Distribuições discretas: Uniforme Discreta, Bernoulli, Binomial e Poisson

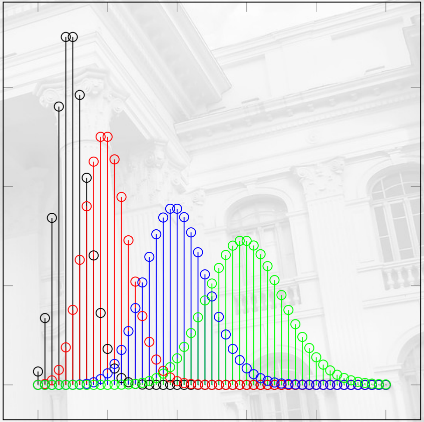
Prof. Walmes M. Zeviani

Departamento de Estatística
Universidade Federal do Paraná

Conteúdo

Neste vídeo

- Principais modelos probabilísticos discretos.
  - Uniforme Discreta.
  - Bernoulli.
  - Binomial.
  - Poisson.
- Fundamentação e propriedades.
- Exemplos de aplicação.



Distribuição Uniforme Discreta

Distribuição Uniforme Discreta

É a variável aleatória discreta mais simples, pois apresenta um número finito de valores possíveis com **igual probabilidade**.

Uma variável aleatória Y tem uma distribuição **Uniforme Discreta** se cada um dos m valores em seu suporte, isto é, $y_1, y_2, \ldots, y_m$, apresentar igual probabilidade, no caso

$$p(y_i) = \frac{1}{m}, \quad i = 1, \ldots, m.$$

Denotamos por $Y \sim \text{UD}(m)$.

Figura 1. A face resultante do lançamento de um dado tem distribuição UD com $m = 6$. Foto de fotografierende no Pexels.

Exemplo de v.a. com distribuição Uniforme Discreta

- O resultado de lançar um dado.
- Último digito da placa de um veículo.
- Um número de sorteio de Bingo.
- Um número de sorteio de Mega Sena.
- A posição da 1º lâmpada que queima em uma fita de luzes pisca-pisca.

Figura 2. Equipamento para jogo de bingo.

Exemplo

O último dígito de uma imagem de CAPTCHA de um site é igualmente provável de ser qualquer um entre 0 e 9. Sendo assim, se Y representa o último dígito, então terá distribuição Uniforme Discreta com $p(y) = 1/10$ para qualquer $y \in \{0, 1, \ldots, 9\}$.

Figura 3. Exemplo de CAPTCHA. Extraído de http://tecnodrom.com.br/146/captchas-numericos-e-o-google-street-view.

Média e variância

Suponha que Y tenha 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 $p = 1/m$.

Dessa forma, tem-se que

- A média é

$$\mu = \mathrm{E}(Y) = \sum_{y=a}^{b} y \cdot \left(\frac{1}{b-a+1}\right) = \frac{b+a}{2}$$

- A variância é

$$\sigma^2 = \mathrm{V}(Y) = \sum_{y=a}^{b} (y-\mu)^2 \cdot \left(\frac{1}{b-a+1}\right) = \frac{(b-a+1)^2 - 1}{12}.$$

Gráficos da distribuição

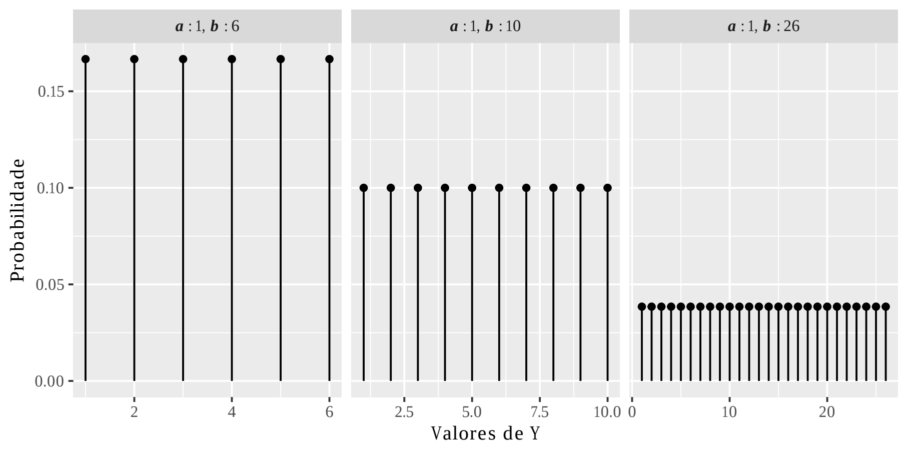


Figura 4. Gráficos para a distribuição Uniforme Discreta.

Distribuição de Bernoulli

Distribuição de Bernoulli

A 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**. Dessa forma, a função de probabilidade é

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

Denotamos por $Y \sim \text{Ber}(p)$.

Com isso, Y apresenta:

- $\mu = \text{E}(Y) = p$.
- $\sigma^2 = \text{V}(Y) = p \cdot (1-p)$.

Exemplos de v.a. com distribuição de Bernoulli

1. Se a face resultante do lançamento de uma moeda: se face cara $\rightarrow y = 1$.
2. Se um anúncio para um cliente é convertido em venda: se converter $\rightarrow y = 1$.
3. Se uma semente germina: se germinar $\rightarrow y = 1$.
4. Se o retorno mensal de um investimento é positivo: se lucro $\rightarrow y = 1$.
5. Se o jogador faz ponto no arremesso à cesta: se acerto $\rightarrow y = 1$.
6. Se um robô consegue resolver um CAPTCHA: se resolver $\rightarrow y = 1$.
7. Se o paciente é diagnosticado com Covid-19: se sim $\rightarrow y = 1$.
8. Se um réu é condenado após julgamento: se condenado $\rightarrow y = 1$.

*Nem sempre o que é convencionado como "sucesso" ($y = 1$) é algo positivo. É apenas uma convenção.

Gráfico da distribuição de Bernoulli

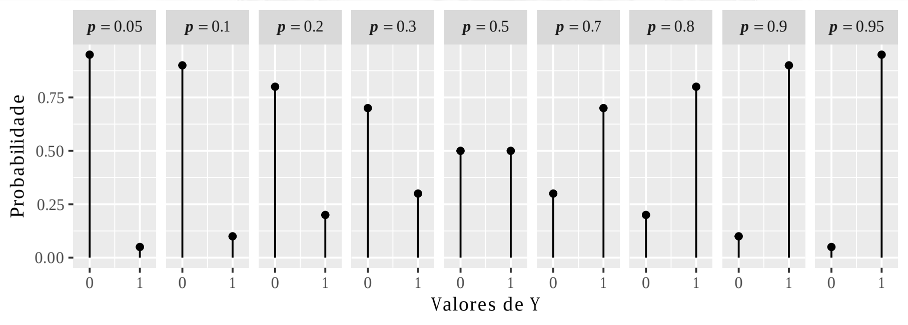


Figura 5. Gráficos para a distribuição de Bernoulli.

Importância da v.a. com distribuição de Bernoulli

A Bernoulli é a base para definição de outras variáveis aleatórias.

1. Número de sucessos em determinado número de tentativas ($n > 0$).
   Número de arremessos que acertou no total de 3 arremessos livres.
2. Número de tentativas até $r > 0$ sucessos ocorrerem.
   Número de arremessos até acertar 3 vezes.
3. Número de ocorrências de cada classe $y_1, y_2, \ldots, y_k$, $k > 1$.
   Número de arremessos errados, de 1, 2 e 3 pontos ($k = 4$).

Figura 6. Jogador de basquete marcando ponto. Foto de Wallace Chuck no Pexels.

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

Figura 7. Arremesso em uma partida de basebol. Foto de Pixabey no Pexels.

Distribuição Binomial

A 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

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

Denotamos por $Y \sim \text{Bin}(n, p)$.

Dessa forma, Y apresenta:

- $\mu = \text{E}(Y) = n \cdot p$.
- $\sigma^2 = \text{V}(Y) = n \cdot p \cdot (1-p)$.

Lembre-se:

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

Gráfico da distribuição Binomial

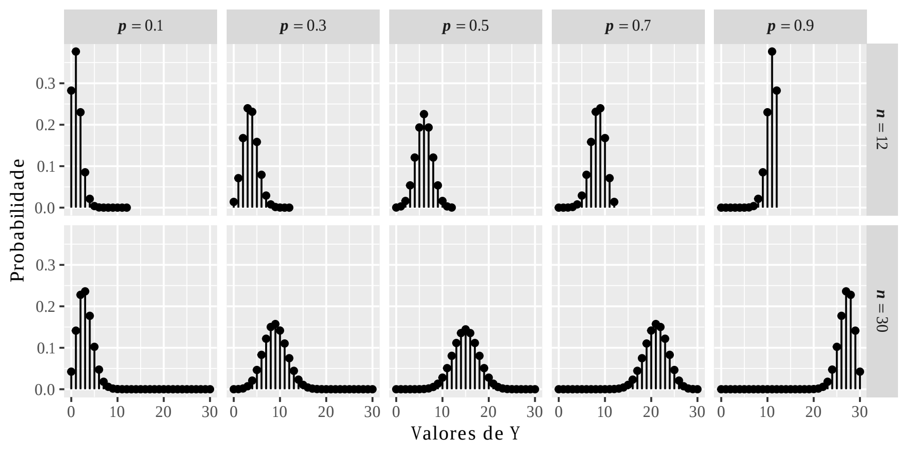


Figura 8. Gráficos para a distribuição Binomial.

Exemplos de v.a. com distribuição Binomial

1. Número de caras ao se lançar 10 moedas.
2. Número de ovos danificados em uma caixa com uma dúzia de ovos.
3. Número de peças defeituosas por lote de 100 peças.
4. Número de confeitos de cor azul em 25 confeitos de M&Ms.
5. Número de arremessos à cesta que acertou no total de 3 arremessos livres.
6. Número de clientes que participaram de uma promoção de 1000 clientes que receberam o anúncio.
7. Número de intenções de voto para o candidato A em 2000 entrevistados.
8. Número de processos judiciais concluídos em um mês do total de processos que foram submetidos.
9. Número de dias com chuva em um ano.

*As suposições precisam ser atendidas.

Exemplo: transporte de ovos de galinha

Uma empresa desenvolve embalagens para armazenamento e transporte de produtos alimentícios. Nos ensaios sob **condições extremas** de transporte e manuseio, a nova embalagem desenvolvida para ovos de galinha protegeu mais que as concorrentes apresentando apenas 10% ($p = 0.10$) de ovos danificados por caixa com uma dúzia ($n = 12$).

1. Qual a probabilidade de 1 caixa ter 2 ovos danificados nas condições acima?
2. Considere que em condições normais de transporte $p = 0.01$. Qual a probabilidade de, um carregamento de 100 caixas, ter 90 ou mais delas sem ovos danificados?

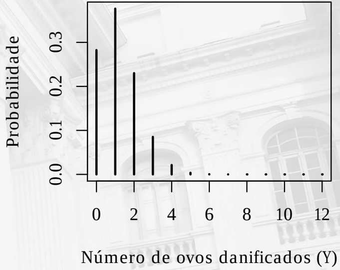


Figura 9. Número de ovos danificados por dúzia em condições extremas de transporte e manuseio.

Solução

1. Basta aplicar a fórmula

$$p(2) = \binom{12}{2} \cdot 0.1^2 \cdot (1 - 0.1)^{12-2} = 0.2301.$$

2. A probabilidade de uma cartela sem ovos danificados é

$$p(0) = \binom{12}{0} \cdot 0.01^0 \cdot (1 - 0.01)^{12-0} = 0.8864.$$

Assim, determina-se a probabilidade de 90 ou mais cartelas intactas por

$$\text{P}(Y \geq 90) = \textstyle\sum_{y=90}^{y=n=100} \binom{n}{y} \cdot p^y \cdot (1-p)^{n-y} = p(90) + p(91) + \cdots + p(100).$$

Aplicando-se a fórmula, tem-se

$$\text{P}(Y \geq 90) = \textstyle\sum_{y=90}^{y=n=100} \binom{100}{y} \cdot 0.8864^y \cdot (1 - 0.8864)^{100-y} = 0.4082.$$

Para pensar sobre as suposições

- Os ovos tem a mesma resistência a impactos?
- Todas as posições na caixa oferecem mesmo risco de quebrar?
- Eles quebram de forma independente uns dos outros ou existe dependência espacial?
- E se as suposições p constante e tentativas independentes não forem atendidas?

Figura 10. Caixa com uma dúzia de ovos.

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. Número de eventos em um domínio.
2. Taxa de ocorrência constante.
3. Independência entre domínios disjuntos.
4. Taxa proporcional ao tamanho do domínio.

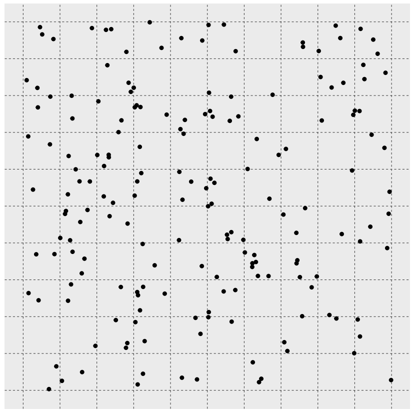


Figura 11. Ocorrência de eventos no espaço contínuo.

Características de uma v.a. com distribuição de Poisson

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

1. Consistir no **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 aconteça é a mesma para qualquer unidade de mesma dimensão $\rightarrow$ 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** à dimensão do subdomínio.

Exemplos de v.a. com distribuição de Poisson

1. Número de novos casos de Covid-19 por dia em Curitiba.
2. Número de interrupções de operação das máquinas por mês em uma fábrica.
3. Número de insetos por armadilha em uma lavoura.
4. Número de chutes a gol em uma partida de futebol.
5. Número de reclamações de clientes em um dia se serviço no SAC.
6. Número de links para outras páginas em uma página do wikipedia.
7. Número de referências bibliográficas em um artigo publicado.
8. Número de compartilhamentos de uma postagem de rede social em um dia.
9. Número de pessoas alcançadas com uma campanha de marketing digital nas primeiras 6 horas.

*As suposições precisam ser atendidas.

Distribuição de Poisson

A variável aleatória Y tem distribuição de Poisson com parâmetro (taxa) $\lambda > 0$ quando sua função de probabilidades é expressa por

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

Denotamos por $Y \sim \text{Pois}(\lambda)$.

Assim, Y tem média e variância dadas por

- $\mu = \text{E}(Y) = \lambda$ é a média de eventos por unidade.
- $\sigma^2 = \text{V}(Y) = \lambda$.

Ou seja, $\mu = \sigma^2$ e isso é **equidispersão**.

Gráfico da distribuição de Poisson

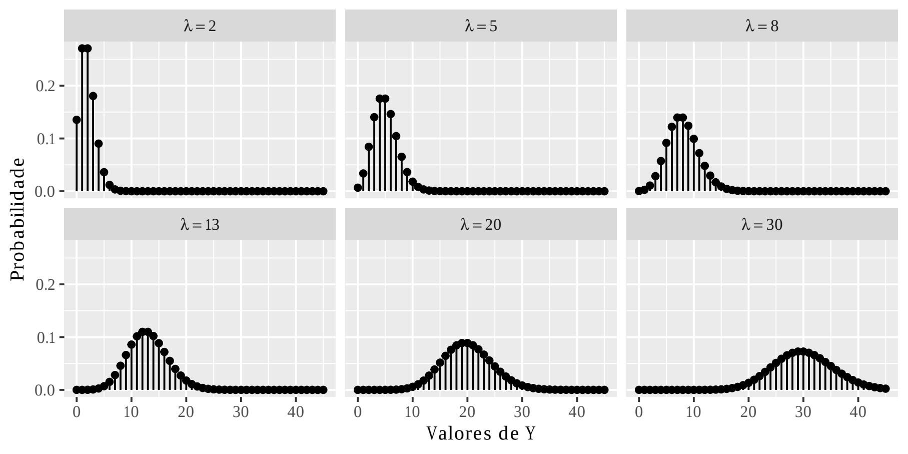


Figura 12. Gráficos para a distribuição de Poisson.

Exemplo: número de clientes de um restaurante

Como parte de uma iniciativa para melhorar o serviço prestado por um restaurante local, a gerente monitorou a chegada de clientes para o almoço em vários dias. A conclusão foi de que o número de clientes que chegam a cada 10 minutos segue uma distribuição de Poisson. As probabilidades, segundo a distribuição, são dadas na tabela.

1. Determine λ.
2. Calcule as probabilidades faltantes na tabela.
3. O dono do restaurante diz que o número médio de clientes por hora é maior que 60. Está correto?

Tabela 1. Distribuição de probabilidades para o número de clientes a cada 10 minutos.

y	$P(Y = y)$	y	$P(Y = y)$
0	0.0020	7	0.1421
1	0.0124	8	0.1104
2	—	9	0.0762
3	0.0800	10	—
4	0.1243	11	0.0268
5	0.1545	12	0.0139
6	—	13	0.0066

Solução

1. Como são dadas as probabilidades, resolva alguma delas em λ, por exemplo

$$P(Y = 0) = \frac{\exp\{-\lambda\} \cdot \lambda^0}{0!} \qquad \Rightarrow \lambda = -\ln(P(0))$$
$$\lambda = -\ln(0.002) = 6.2146.$$

2. Use a função de probabilidades

$$P(Y = 6) = \frac{\exp\{-6.2146\} \cdot 6.2146^6}{6!}$$
$$= 0.16$$

3. Se a cada 10 minutos a média é $\mu = \lambda = 6.2146$, então em 60 minutos será

$$\mu_{\text{hora}} = \frac{60}{10} \cdot \mu_{\text{10 minutos}} = 37.2876.$$

Distribuição de Poisson para unidades de dimensão variável

Como **propriedade** da distribuição de 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 seguinte forma

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

A média e variância são

- $\mu = \text{E}(Y) = \lambda t$ para unidades de tamanho t.
- $\sigma^2 = \text{V}(Y) = \lambda t$.

Para pensar nas suposições

- Considere o número de crimes por bairro em New York, NY.
- Como acomodar bairros de dimensão diferente:
  - Área territorial ou
  - Número de habitantes?
- A taxa de crime é constante ou depende, por exemplo, da concentração de riquezas?
- A ocorrência de um crime é independente das ocorrências em localidades vizinhas e datas vizinhas?

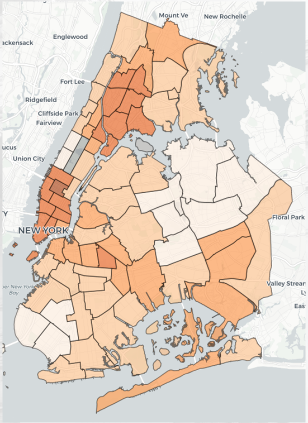


Figura 13. Crimes em New York para o período de 1 a 30/Jul/2020. Extraído de https://maps.nyc.gov/crime/.

Tipos dispersão e relação média variância

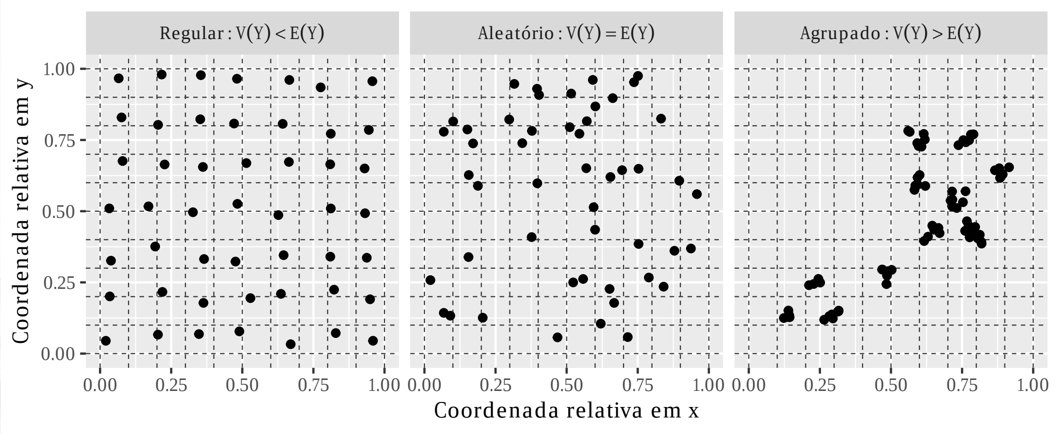


Figura 14. Padrões de dispersão espacial e a relação de média e variância da contagem.

Relação entre Binomial e Poisson

Considere que $Y_1 \sim \text{Bin}(n, p)$ e $Y_2 \sim \text{Pois}(\lambda)$ de tal forma que $\lambda = np$. Dessa forma, pode-se escrever

$$p(y) = \binom{n}{y} p^y (1 - p)^{n-y} = \binom{n}{y} \left(\frac{\lambda}{n}\right)^y \left(1 - \frac{\lambda}{n}\right)^{n-y}.$$

Agora considere que $n \to \infty$ ao mesmo tempo que $p \to 0$ mas de tal modo que λ permaneça constante. Com algum trabalho pode ser mostrado que

$$\lim_{n\to\infty} \binom{n}{y} \left(\frac{\lambda}{n}\right)^y \left(1 - \frac{\lambda}{n}\right)^{n-y} = \frac{\exp\{-\lambda\} \cdot \lambda^y}{y!}.$$

O que isso quer dizer na prática?

1. Que a distribuição de Poisson pode ser vista como um **caso limite** da distribuição Binomial.
2. Que podemos **aproximar** a distribuição Binomial pela distribuição de Poisson sempre que n for grande e p for pequeno.

Exemplo

- n: número total de taças em uma grande festa de formatura. n é muito grande.
- p: probabilidade de uma taça quebrar. p é muito pequeno.
- Y_1: número de taças quebradas do total de taças em circulação (n).
- Y_2: número de taças quebradas na festa de formatura.
- $Y_1 \sim \text{Bin}(n, p) \rightarrow Y_2 \sim \text{Pois}(\lambda = np)$.

Considerações finais

Considerações finais

Revisão

- Principais modelos probabilísticos discretos: Uniforme Discreta, Bernoulli, Binomial e Poisson.
- Fundamentação e propriedades $\rightarrow$ as suposições de cada fenômeno.
- São de extrema importância: Binomial e Poisson.
- Exemplos de aplicação.