ME414 - Estatística para Experimentalistas

Parte 11

Distribuição Binomial

Exemplo: Sherlock

Um inimigo de Sherlock propõe um jogo, que consiste no lançamento de uma moeda honesta várias vezes, em quatro versões:

1. Se a proporção de caras for maior ou igual do que 0.60, Sherlock vence.
2. Se a proporção de caras for maior ou igual do que 0.40, Sherlock vence.
3. Se a proporção de caras estiver entre 0.40 e 0.60 (inclusive), Sherlock vence.
4. Se a proporção de caras for menor ou igual do que 0.30, Sherlock vence.

O inimigo escolhe primeiro qual a versão do jogo e depois Sherlock terá que escolher se quer jogar com 10 ou 100 lançamentos da moeda.

Exemplo: Sherlock

Se o inimigo escolhe a versão 1, Sherlock deve escolher 10 ou 100 lançamentos?

Seja X_i a v.a. que indica o resultado do i-ésimo lançamento da moeda, ou seja,

$$X_i = \begin{cases} 1, & \text{se sair cara} \\ 0, & \text{se sair coroa} \end{cases} \quad \text{e} \quad P(X_i = 0) = P(X_i = 1) = 0.5$$

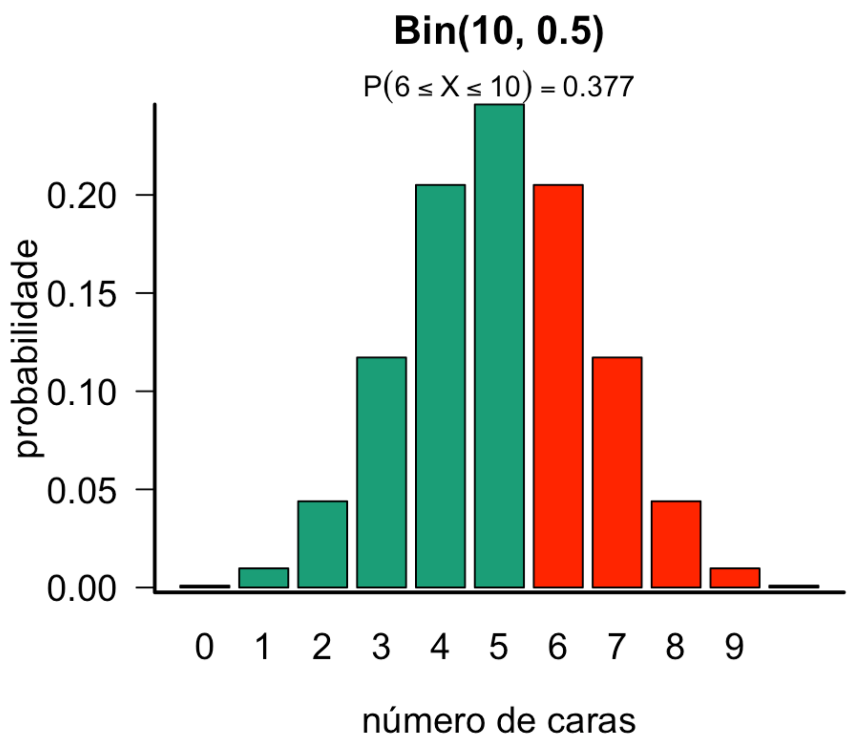
Seja $X = \sum_{i=1}^{n} X_i$ a v.a. que indica o número de caras em n lançamentos da moeda. Então, $X \sim \text{Bin}(n, 0.5)$.

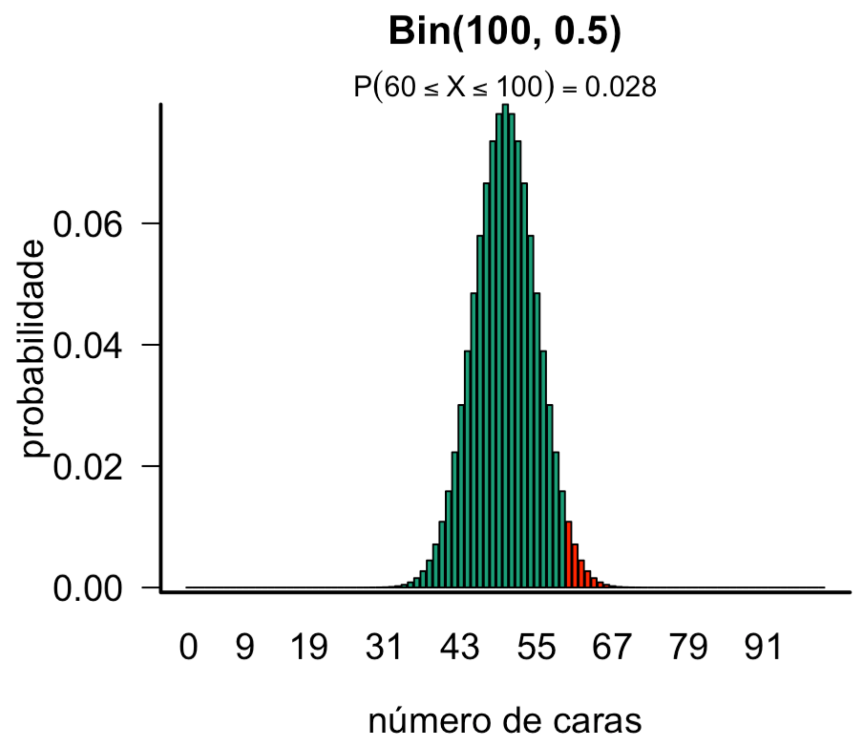
Na versão 1, Sherlock vence se a proporção de caras é maior do que 0.60, ou seja, se $X \geq n \times 0.6$.

Basta então Sherlock comparar $P(X \geq n \times 0.6)$ para $n = 10$ ou $n = 100$ e escolher o que resultar em maior probabilidade.

Exemplo: Sherlock - Versão 1

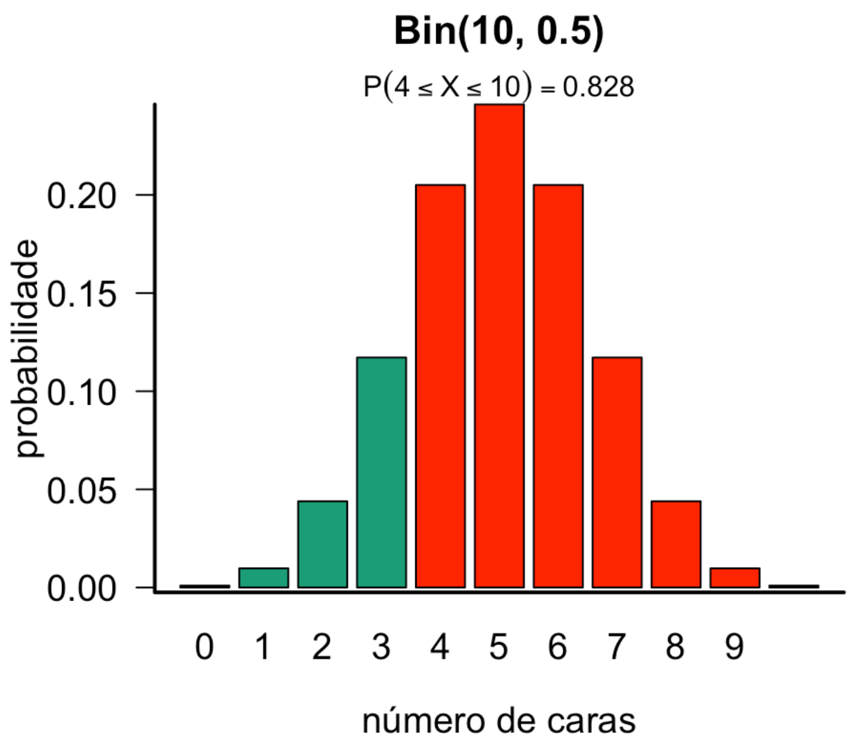
Se a proporção de caras for maior ou igual do que 0.60, Sherlock vence.

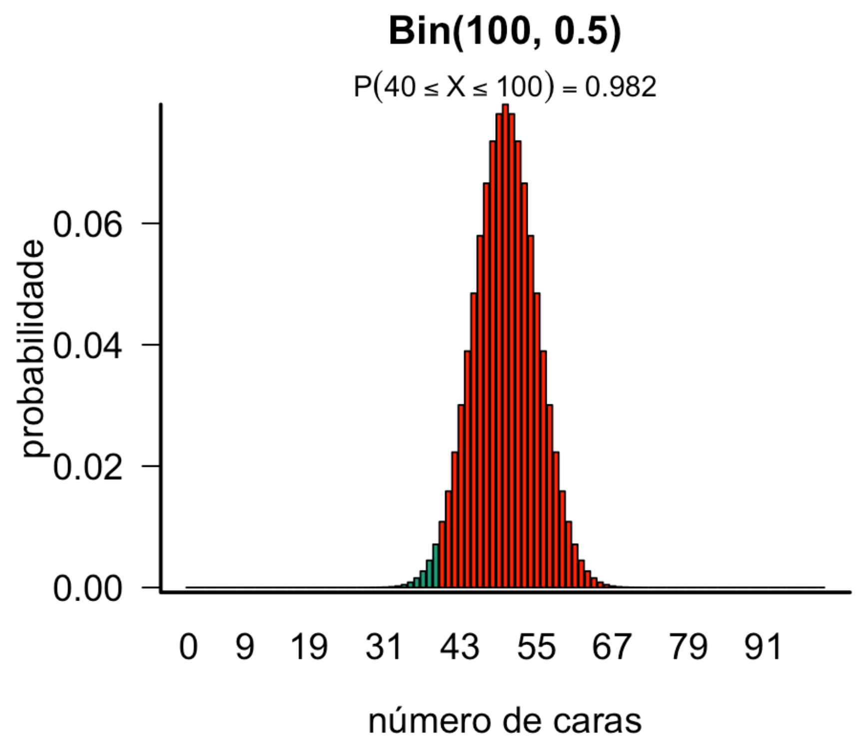




Exemplo: Sherlock - Versão 2

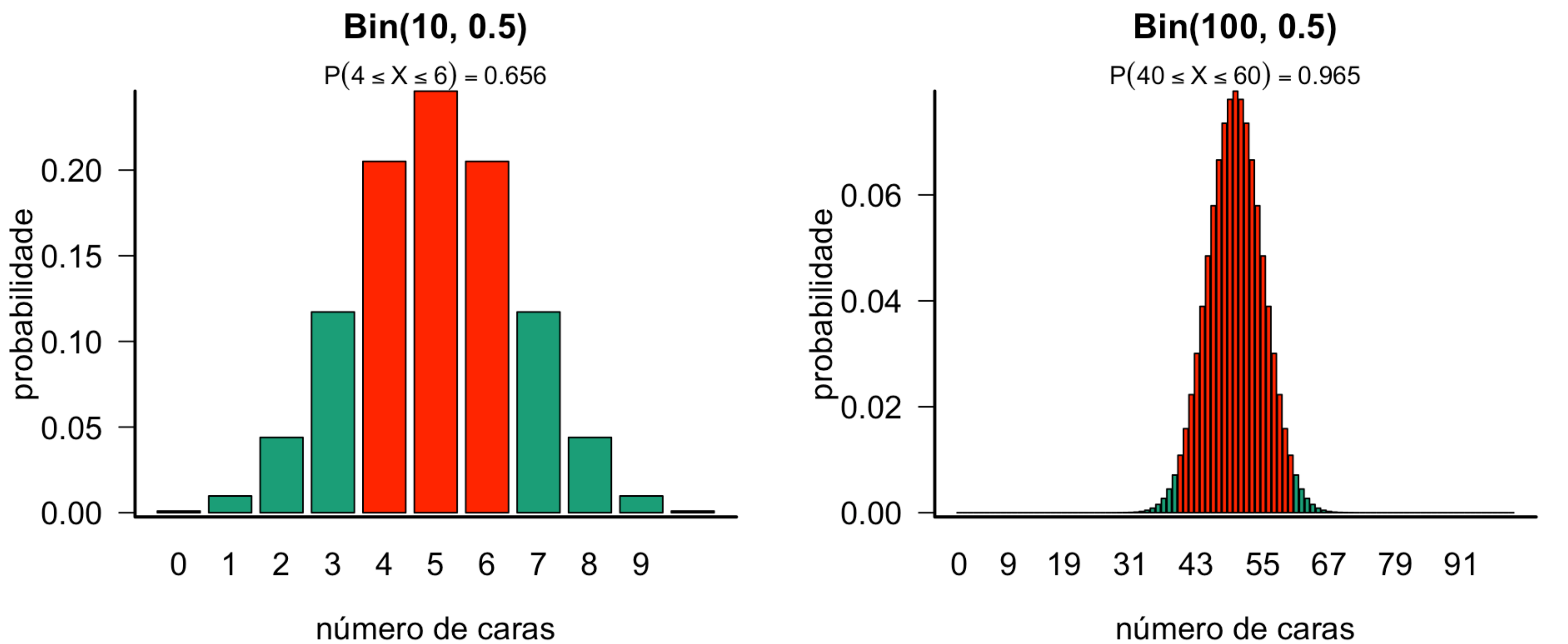
Se a proporção de caras for maior ou igual do que 0.40, Sherlock vence.





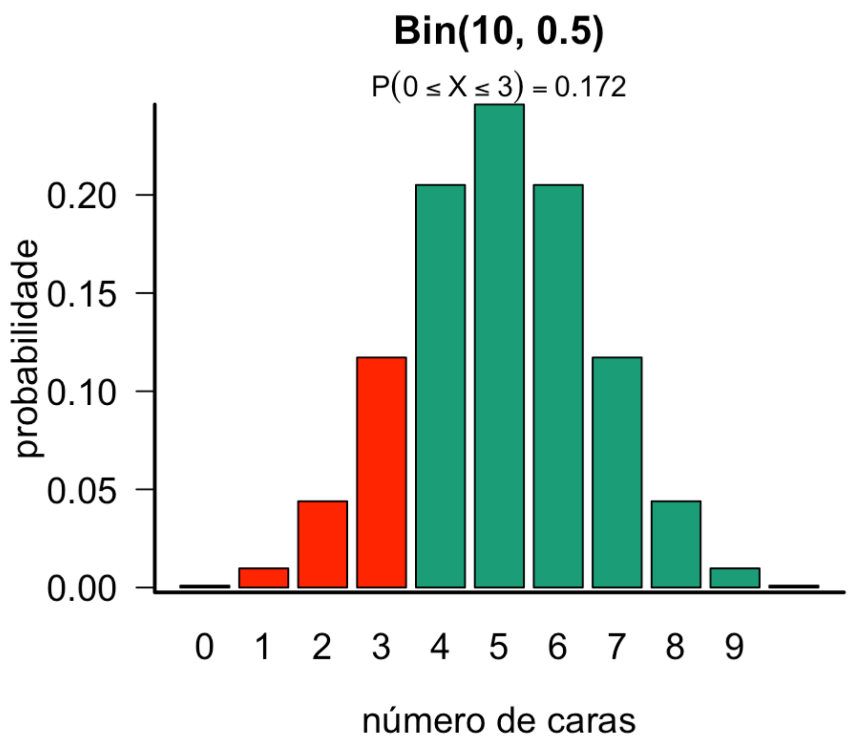
Exemplo: Sherlock - Versão 3

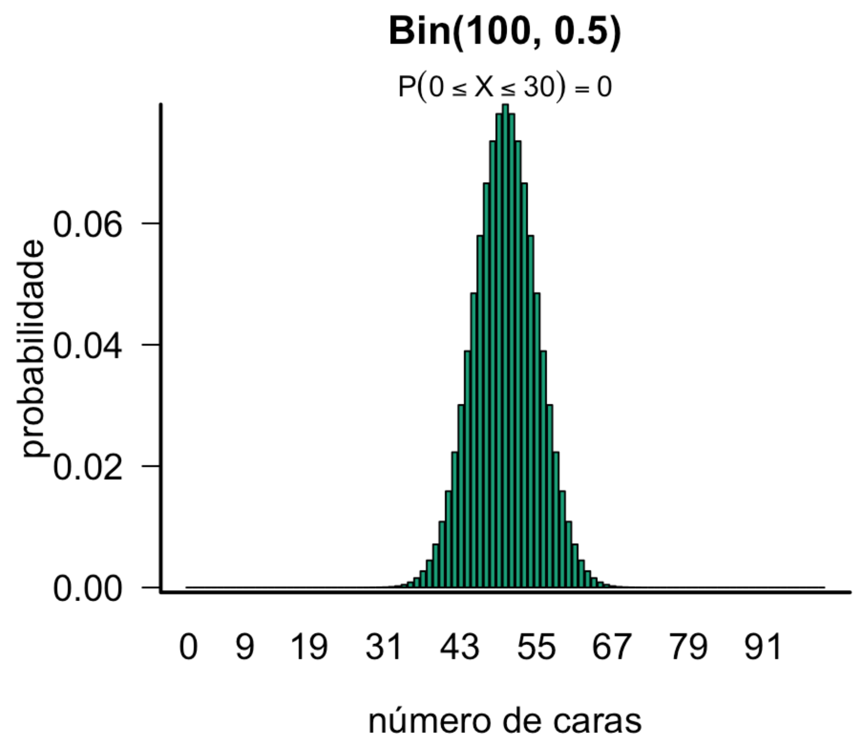
Se a proporção de caras estiver entre 0.40 e 0.60 (inclusive), Sherlock vence.



Exemplo: Sherlock - Versão 4

Se a proporção de caras for menor ou igual do que 0.30, Sherlock vence.





Exemplo: Sherlock

Na prática, para tomar uma decisão rápida, Sherlock deve considerar que a proporção esperada de caras é sempre 0.5, mas que quando n é menor, existe maior variabilidade em torno desse valor esperado.

- Se a proporção de caras for maior do que 0.60, Sherlock vence.

Aqui Sherlock deve escolher $n = 10$, pois a variância de X/n (proporção de caras) é maior com $n = 10$.

- Se a proporção de caras for maior do que 0.40, Sherlock vence.

Aqui Sherlock deve escolher $n = 100$, pois a variância de X/n é menor com $n = 100$, portanto chances maiores da proporção observada estar próxima da proporção esperada.

Exemplo: Sherlock

- Se a proporção de caras estiver entre 0.40 e 0.60, Sherlock vence.

Mesmo raciocínio do item anterior.

- Se a proporção de caras for menor do que 0.30, Sherlock vence.

Mesmo raciocínio do item 1.

Distribuição Geométrica

Distribuição Geométrica

Consideremos novamente um experimento aleatório com espaço de resultados Ω e o evento A.

Vamos dizer que ocorreu sucesso se o evento A aconteceu e $p = P(\text{sucesso})$.

Repetimos o experimento até o primeiro sucesso.

Seja X o número de repetições até o primeiro sucesso.

Exemplo: lançar uma moeda repetidas vezes até a primeira cara e $p = P(cara)$.

Os valores possíveis de X são $\{1, 2, 3, \dots\}$.

$$
\begin{aligned}
P(X = 1) &= p && \text{(sucesso logo na primeira tentativa)} \\
P(X = 2) &= (1 - p)p && \text{(1 fracasso seguido de 1 sucesso)} \\
P(X = k) &= (1 - p)^{k-1}p && (k - 1 \text{ fracassos sucessivos e 1 sucesso})
\end{aligned}
$$

Distribuição Geométrica

Modelo Geral: Suponha uma sequência de ensaios de Bernoulli independentes com probabilidade de sucesso p.

Seja X a v.a. que representa o número de ensaios de Bernoulli até a ocorrência do primeiro sucesso. Então dizemos que X segue uma distribuição **Geométrica** com parâmetro p, ou seja, $X \sim G(p)$.

A probabilidade de se observar x é dada por:

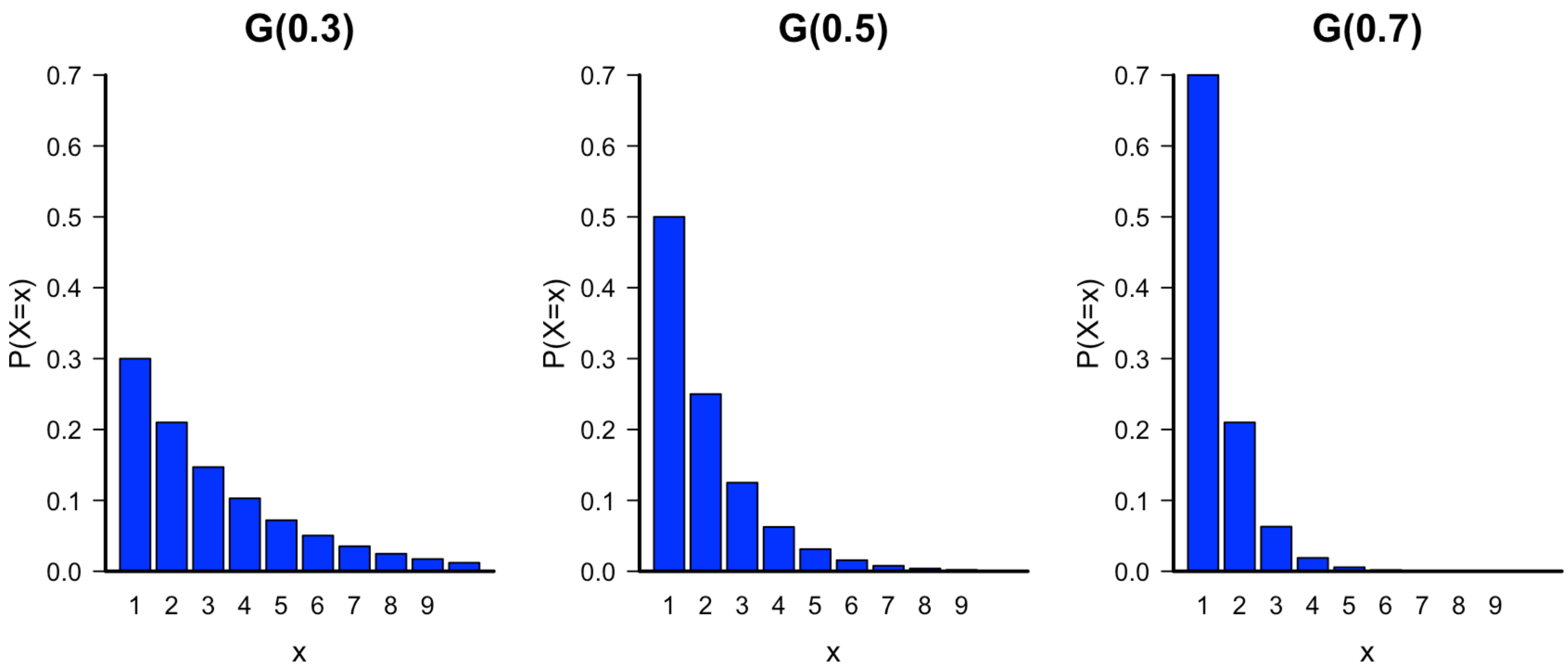
$$P(X = x) = (1 - p)^{x-1} p, \qquad x = 1, 2, \dots$$

A esperança e variância de uma v.a. Geométrica são dadas por:

$$\mathbb{E}(X) = \frac{1}{p} \qquad \text{e} \qquad Var(X) = \frac{1 - p}{p^2}$$

Distribuição Geométrica

Distribuição de probabilidade de uma $G(p)$, com $p = 0.3, 0.5$ e 0.7.



Distribuição Geométrica

A função de distribuição acumulada de uma v.a. $G(p)$ é dada por:

$$F(x) = P(X \leq x) = 1 - (1-p)^x$$

A distribuição geométrica tem uma propriedade que serve para caracterizá-la no conjunto das distribuições discretas: a propriedade de perda de memória!

Propriedade de Perda de Memória

$$P(X > x + m \mid X > m) = P(X > x)$$

Interpretação: O fato de já termos observado m fracassos sucessivos não muda a probabilidade do número de ensaios até o primeiro sucesso ocorrer.

Propriedade de perda de memória

$$P(X > x + m \mid X > m) = P(X > x)$$

Demonstração:

Lembre-se que:

$$F(x) = P(X \leq x) = 1 - (1-p)^x \quad \Longrightarrow \quad P(X > x) = (1-p)^x$$

Então,

$$\begin{aligned} P(X > x + m \mid X > m) &= \frac{P(X > x + m, X > m)}{P(X > m)} \\ &= \frac{P(X > x + m)}{P(X > m)} = \frac{(1-p)^{x+m}}{(1-p)^m} \\ &= (1-p)^x = P(X > x) \end{aligned}$$

Exemplo: Sinal de trânsito

A probabilidade de se encontrar aberto o sinal de trânsito numa esquina é 0.2.

Qual a probabilidade de que seja necessário passar pelo local 5 vezes para encontrar o sinal aberto pela primeira vez?

$X =$ número de vezes necessárias para encontrar o sinal aberto.

$p = P(\text{sinal aberto}) = 0.2$

$$\begin{aligned} P(X = 5) &= (1 - p)^4 p \\ &= 0.8^4 \times 0.2 \\ &= 0.0819 \end{aligned}$$

Exemplo: lançamento de um dado

Qual a probabilidade de que um dado deva ser lançado 15 vezes para que ocorra a face 6 pela primeira vez?

$X =$ número de vezes necessárias para ocorrer o resultado 6.

$p = P(\text{face } 6) = 1/6$

$$\begin{aligned} P(X = 15) &= (1 - p)^{15-1} p \\ &= \left(\frac{5}{6}\right)^{14} \frac{1}{6} \\ &= 0.01298 \end{aligned}$$

Exemplo: Banco de Sangue

Um banco de sangue necessita sangue do tipo O negativo. Suponha que a probabilidade de uma pessoa ter este tipo de sangue seja 0.10. Doadores permanentes chegam ao hemocentro para fazer sua doação rotineira. Calcule a probabilidade de que o primeiro doador com sangue do tipo O negativo seja:

- o primeiro a chegar;
- o segundo;
- o sétimo.
- Quantos doadores esperamos passar pelo hospital até encontrarmos um com sangue O negativo?

Fonte: Prof. Mario Gneri, Notas de Aula.

Exemplo: Banco de Sangue

Seja X o número de doadores que chegam no hemocentro até a chegada do primeiro doador com sangue O negativo.

Novamente temos um experimento com distribuição geométrica. Usando a fórmula para a função de probabilidade, send $X \sim G(0.1)$:

$$P(X = x) = 0.9^{x-1}0.1, \qquad x = 1, 2, \ldots$$

Temos que

- $P(X = 1) = 0.1$
- $P(X = 2) = 0.9 \times 0.1 = 0.09$
- $P(X = 7) = 0.9^6 \times 0.1 = 0.053$
- $\mathbb{E}(X) = 1/0.1 = 10$. Neste caso, esperamos que dez doadores passem pelo hospital, em média, para encontrarmos o primeiro com sangue O negativo.

Hipergeométrica

Distribuição Hipergeométrica

- População dividida em duas características
- Extrações casuais sem reposição

Detalhes:

- N objetos
- r têm a característica A
- $N - r$ têm a característica B
- um grupo de n elementos é escolhido ao acaso, dentre os N possíveis, sem reposição.

Objetivo: calcular a probabilidade de que este grupo de n elementos contenha x elementos com a característica A.

Distribuição Hipergeométrica

Seja X a v.a. que representa o número de elementos com a característica A dentre os n selecionados.

Então dizemos que X segue uma distribuição **Hipergeométrica** com parâmetros N, n, r, ou seja, $X \sim Hip(N, n, r)$.

A probabilidade de se observar x é dada por:

$$P(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}, \qquad 0 \leq x \leq min\{r, n\}$$

A esperança e variância são, respectivamente:

$$\mathbb{E}(X) = \frac{nr}{N} \qquad \text{e} \qquad Var(X) = \frac{nr}{N}\left(1 - \frac{r}{N}\right)\frac{(N-n)}{(N-1)}$$

Exemplo: Urna

Uma urna contém 10 bolas: 6 brancas e 4 pretas.

Qual a probabilidade de obter 3 bolas brancas dentre 4 bolas retiradas?

Seja X o número de bolas brancas dentre as 4 bolas retiradas

Então, $X \sim Hip(N = 10, n = 4, r = 6)$ e

$$P(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}, \qquad 0 \leq x \leq min\{r, n\}$$

Portanto,

$$P(X = 3) = \frac{\binom{6}{3}\binom{4}{1}}{\binom{10}{4}} = \frac{8}{21}$$

Exemplo: Comissão

Voltando ao exemplo: O Departamento de Estatística é formado por 25 professores, sendo 17 homens e 8 mulheres. Uma comissão será formada por 3 professores. Queremos saber qual é a probabilidade da comissão ser formada por pelo menos duas mulheres?

Seja X o número de mulheres na comissão, então
$X \sim Hip(N = 25, n = 3, r = 8)$

$$P(X = 2) = \frac{\binom{8}{2}\binom{17}{1}}{\binom{25}{3}} = 0.21$$

$$P(X = 3) = \frac{\binom{8}{3}\binom{17}{0}}{\binom{25}{3}} = 0.02$$

$$P(X \leq 2) = P(X = 2) + P(X = 3) = 0.21 + 0.02 = 0.23$$

Exemplo: Loteria

- Um jogo de loteria consiste em selecionar 6 dezenas de 00 a 99, com uma bola para cada dezena e sem reposição
- Numa aposta o jogador pode escolher de 6 a 10 dezenas
- Qual a probabilidade de acertar a quina (5 dezenas) marcando-se 10 dezenas na aposta?

- N = 100 (total de dezenas)
- n = 6 (dezenas sorteadas)
- r = 10 (dezenas escolhidas pelo jogador)
- x = 5 (número de sucessos, queremos 5)

$$P(X = 5) = \frac{\binom{10}{5}\binom{100-10}{6-5}}{\binom{100}{6}} = 0.000019$$

Exemplo: Mega-Sena

Qual a probabilidade de um jogador ganhar na Mega-Sena jogando 6 dezenas?

- $N = 60$ (dezenas de 01 a 60)
- $n = 6$ (dezenas sorteadas)
- $r = 6$ (dezenas escolhidas pelo jogador)
- $x = 6$ (número de sucessos, queremos 6)

Então, a probabilidade de ganhar na Mega-Sena é:

$$P(X = 6) = \frac{\binom{6}{6}\binom{54}{0}}{\binom{60}{6}} = \frac{1}{50063860}$$

Aplicação: Controle de Qualidade

Suponha um lote com $N = 100$ elementos a ser analisado.

São escolhidas $n = 5$ peças sem reposição.

Sabendo que neste lote de 100 elementos, $r = 10$ são defeituosos.

Se X é o número de peças defeituosas em 5 escolhidas, então

$$X \sim Hip(N = 100, n = 5, r = 10)$$

A probabilidade de nenhuma peça defeituosa na amostra retirada é:

$$P(X = 0) = \frac{\binom{10}{0}\binom{100-10}{5-0}}{\binom{100}{5}} = \frac{\binom{90}{5}}{\binom{100}{5}} \approx 0.584$$

Aplicação: Controle de Qualidade

A probabilidade de pelo menos uma peça defeituosa é:

$$P(X \geq 1) = \sum_{i=1}^{5} P(X = i) = 1 - P(X = 0) \approx 0.416$$

A média e a variância são:

$$\begin{aligned}
\mathbb{E}(X) &= \frac{nr}{N} = \frac{5 \times 10}{100} = 0.5 \\
Var(X) &= \frac{nr}{N}\left(1 - \frac{r}{N}\right)\frac{(N - n)}{(N - 1)} \\
&= \frac{5 \times 10}{100}\left(1 - \frac{10}{100}\right)\frac{(100 - 5)}{(100 - 1)} \approx 0.409
\end{aligned}$$

Exemplo

Pequenos motores são guardados em caixas de 50 unidades. Um inspetor de qualidade examina cada caixa, antes da remessa, testando 5 motores. Se nenhum motor for defeituoso, a caixa é aceita. Se pelo menos um for defeituoso, todos os 50 motores são testados. Há 6 motores defeituosos numa caixa. Qual a probabilidade de que seja necessário examinar todos os motores?

$X =$ número de motores defeituosos da amostra.

$N = 50, n = 5$ e $r = 6$. Então $X \sim Hip(N = 50, n = 5, r = 6)$

Se pelo menos 1 é defeituoso, inspeciona todos os 50.

$$\begin{aligned} P(X \geq 1) &= 1 - P(X < 1) = 1 - P(X = 0) \\ &= 1 - \frac{\binom{6}{0}\binom{44}{5}}{\binom{50}{5}} = 1 - 0.5126 = 0.4874 \end{aligned}$$

Exemplo

Uma firma compra lâmpadas por centenas. Examina sempre uma amostra de 15 lâmpadas para verificar se estão boas. Se uma centena inclui 12 lâmpadas queimadas, qual a probabilidade de se escolher uma amostra com pelo menos uma lâmpada queimada?

$X =$ número de lâmpadas queimadas da amostra.

$N = 100, n = 15$ e $r = 12$. Então $X \sim Hip(N = 100, n = 15, r = 12)$

Pelo menos uma queimada:

$$\begin{aligned} P(X \geq 1) &= 1 - P(X < 1) = 1 - P(X = 0) \\ &= 1 - \frac{\binom{12}{0}\binom{88}{15}}{\binom{100}{15}} = 0.8747 \end{aligned}$$

Aproximação da Binomial pela Poisson

Distribuição de Poisson

- Muitas vezes, em problemas em que seria natural usar a distribuição binomial, temos n muito grande ($n \to \infty$) e p muito pequeno ($p \to 0$).

- Nesses casos, o cálculo fica difícil com calculadoras comuns.

- Considerando uma v.a. $X \sim \text{Bin}(n, p)$, quando temos grandes valores para n e p pequeno (mantendo-se o produto $np = \lambda$ constante), podemos usar a seguinte aproximação para a probabilidade:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \approx \frac{e^{-np}(np)^x}{x!}, \quad x = 0, 1, 2, \ldots n$$

- Geralmente considera-se o critério $np \leq 7$ para usar essa aproximação.

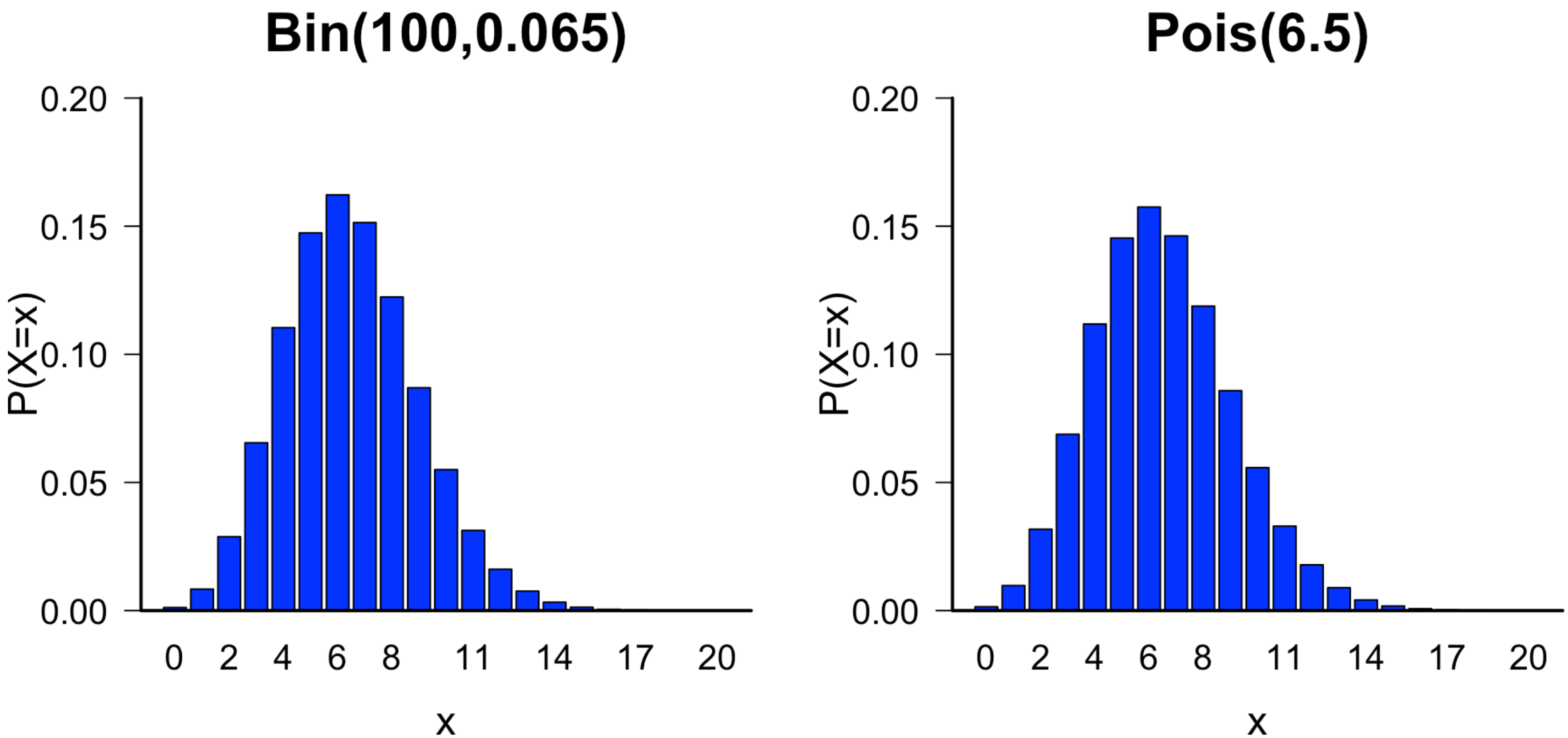
- Demonstração

Exemplo

$X \sim Bin(100, 0.065)$, deseja-se obter $P(X = 10)$

- No modelo Binomial: $P(X = 10) = \binom{100}{10}(0.065)^{10}(0.935)^{100-10} = 0.055$
- $\lambda = np = 100 \times 0.065 = 6.5 \leq 7$
- No modelo Poisson: $P(X = 10) = \frac{e^{-6.5}(6.5)^{10}}{10!} \approx 0.056$

Poisson para Aproximar uma Binomial



Exemplo

A probabilidade de uma lâmpada se queimar ao ser ligada é $1/100$. Numa instalação com 100 lâmpadas, qual a probabilidade de 2 lâmpadas se queimarem ao serem ligadas?

- No modelo Binomial: $X \sim \text{Bin}(100, 0.01)$

$P(X = 2) = \binom{100}{2}(0.01)^2(0.99)^{100-2}$=0.1849

- $\lambda = np = 100 \times 0.01 = 1 \leq 7$

- No modelo Poisson: $P(X = 2) = \frac{e^{-1}(1)^2}{2!} \approx 0.1839$

Distribuição de Poisson

Outro caso em que a distribuição de Poisson é utilizada:

- Considere a probabilidade de ocorrência de sucessos em um determinado intervalo.
- A probabilidade de ocorrência de um sucesso no intervalo é proporcional ao intervalo.
- A probabilidade de mais de um sucesso nesse intervalo é bastante pequena com relação à probabilidade de apenas um sucesso.

Distribuição de Poisson

- Seja X o número de sucessos no intervalo, então:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

  onde λ é a esperança.

- Notação: $X \sim \text{Poisson}(\lambda)$.

Distribuição de Poisson

A distribuição de Poisson é muito usada na distribuição do número de:

- carros que passam por um cruzamento por minuto, durante uma certa hora do dia;
- erros tipográficos por página, em um livro;
- defeitos por unidade (m^3, m^2, m, etc…) por peça fabricada;
- colônias de bactérias numa dada cultura por $0.01mm^2$, numa plaqueta de microscópio;
- mortes por ataque de coração por ano, em um certo bairro
- etc…

Distribuição de Poisson

Para uma v.a. quantificando eventos raros, sob algumas suposições, podemos usar a distribuição de Poisson.

- Uma variável aleatória X tem distribuição de Poisson com parâmetro $\lambda > 0$, se sua função de probabilidade é dada por:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

- λ é chamado de taxa de ocorrência
- $\mathbb{E}(X) = Var(X) = \lambda$
- Notação: $X \sim P(\lambda)$

Exemplo: Erros em um livro

Num livro de 800 páginas, há 800 erros de impressão. Qual a probabilidade de que uma página contenha pelo menos 3 erros?

$X =$ número de erros por página

Taxa de ocorrência: $\lambda = 1$

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) \\ &= 1 - [P(X = 0) + P(X = 1) + P(X = 2)] \\ &= 1 - \left\{ \frac{e^{-1} 1^0}{0!} + \frac{e^{-1} 1^1}{1!} + \frac{e^{-1} 1^2}{2!} \right\} \\ &= 0.08 \end{aligned}$$

Exemplo: Mensagens no Facebook

Uma firma recebe 720 mensagens em sua página do Facebook durante as 8 horas de horário comercial. Qual a probabilidade de que em 6 minutos no horário comercial a firma receba pelo menos 4 mensagens no Facebook?

$$
\begin{aligned}
720 \text{ mensagens} &\rightarrow 480 \text{ min} \\
\lambda &\rightarrow 6 \text{ min}
\end{aligned}
$$

Então, usando “regra de três”, $\lambda = 9$ e

$$
\begin{aligned}
P(X \geq 4) = 1 - P(X < 4) &= 1 - P(X \leq 3) \\
&= 1 - \left[\frac{e^{-9} 9^0}{0!} + \frac{e^{-9} 9^1}{1!} + \frac{e^{-9} 9^2}{2!} + \frac{e^{-9} 9^3}{3!} \right] \\
&= 0.979
\end{aligned}
$$

Exemplo: SAC

Numa central de SAC (serviço de atendimento ao consumidor) chegam 300 telefonemas por hora. Qual a probabilidade de que:

- num minuto não haja nenhuma chamada?
- em 2 minutos haja 2 chamadas?
- em t minutos não haja chamadas?

Exemplo: SAC

X = número de chamadas por minuto.

- Taxa de ocorrência por minuto: $\lambda = 300/60 = 5$

$$P(X = 0) = \frac{e^{-5} 5^0}{0!} = 0.0067$$

X = número de chamadas a cada 2 minutos.

- Taxa de ocorrência em 2 minutos: $\lambda = 10$

$$P(X = 2) = \frac{e^{-10} 10^2}{2!} = 0.00227$$

X = número de chamadas a cada t minutos.

- Taxa de ocorrência em t minutos: $\lambda = 5t$

$$P(X = 0) = \frac{e^{-5t} (5t)^0}{0!} = e^{-5t}$$

Exemplo: Lâmpadas

A experiência mostra que de cada 400 lâmpadas, 2 se queimam ao serem ligadas. Qual a probabilidade de que numa instalação de:

- 600 lâmpadas, no mínimo 3 se queimem?
- 900 lâmpadas, exatamente 8 se queimem?

Exemplo: Lâmpadas

$X =$ número de lâmpadas que se queimam numa instalação de 600 lâmpadas.

De 400, 2 se queimam, ou seja, de 200, 1 se queima.

Taxa de ocorrência para 600 lâmpadas: $\lambda = 600/200 = 3$

600 lâmpadas, no mínimo 3 se queimem:

$$
\begin{aligned}
P(X \geq 3) = \sum_{x=3}^{\infty} \frac{e^{-3} 3^x}{x!} = 1 - P(X < 3) = \\
&= 1 - [P(X = 0) + P(X = 1) + P(X = 2)] \\
&= 1 - \left[\frac{e^{-3} 3^0}{0!} + \frac{e^{-3} 3^1}{1!} + \frac{e^{-3} 3^2}{2!} \right] = 0.5768
\end{aligned}
$$

Exemplo: Lâmpadas

$X =$ número de lâmpadas que se queimam numa instalação de 900 lâmpadas.

Taxa de ocorrência para 900 lâmpadas: $\lambda = 900/200 = 4.5$

900 lâmpadas, 8 se queimem:

$$P(X = 8) = \frac{e^{-4.5}(4.5)^8}{8!} = 0.0463$$

Exemplo: Twitter

O número citações de uma certa conta do Twitter ocorre segundo uma distribuição de Poisson, com a média de oito citações por minuto.

Determinar qual a probabilidade de que num minuto se tenha:

1. dez ou mais citações;
2. menos que nove citações;
3. entre sete (inclusive) e nove (exclusive) citações.

Exemplo: Twitter

Sabemos que se $X \sim \text{Poisson}(\lambda)$, então sua função de probabilidade é

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, \ldots$$

Além disso, $\mathbb{E}(X) = \lambda$.

O enunciado diz *média de oito citações por minuto*, então a variável aleatória X = *número de citações por minuto* tem distribuição $\text{Poisson}(8)$.

- A probabilidade de dez ou mais chamadas é dada por:

$$\begin{aligned} P(X \geq 10) &= 1 - P(X < 10) = 1 - P(X \leq 9) \\ &= 1 - \sum_{k=0}^{9} \frac{e^{-8}8^k}{k!} = 1 - e^{-8} - \ldots - \frac{e^{-8}8^9}{9!} = 0.2833 \end{aligned}$$

Exemplo: Twitter

- A probabilidade de termos menos que nove citações em um minuto é dada por:

$$P(X < 9) = P(X \leq 8) = e^{-8} + \ldots + \frac{e^{-8}8^8}{8!} = 0.5926$$

- A probabilidade de termos entre sete (inclusive) e nove (exclusive) citações em um minuto é dada por:

$$\begin{aligned} P(7 \leq X < 9) &= P(7 \leq X \leq 8) = P(X = 7) + P(X = 8) \\ &= \frac{e^{-8}8^7}{7!} + \frac{e^{-8}8^8}{8!} = 0.2792 \end{aligned}$$

Leituras

- Ross: capítulo 5
- Magalhães: capítulo 3
- OpenIntro: seções 3.3, 3.4, 3.5.2

Slides produzidos pelos professores:

- Samara Kiihl
- Tatiana Benaglia
- Larissa Matos
- Benilton Carvalho