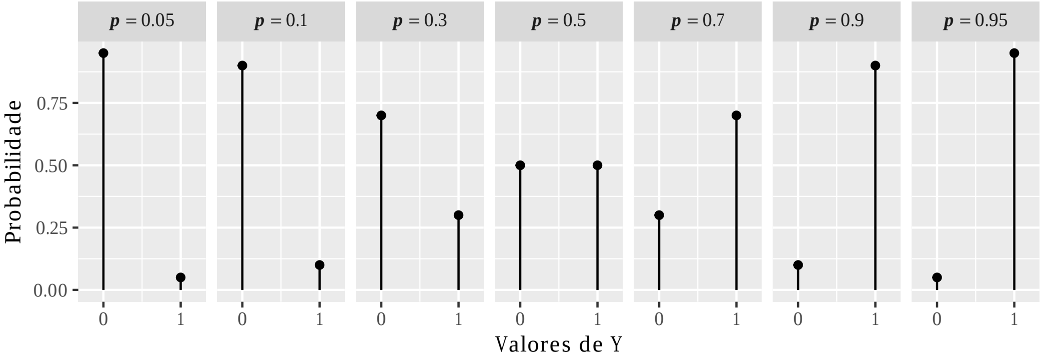

Figura 2. Gráficos para a distribuição de Bernoulli.

Distribuição Binomial

Distribuição Binomial

- **Características** de uma v.a. com distribuição Binomial:

Um experimento aleatório consiste em $n > 0$ **tentativas de Bernoulli**, de modo que

1. As tentativas são **independentes**.
2. A probabilidade de um sucesso em cada tentativa, $0 < p < 1$, é **constante**.
3. Cada tentativa apresenta apenas um de dois resultados possíveis (0: *fracasso* ou 1: *sucesso*).

Distribuição Binomial

- Variável aleatória Y, que é igual ao **número de tentativas que resultam em sucesso**, é uma v.a. Binomial com parâmetros p e n.
- A **função de probabilidade** de Y é

$$P(Y = y) = \binom{n}{y} \cdot p^y \cdot (1 - p)^{n-y}, \quad y \in \{0, \ldots, n\}.$$

- Denotamos por $Y \sim \text{Bin}(n, p)$.
- Y apresenta **média** e **variância** dadas por
  - $\mu = \text{E}(Y) = n \cdot p$.
  - $\sigma^2 = \text{V}(Y) = n \cdot p \cdot (1 - p)$.
- Lembre-se que $\binom{n}{y} = \frac{n!}{y!(n-y)!}$.

Gráfico da distribuição Binomial

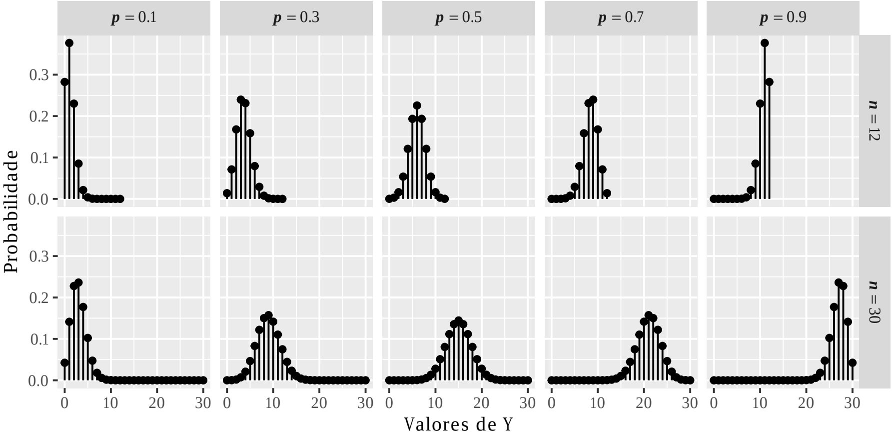

Figura 3. Gráficos para a distribuição Binomial.

Exemplos de v.a. com distribuição Binomial

1. Número de ocorrências de árvores mortas em 100 árvores.
2. Número de sementes germinadas em um saco com 1000 sementes.
3. Número de árvores atacadas por pragas florestais em 1500 árvores.
4. Número de árvores bifurcadas em uma unidade amostral com 150 árvores.
5. Número de espécies de interesse comercial em uma unidade de exploração com 150 espécies.

As suposições precisam ser atendidas para o modelo ser válido.

Distribuição de Poisson

Distribuição de Poisson

- **Características** de uma v.a. com distribuição de Poisson:

Uma variável aleatória Y apresenta distribuição de Poisson se as seguintes suposições são atendidas

1. Variável é o número de eventos que ocorrem em um certo domínio, como intervalo de tempo, área, volume, distância ou outra unidade de medida (pessoa, website, bairro).
2. A probabilidade de que um evento ocorra é a mesma para qualquer unidade de mesma dimensão, ou seja, $\rightarrow$ taxa de ocorrência constante.
3. O número de eventos que ocorrem em uma unidade é independente do número de eventos que ocorrem em outra unidade mutuamente exclusiva.
4. A probabilidade do evento ocorrer em um subdomínio é igual para todos os possíveis subdomínios de mesma dimensão e é proporcional a dimensão do subdomínio.

Distribuição Poisson

- Variável aleatória Y tem distribuição de Poisson com parâmetro (**taxa**) $\lambda > 0$ se sua função de probabilidade é expressa por

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

- Denotamos por $Y \sim \text{Pois}(\lambda)$.
- Y apresenta **média** e **variância** dadas por
  - $\mu = \text{E}(Y) = \lambda$.
  - $\sigma^2 = \text{V}(Y) = \lambda$.
- Se $\mu = \sigma^2$, então a relação configura uma **equidispersão**.

Gráfico da distribuição Poisson

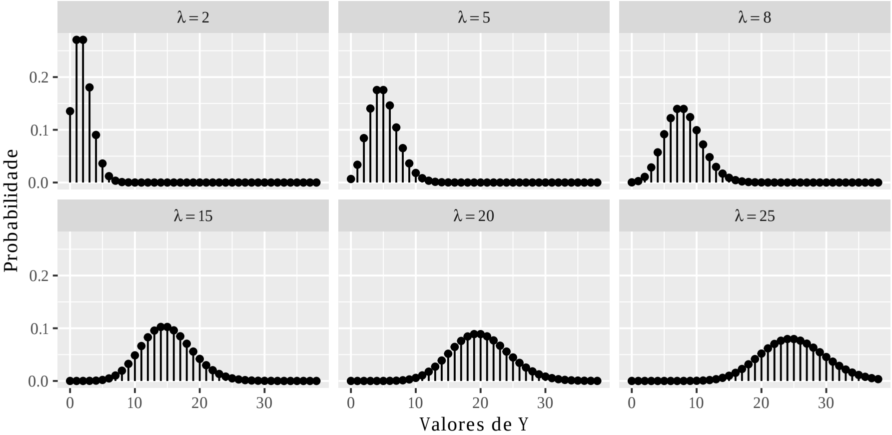

Figura 4. Gráficos para a distribuição de Poisson.

Exemplos de v.a. com distribuição de Poisson

1. Número de árvores mortas por hectare.
2. Número de árvores colhidas no primeiro desbaste.
3. Número de árvores bifurcadas por unidade amostral.
4. Número de espécies por hectare em uma floresta nativa.
5. Número de árvores atacadas por macaco-prego por hectare.

As suposições precisam ser atendidas para o modelo ser válido.

Tipos dispersão e relação média variância

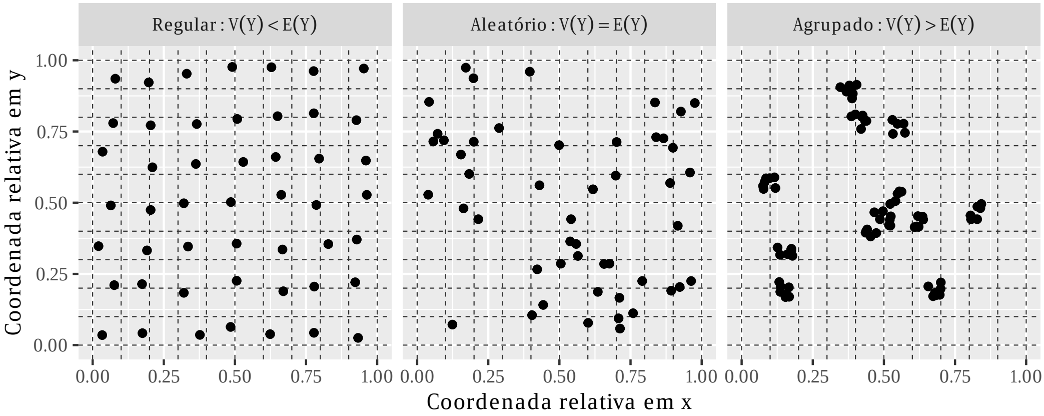

Figura 5. Padrões de dispersão espacial e a relação de média e variância da contagem.

Distribuição Poisson para dimensão não constante

- **Propriedade da Poisson**: se a dimensão da unidade de observação em que ocorrem os eventos for multiplicada por uma constante $t > 0$, então a v.a. resultante terá distribuição de Poisson com a função de probabilidade da forma

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

- Y apresenta **média** e **variância** dadas para unidades de tamanho t por
  - $\mu = \mathrm{E}(Y) = \lambda t$.
  - $\sigma^2 = \mathrm{V}(Y) = \lambda t$.

Distribuição Normal

Distribuição Normal

- A distribuição Normal é a **mais importante** distribuição contínua.
- **Principais** motivos:
  1. Modela um grande número de variáveis aleatórias contínuas.
  2. Serve de aproximação para diversas distribuições contínuas e discretas.
  3. Desempenha papel importante na Teoria Estatística, fundamentando a obtenção de inferências em diferentes contextos.

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}$$

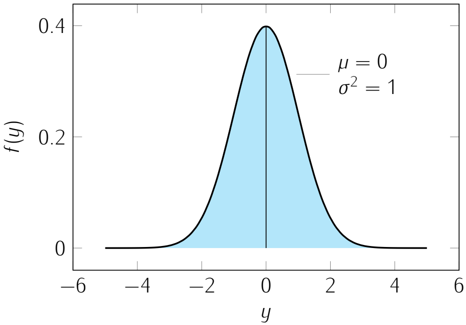

Figura 6. Distribuição Gama.

Distribuição Normal

- **Características** de uma v.a. com distribuição Normal:
  - Variável aleatória contínua não limitada.
  - Possui comportamento simétrico em relação a μ.
  - Ponto máximo (moda) de $f(y)$ é o ponto $x = \mu$.
  - Pontos de inflexão da função são $\mu - \sigma$ e $\mu + \sigma$.
  - A curva é assintótica em relação ao eixo y.
  - Resultado do efeito de muitos fatores de pequena contribuição.

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}$$

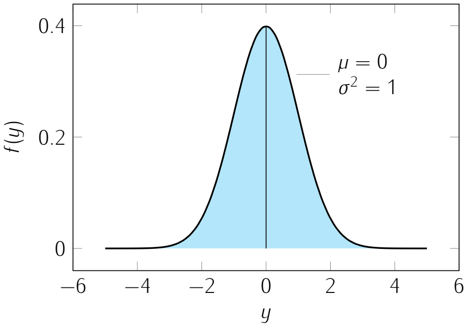

Figura 7. Distribuição Gama.

Distribuição Normal

- Variável aleatória Y tem distribuição Normal com parâmetros $\mu \in \mathbb{R}$ e $\sigma^2 > 0$ se sua **função densidade de probabilidade** é dada por

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

- Denotamos por $Y \sim \mathrm{N}(\mu, \sigma^2)$.
- Y apresenta **média** e **variância** dadas por
  - $\mathrm{E}(Y) = \mu$.
  - $\mathrm{V}(Y) = \sigma^2$.
- Média (μ) e variância (σ^2) são os **parâmetros da distribuição**:

$$\mu = \frac{\sum_{i=1}^{n} y_i}{n} \quad \text{e} \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2.$$

Gráfico da distribuição Normal

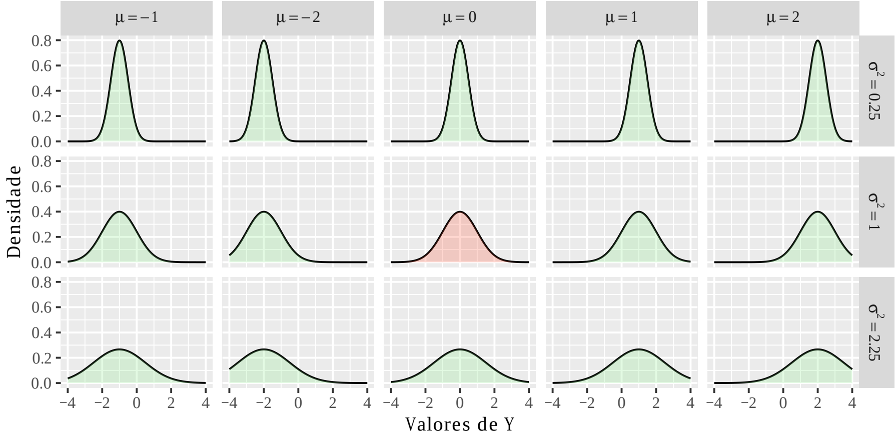

Figura 8. Gráficos para a distribuição Normal.

Distribuição Normal Padrão

- Probabilidades associadas à distribuição Normal não podem ser obtidas analiticamente (mas sim numericamente), pois a integral correspondente **não tem forma fechada**.
- Pode-se usar *softwares* estatísticos ou consultar **tabelas da distribuição Normal Padrão** ($\mu = 0, \sigma^2 = 1$).
- Basta determinar probabilidades na **Normal Padrão** para calcular probabilidades de qualquer v.a. Normal.
- A **consulta** a tabela da Normal Padrão se justifica pelo fato que se $Y \sim \mathrm{N}(\mu, \sigma^2)$, então

$$Z = \frac{Y - \mu}{\sigma} \sim \mathrm{N}(\mu = 0, \sigma^2 = 1).$$

  Então, Z representa o caso em que $Y \sim \mathrm{N}(0, 1)$.
- Dessa forma, pelo **caminho contrário**, tem-se que

$$Y = \mu + \sigma \cdot Z \sim \mathrm{N}(\mu, \sigma^2).$$

Exemplos de v.a. com distribuição Normal

1. Altura total de árvores.
2. Diâmetro à altura do peito.
3. Volume individual de árvores.
4. Volume total de árvores por hectare.
5. Volume de madeira atacada por pragas florestais.

Distribuição Gama

Distribuição Gama

- **Características** de uma v.a. com distribuição Gama
  - Seja $Y_{\mathrm{E}i} \sim \mathrm{Exp}(\lambda)$ $(i \in \{1, \ldots, k\})$ uma variável com **distribuição Exponencial**. Então, $Y = Y_{\mathrm{E}1} + Y_{\mathrm{E}2} + \cdots + Y_{\mathrm{E}k}$ tem distribuição Gama.
  - A Gama tem suporte no conjunto dos **reais positivos**, assumindo formas assimétricas.
  - Ela tem ampla aplicações na área de **confiabilidade** e análise de **sobrevivência**.

$$f(y) = \frac{\lambda^r}{\Gamma(r)} \cdot y^{r-1} \cdot \exp\{-\lambda y\}$$

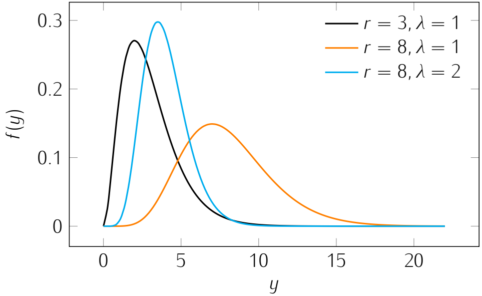

Figura 9. Distribuição Gama.

Distribuição Gama

- Variável aleatória Y tem distribuição Gama de parâmetros $r > 0$ e $\lambda > 0$ se sua **função densidade de probabilidade** é dada por

$$f(y) = \frac{\lambda^r}{\Gamma(r)} \cdot y^{(r-1)} \cdot \exp\{-\lambda y\}, \quad y > 0,$$

em que

$$\Gamma(r) = \int_0^\infty u^{r-1} \exp\{-u\}\, \mathrm{d}u \quad \text{com} \quad \Gamma(r) = (r-1)! \quad \text{quando } r \in \mathbb{N},$$

é a **função gama** (por isso distribuição Gama).

- Denotamos por $Y \sim \text{Gama}(r, \lambda)$.
- Parâmetros λ e r são chamados de **taxa** e **forma**, respectivamente.
- Y apresenta **média** e **variância** dadas por

$$\mu = \mathrm{E}(Y) = \frac{r}{\lambda} \quad \text{e} \quad \sigma^2 = \mathrm{V}(Y) = \frac{r}{\lambda^2}.$$

Gráfico de densidade da Gama

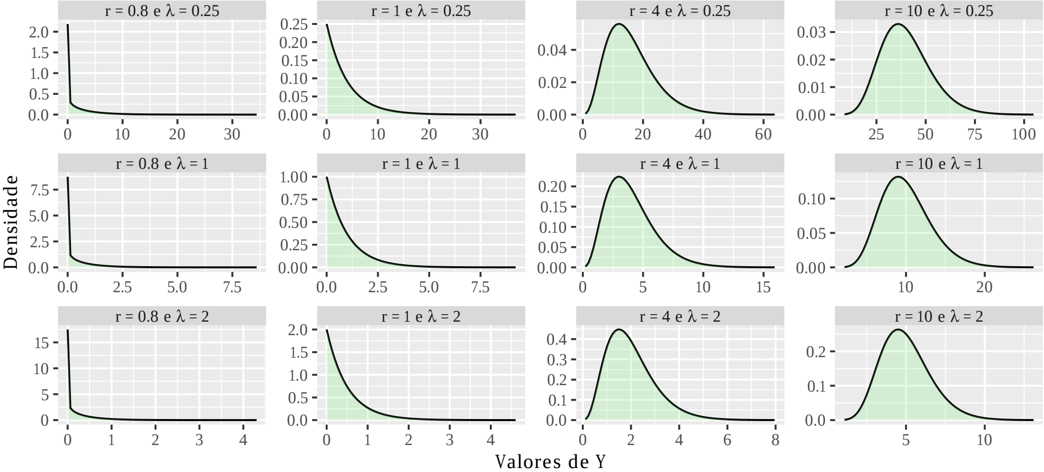

Figura 10. Gráficos de densidade para a Gama.

Gráfico de distribuição da Gama

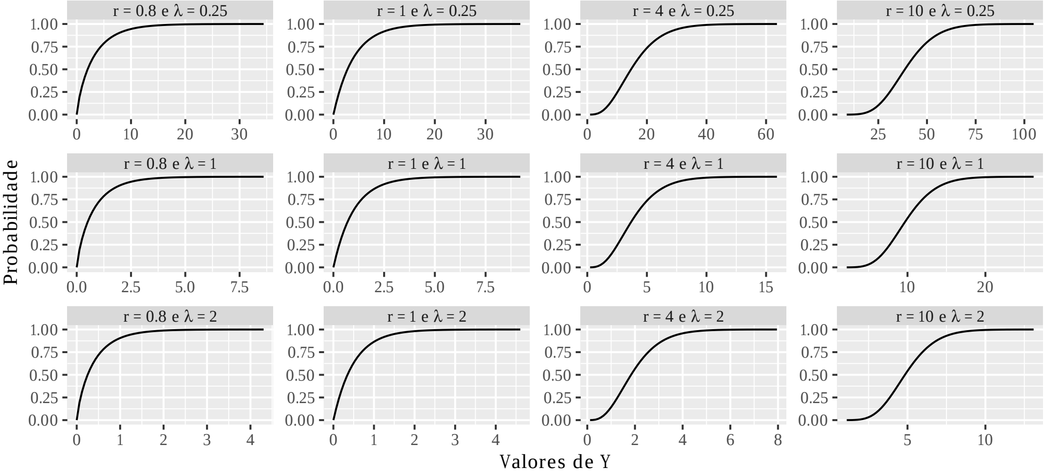

Figura 11. Gráficos de distribuição para a Gama.

Exemplos de v.a. com distribuição Gama

1. Volume total por parcela.
2. Volume individual de árvores.
3. Tempo de sobrevivência da árvore em uma floresta.
4. Tempo de sobrevivência de mudas em uma casa de vegetação.
5. Razão entre o diâmetro mensurado em diferentes alturas do fuste da árvore e o diâmetro à altura do peito.

Considerações finais

Considerações finais

- A **escolha** de uma distribuição de probabilidade adequada é fundamental para uma análise correta do problema.
- Há **diversas** distribuições na literatura.
- A utilização de *softwares* é **fundamental** para a aplicação dos modelos.

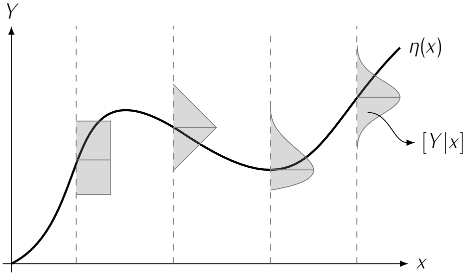

Figura 12. Modelo de probabilidade.