

# ACH2053 INTRODUÇÃO À ESTATÍSTICA

1º Sem/2018

## Estimação

Prof. Marcelo S. Lauretto  
marcelolauretto@usp.br  
www.each.usp.br/lauretto

Referência:

W.O.Bussab, P.A.Morettin. Estatística Básica, 6ª Edição.  
São Paulo: Saraiva, 2010 – Capítulo 11

# 1. Estimação: Introdução

- Vimos anteriormente que a Inferência Estatística tem por objetivo fazer generalizações sobre uma população, com base nos dados de uma amostra
- Principais problemas da inferência estatística:
  - Estimação de parâmetros
  - Testes de hipóteses sobre parâmetros
  - Previsão
- Relembrando:
  - *Parâmetros*: funções de valores populacionais
  - *Estatísticas*: funções de valores amostrais
- Neste capítulo: ideias básicas sobre **estimação**

# 1. Estimação: Introdução

- Exemplo 11.1:
  - Uma amostra de  $n=500$  pessoas de uma cidade é escolhida
  - Pergunta-se a pergunta entrevistado se ele é favorável ou não a um certo projeto proposto pela prefeitura
    - Respostas possíveis: SIM / NÃO
  - Deseja-se estimar a proporção,  $p$ , de pessoas na cidade favoráveis à proposta (ou seja, que responderiam SIM)
  - Se 300 pessoas responderam SIM, uma estimativa natural para  $p$  seria  $300/500$ , ou 60%
    - Essa estimativa é baseada na suposição de que a amostra é representativa da população
  - Vimos no capítulo anterior algumas propriedades desse estimador

# 1. Estimaco: Introduco

- Exemplo 11.1 (cont):

- Definamos as v.a.  $X_1, \dots, X_n$  tais que:

$$X_i = \begin{cases} 1, & \text{se a } i\text{sima pessoa na amostra responder SIM} \\ 0, & \text{se a } i\text{sima pessoa na amostra responder NO,} \end{cases}$$

- e seja  $p = \Pr(\text{SIM})$

- Se  $Y_n = \sum_{i=1}^n X_i$ , sabemos que  $Y_n$  tem distribuico binomial com parmetros  $n$  e  $p$ , e o problema consiste em estimar  $p$

- Como  $Y_n$  representa o nmero de pessoas na amostra que responderam SIM, um possvel **estimador** de  $p$  

$$\hat{p} = \frac{Y_n}{n} = \frac{\sum_{i=1}^n X_i}{n} = \frac{\text{nmero de SIM}}{\text{nmero de indivduos}}. \quad (11.1)$$

- Ou seja, se  $Y_n = k$ , ou seja, observarmos o valor  $k$  da varivel  $Y_n$ , obteremos  $\hat{p} = k/n$  como uma **estimativa** de  $p$

# 1. Estimação: Introdução

- Como visto na Seção 10.9, pelo TLC temos que, para tamanhos grandes de  $n$ ,  $\hat{p}$  tem distribuição aproximadamente normal com parâmetros:

$$E(\hat{p}) = p, \quad (11.2)$$

$$\text{Var}(\hat{p}) = p(1 - p)/n. \quad (11.3)$$

- Esses resultados nos ajudam a avaliar as qualidades do estimador  $\hat{p}$ 
  - Eq. (11.2) indica que  $\hat{p}$ , em média, “acerta”  $p$ 
    - Estimador **não-viesado** (ou **não-viciado**) de  $p$
  - Eq. (11.3) indica que, para amostras grandes, a diferença entre  $\hat{p}$  e  $p$  tende a ser pequena (para  $n \rightarrow \infty$ ,  $\text{Var}(\hat{p}) \rightarrow 0$ )
    - Estimador **consistente** de  $p$
- Nota: essas duas propriedades indicam uma *tendência* do estimador sobre o conjunto de todas as amostras, não uma *garantia* para qualquer amostra

## 2. Propriedades de estimadores

- Consideremos uma amostra  $(X_1, X_2, \dots, X_n)$  de uma v.a. que descreve uma característica de interesse de uma população
- Seja  $\theta$  um parâmetro que desejamos estimar, como por exemplo a média  $\mu = E(X)$  ou a variância  $\sigma^2 = Var(X)$
- **Definição:** Um **estimador**  $T$  do parâmetro  $\theta$  é qualquer função das observações da amostra, ou seja,  $T = g(X_1, X_2, \dots, X_n)$
- Segundo a definição acima, um estimador é o que havíamos chamado anteriormente de *estatística*; porém, aqui associamos essa estatística a um parâmetro populacional
- Uma forma usual de denotar um estimador para  $\theta$  é  $\hat{\theta}$

## 2. Propriedades de estimadores

- O problema da estimação é, então, determinar uma função  $T = g(X_1, X_2, \dots, X_n)$  que seja “próxima” de  $\theta$ , segundo algum critério
- Um dos critérios mais utilizados é dado a seguir:

- **Definição:** O estimador  $T$  é **não-viesado** para  $\theta$  se

$$E(T) = \theta, \quad (11.4)$$

para todo  $\theta$ .

- Se (11.4) não valer,  $T$  diz-se **viesado** e a diferença  $V(T) = E(T) - \theta$  é denominada o **viés** de  $T$ .
- Note que a esperança de  $T$  em (11.4) é calculada sobre a distribuição amostral de  $T$ , conforme vimos no capítulo anterior

## 2. Propriedades de estimadores

- **Definição: Estimativa** é o valor assumido pelo estimador em uma particular amostra.
- No exemplo 11.1,  $\hat{p}$  é um *estimador* de  $p$ , enquanto 60% é uma *estimativa* de  $p$
- Exemplo 11.3:
  - Pelo Teorema 10.1, sabemos que a média amostral  $\bar{X}$  é um estimador não-viesado de  $\mu = E(X)$ , pois  $E(\bar{X}) = \mu$   
Da mesma forma, a proporção amostral  $\hat{p}$  é um estimador não viesado da proporção  $p$  de indivíduos de uma população com certa característica em comum (ver Seção 10.9)

## 2. Propriedades de estimadores

- Exemplo 11.4:
  - Considere uma população com  $N$  elementos e a variância populacional

$$\sigma^2 = E[(X - \mu)^2] = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2, \quad (11.5)$$

onde  $\mu = \frac{1}{N} \sum_{i=1}^N X_i$  é a média populacional

- Um possível estimador para  $\sigma^2$ , baseado em uma AAS de tamanho  $n$  extraída dessa população, é

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (11.6)$$

- Mostraremos que o estimador  $\hat{\sigma}^2$  é viesado.

## 2. Propriedades de estimadores

- Estimador  $\hat{\sigma}^2$  é viesado (demonstração apenas de caráter informativo):

– Note que

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{1}{n} \sum_{i=1}^n 2X_i\bar{X} + \frac{1}{n} \sum_{i=1}^n \bar{X}^2 \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{2}{n} \bar{X} \sum_{i=1}^n X_i + \frac{n}{n} \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - 2\bar{X}\bar{X} + \bar{X}^2 \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2,\end{aligned}$$

logo,

$$E[\hat{\sigma}^2] = E\left[\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2\right] = \frac{1}{n} \sum_{i=1}^n E[X_i^2] - E[\bar{X}^2].$$

## 2. Propriedades de estimadores

- Estimador  $\hat{\sigma}^2$  é viesado (cont):
  - Mas pela definição de AAS e definição de variância de uma v.a.,

$$\begin{aligned}\text{Var}(X_i) &= E(X_i^2) - [E(X_i)]^2 \\ \Rightarrow E(X_i^2) &= \text{Var}(X_i) + [E(X_i)]^2 = \sigma^2 + \mu^2.\end{aligned}$$

Também, usando o Teorema 10.1, temos que

$$E(\bar{X}^2) = \text{Var}(\bar{X}) + [E(\bar{X})]^2 = \frac{\sigma^2}{n} + \mu^2.$$

Segue-se que

$$E[\hat{\sigma}^2] = \frac{1}{n} \sum_{i=1}^n (\sigma^2 + \mu^2) - \left( \frac{\sigma^2}{n} + \mu^2 \right),$$

ou seja,

$$\begin{aligned}E[\hat{\sigma}^2] &= \frac{1}{n} n(\sigma^2 + \mu^2) - \left( \frac{\sigma^2}{n} + \mu^2 \right) = \sigma^2 - \frac{\sigma^2}{n} = \sigma^2 \left( 1 - \frac{1}{n} \right) \\ &= \frac{n-1}{n} \sigma^2.\end{aligned}\tag{11.7}$$

## 2. Propriedades de estimadores

- Estimador  $\hat{\sigma}^2$  é viesado (cont):
  - De (11.7) vemos que  $\hat{\sigma}^2$  é viesado para  $\sigma^2$  e o viés é dado por
$$V = V(\hat{\sigma}^2) = E(\hat{\sigma}^2) - \sigma^2 = \left(\frac{n-1}{n}\right)\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n}. \quad (11.8)$$
  - Como esse viés é negativo, o estimador  $\hat{\sigma}^2$  em geral subestima o verdadeiro parâmetro  $\sigma^2$ .
  - Por outro lado, por (11.8) o viés diminui com  $n$ , ou seja, formalmente, para  $n \rightarrow \infty$ , o viés de  $\hat{\sigma}^2$  tende a zero.
  - É fácil ver que para obter um estimador não-viesado de  $\sigma^2$ , basta considerar  $\frac{n}{n-1}\hat{\sigma}^2$ , pois de (11.7) segue que

$$E\left[\frac{n}{n-1}\hat{\sigma}^2\right] = \frac{n}{n-1}E[\hat{\sigma}^2] = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2.$$

## 2. Propriedades de estimadores

- Do resultado anterior segue que, se definirmos

$$S^2 = \frac{n}{n-1} \hat{\sigma}^2 = \frac{n}{n-1} \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad (1.19)$$

então  $E(S^2) = \sigma^2 \Rightarrow S^2$  é um estimador não viesado para  $\sigma^2$

- Essa é a razão para se usar  $n - 1$  no denominador ao invés de  $n$

- No Excel:

- VAR.P: variância sobre uma população

$$\sigma^2 = E[(X - \mu)^2] = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2$$

- VAR.A: estimador não viesado da variância sobre uma amostra

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

# 3. Estimadores de máxima verossimilhança

- Significado geral de verossimilhança
  - *Verossímil*: aquilo que é semelhante à verdade, provável
  - *Verossimilhança*: qualidade ou caráter de verossímil
- Amostra verossímil:
  - Aquela que fornece a melhor informação possível sobre um parâmetro  $\theta$  de interesse da população, desconhecido, que queremos estimar
- Princípio da verossimilhança:
  - Afirma que devemos escolher o valor de  $\theta$  que maximiza a probabilidade de obter a amostra particular observada, ou seja, o valor que torna aquela amostra a “mais provável”

# 3. Estimadores de máxima verossimilhança

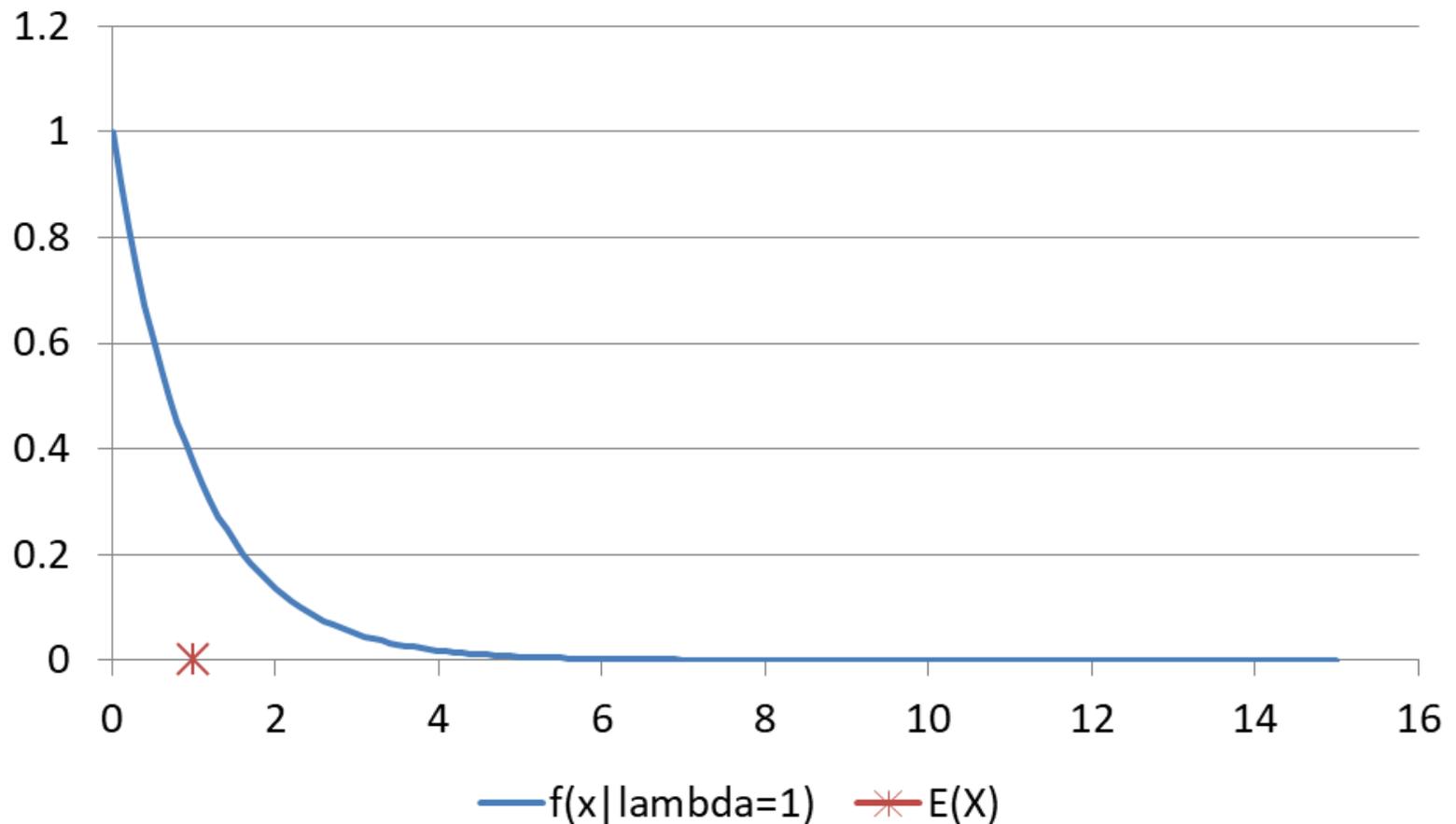
- Exemplo 1 (durabilidade de componentes eletrônicos):
  - Uma companhia produz componentes eletrônicos e está interessada em conhecer quanto cada componente deve durar, em média.
  - Os engenheiros pretendem modelar os componentes considerando a mesma taxa de falha desconhecida  $\lambda$  (constante no tempo)
  - Parte-se do princípio de que, se  $\lambda$  fosse conhecida, então a duração individual de cada componente ( $X$ ) teria distribuição exponencial com parâmetro  $\lambda$  (taxa)

# 3. Estimadores de máxima verossimilhança

- Exemplo 1 (cont):
  - Distribuição exponencial:
    - função de densidade de probabilidade (f.d.p):  $f(x|\lambda) = \lambda \exp(-\lambda x), x \geq 0$
    - função de distribuição acumulada (f.d.a):  $F(x|\lambda) = 1 - \exp(-\lambda x), x \geq 0$
    - função quantil:  $F^{-1}(p) = -\ln(1 - p)/\lambda$
    - $E(X) = 1/\lambda$  (escala,  $\beta$ )

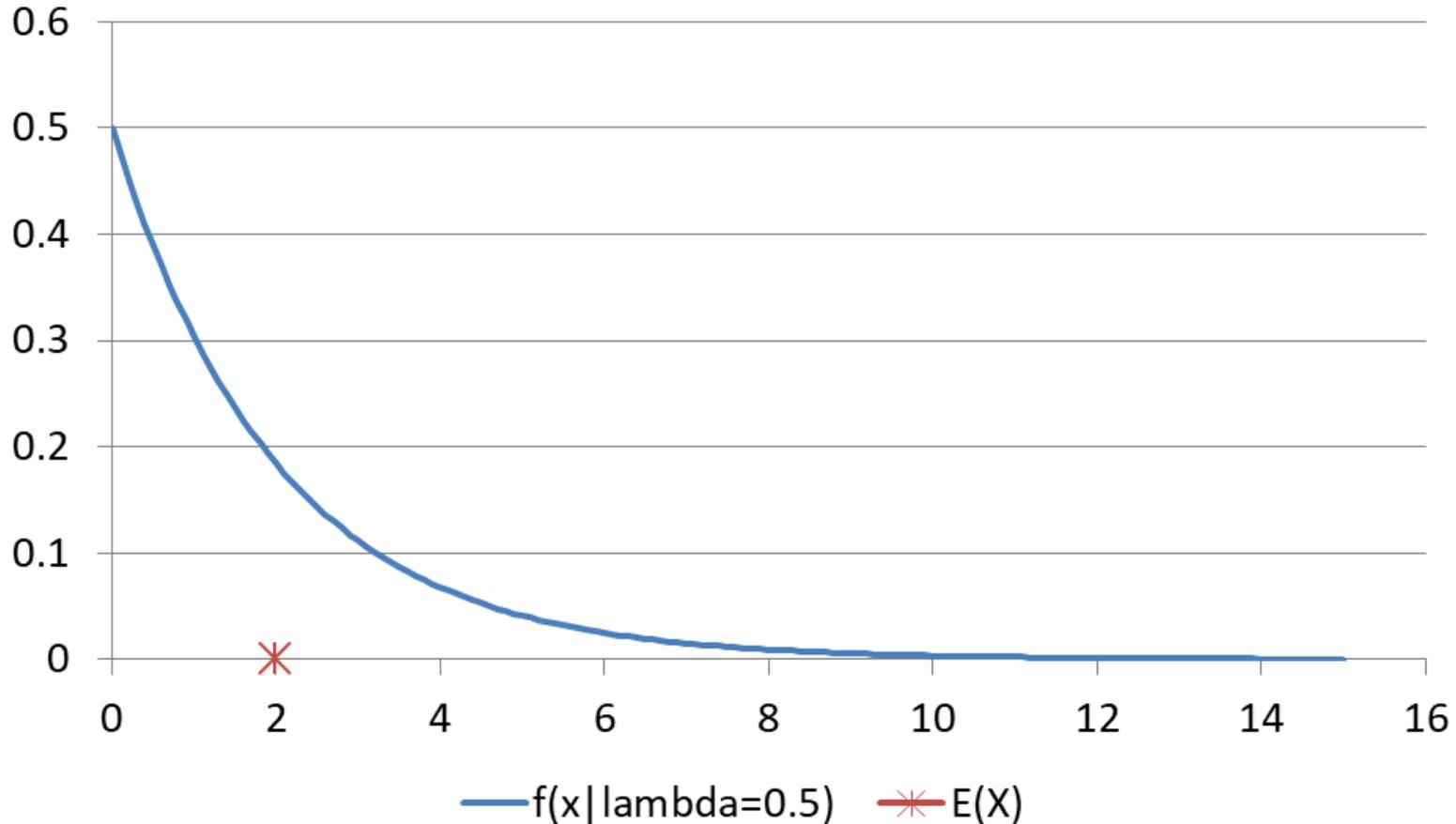
# 3. Estimadores de máxima verossimilhança

- Exemplo 1 (cont):
  - Distribuição exponencial (cont):
    - $\lambda = 1$



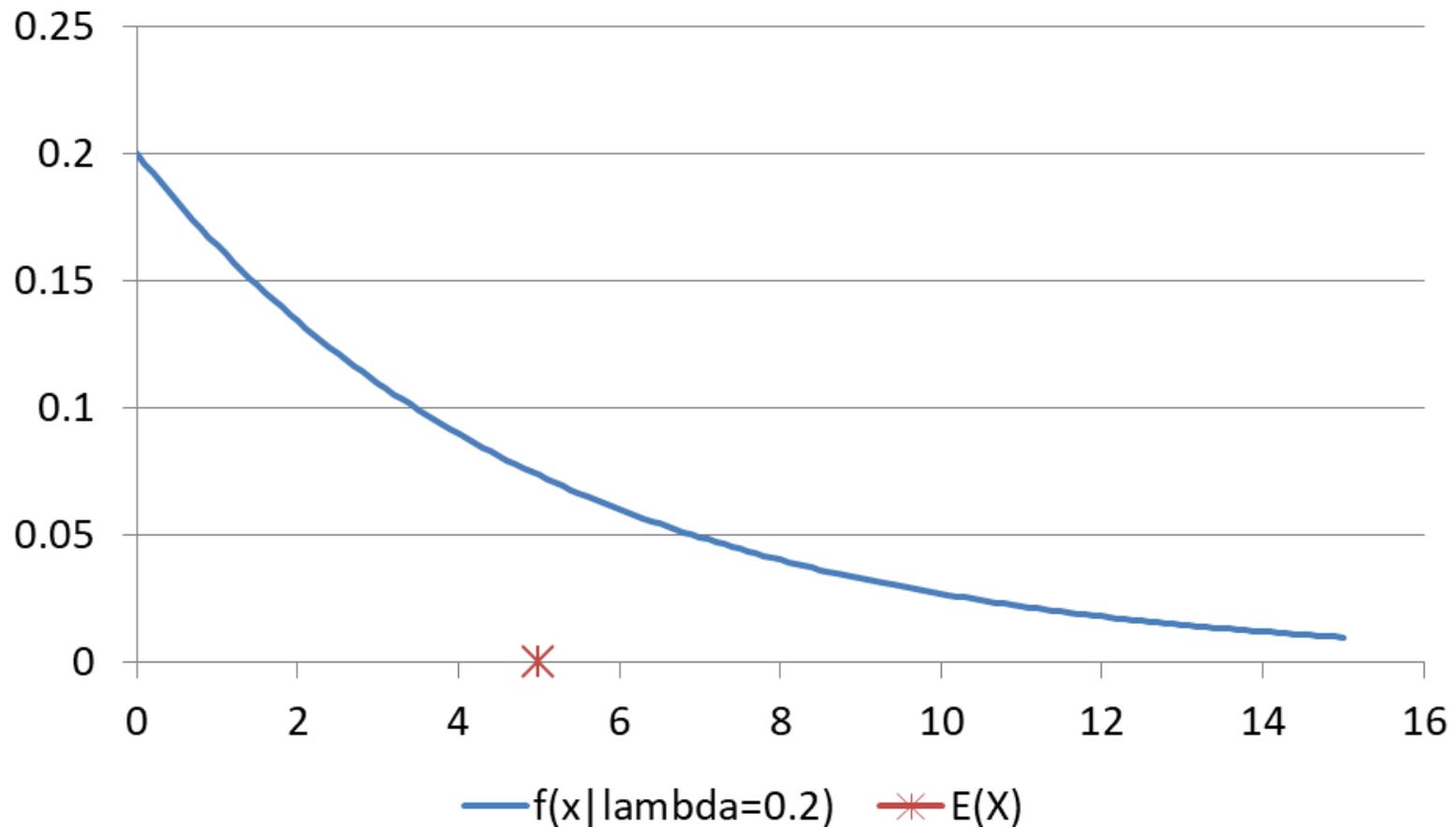
# 3. Estimadores de máxima verossimilhança

- Exemplo 1 (cont):
  - Distribuição exponencial (cont):
    - $\lambda = 1/2$



# 3. Estimadores de máxima verossimilhança

- Exemplo 1 (cont):
  - Distribuição exponencial (cont):
    - $\lambda = 1/5$



# 3. Estimadores de máxima verossimilhança

- Exemplo 1 (cont):
  - Suponha agora que 3 componentes foram usados sob condições típicas, com durações (em anos)  
 $x_1 = 3, x_2 = 1.5, x_3 = 2.1$ .
  - Denotaremos a amostra  $(x_1, x_2, x_3)$  por  $\mathbf{x}$
  - Assumindo que as durações dos 3 componentes possuíam a mesma distribuição exponencial( $\lambda$ ), então a f.d.p. conjunta  $f(\mathbf{x}|\lambda) = f(x_1, x_2, x_3|\lambda)$  é
$$\begin{aligned} f(\mathbf{x}|\lambda) &= f(x_1|\lambda) f(x_2|\lambda) f(x_3|\lambda) \\ &= \lambda \exp(-\lambda x_1) \lambda \exp(-\lambda x_2) \lambda \exp(-\lambda x_3) \\ &= \lambda^3 \exp(-\lambda[x_1 + x_2 + x_3]) = \lambda^3 \exp(-6.6\lambda) \end{aligned}$$

### 3. Estimadores de máxima verossimilhança

- Note que, no exemplo 1 acima,  $f(\mathbf{x}|\lambda)$  é de fato uma função de densidade de probabilidade sobre todas as amostras possíveis  $\mathbf{x} = (x_1, x_2, x_3)$ , i.e.

$$\int_{\mathbf{x}} f(\mathbf{x}|\lambda) = 1.$$

- Por outro lado, nosso interesse é usar o princípio da verossimilhança para encontrar o valor de  $\lambda$  (denotado por  $\hat{\lambda}$ ) que maximiza  $f(\mathbf{x}|\lambda)$ .
  - Ou seja, nosso interesse é considerar  $f(\mathbf{x}|\lambda)$  não como uma função de  $\mathbf{x}$  (amostra que já conhecemos), mas como uma função de  $\lambda$  (desconhecida)

# 3. Estimadores de máxima verossimilhança

- **Definição:**

- Quando a f.d.p. conjunta  $f(\mathbf{X}|\theta)$  das observações em uma amostra aleatória  $\mathbf{X}$  é tratada como uma função de  $\theta$ , então  $L(\theta|\mathbf{X}) \equiv f(\mathbf{X}|\theta)$  é denominada **função de verossimilhança**

- **Definição:**

- O estimador  $\hat{\theta} = \arg \max_{\lambda} L(\theta)$  é denominado **estimador de máxima verossimilhança (EMV)**

- Para maximizar  $L(\theta)$ , usualmente (ou quase sempre) é mais conveniente maximizar o logaritmo de  $L(\theta|\mathbf{X})$  (denominado usualmente *log-verossimilhança*), ou seja,

$$\hat{\theta} = \arg \max_{\lambda} \ell(\theta|\mathbf{X}),$$

onde  $\ell(\theta) = \log_e L(\theta)$

# 3. Estimadores de máxima verossimilhança

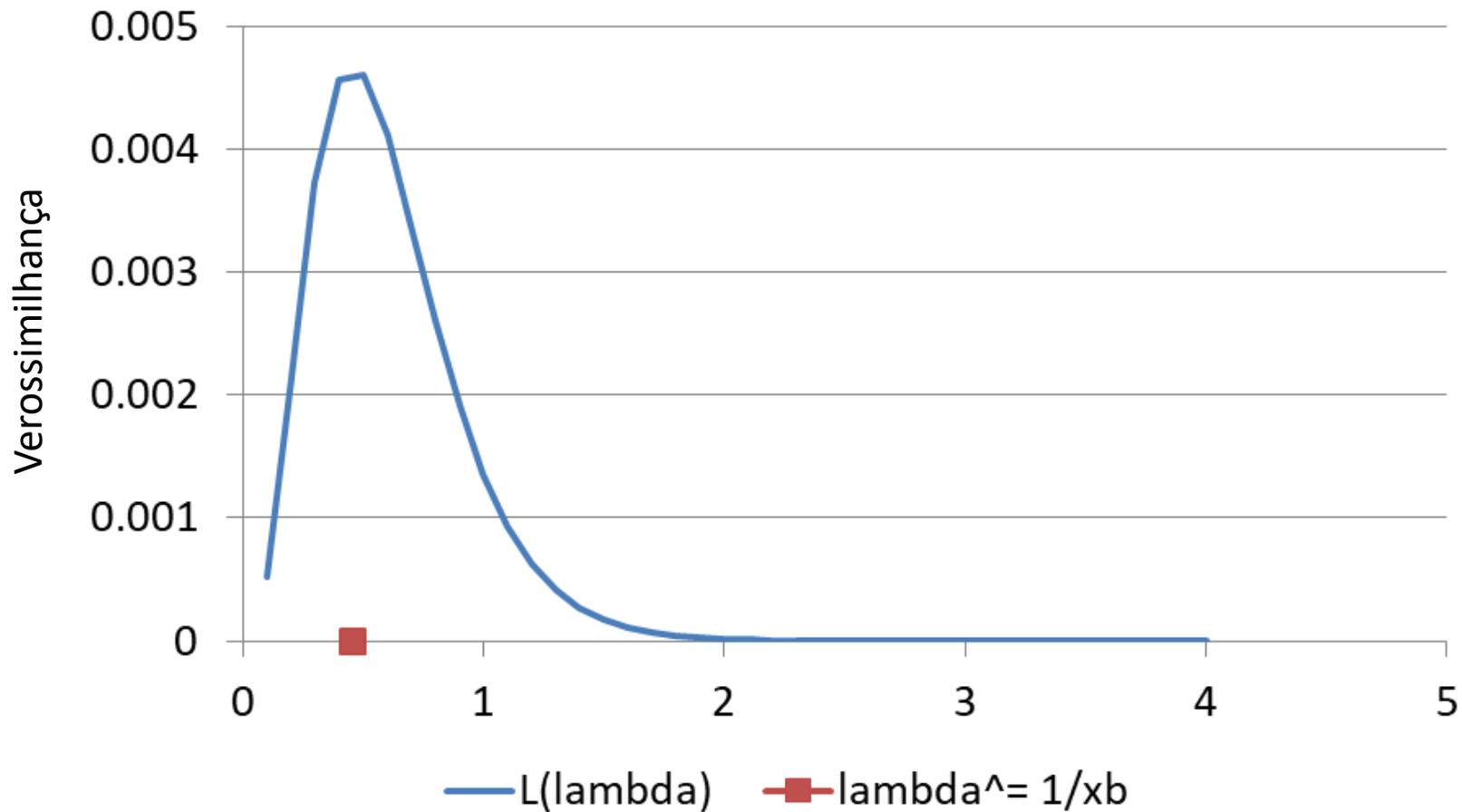
- Tipicamente,  $\hat{\theta}$  é obtido através de métodos numéricos de otimização não linear
  - Em R:
    - Função `fitdistr`: para algumas distribuições pré-determinadas
    - Funções `optim`, `constrOptim`: para otimização de funções de verossimilhança implementadas pelo próprio analista (ou para aquelas não implementadas na função `fitdistr`)
- Otimização é usualmente facilitada quando a derivada de  $\ell(\theta|\mathbf{X})$ , denotada  $\ell'(\theta|\mathbf{X})$ , pode ser obtida analiticamente.
  - Nesse caso, o problema se resume a determinar o valor  $\hat{\theta}$  tal que  $\ell'(\theta|\mathbf{X}) = 0$
  - Em alguns casos, o valor de  $\hat{\theta}$  tal que  $\ell'(\theta|\mathbf{X}) = 0$  também pode ser obtido algebricamente, como veremos a seguir.

# 3. Estimadores de máxima verossimilhança

- Voltando ao exemplo 1 (durabilidade de componentes eletrônicos):
  - $L(\lambda|\mathbf{x}) = f(\mathbf{x}|\lambda) = \lambda^3 \exp(-\lambda[x_1 + x_2 + x_3]) = \lambda^3 \exp(-6.6\lambda)$
  - $\ell(\lambda|\mathbf{x}) = 3 \log \lambda - 6.6\lambda$
  - $\ell'(\lambda|\mathbf{x}) = \frac{3}{\lambda} - 6.6$
  - Igualando  $\ell'(\hat{\lambda}|\mathbf{x}) = 0 \implies \frac{3}{\hat{\lambda}} = 6.6 \implies \hat{\lambda} = \frac{3}{6.6} = \frac{1}{2.2} = 0.455$
- No caso mais geral: se  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  é uma AAS onde  $x_i \sim \text{exponencial}(\lambda)$ , então definimos  $s = \sum_i x_i$  e temos:
  - $L(\lambda|\mathbf{x}) = \lambda^n \exp(-s\lambda)$
  - $\ell(\lambda|\mathbf{x}) = n \log \lambda - s\lambda$
  - $\hat{\lambda} = \frac{n}{s} = \frac{1}{\bar{x}}$ , ou seja,  $\hat{\lambda}$  é o inverso da média amostral!

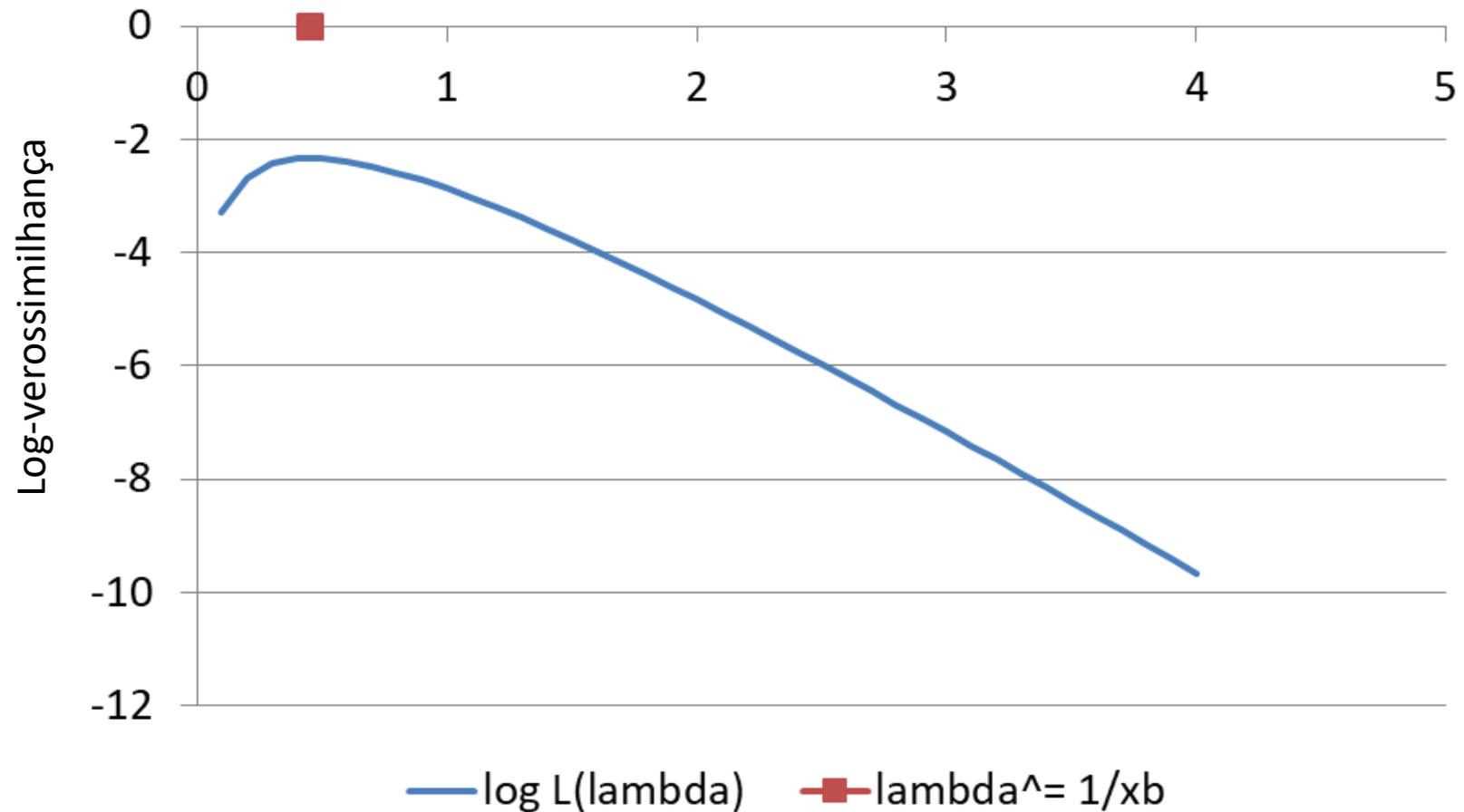
# 3. Estimadores de máxima verossimilhança

- Voltando ao exemplo 1:



# 3. Estimadores de máxima verossimilhança

- Voltando ao exemplo 1:



# 3. Estimadores de máxima verossimilhança

- Exemplo 2 (EMV do parâmetro da distribuição binomial):
  - Considere  $n$  tentativas independentes de Bernoulli com  $P(\text{sucesso}) = p$ ,  $0 < p < 1$  e  $X =$  número de sucessos
  - Dado o número  $x$  de sucessos observados, devemos tomar como estimador  $\hat{p}$  aquele valor de  $p$  que torna a amostra observada a mais provável de ocorrer
  - Como assumimos independência entre as tentativas, temos que  $X \sim \text{Bin}(n, p)$ :

$$f(x|n, p) = L(p|x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

- A função de log-verossimilhança é

$$\ell(p|x) = k + x \log p + (n-x) \log(1-p)$$

onde  $k = \log \frac{n!}{x!(n-x)!}$

# 3. Estimadores de máxima verossimilhança

- Exemplo 2 (cont):

- A função de log-verossimilhança é

$$\ell(p|x) = k + x \log p + (n - x) \log(1 - p)$$

onde  $k = \log \frac{n!}{x!(n-x)!}$

- Logo,

$$\ell'(p|x) = \frac{x}{p} - \frac{n - x}{1 - p}$$

(note que a parcela  $k$  foi suprimida, pois  $k$  é uma constante que não depende de  $p$ )

- Igualando  $\ell'(p|x) = 0$ , temos:

$$\ell'(\hat{p}|x) = \frac{x}{\hat{p}} - \frac{n - x}{1 - \hat{p}} = 0 \implies \frac{x}{\hat{p}} = \frac{n - x}{1 - \hat{p}} \implies x(1 - \hat{p}) = \hat{p}(n - x)$$

$$\implies x - x\hat{p} = n\hat{p} - x\hat{p} \implies x - x\hat{p} = n\hat{p} - x\hat{p} \implies x = n\hat{p} \implies \hat{p} = x/n$$

- Ou seja,  $\hat{p}$  corresponde à proporção observada de sucessos!

# 3. Estimadores de máxima verossimilhança

- Exemplo 3 (EMV da média para distribuição normal):
  - Seja  $X$  uma variável com distribuição normal com média  $\mu$  desconhecida e variância  $\sigma^2$  conhecida
  - Mostramos abaixo como derivar o estimador de máxima verossimilhança  $\hat{\mu}$ , dada uma AAS  $\mathbf{x} = (x_1, x_2, \dots, x_n)$
  - A função de verossimilhança é dada por

$$\begin{aligned} L(\mu|\mathbf{x}, \sigma^2) &= \prod_{i=1}^n f(x_i|\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x_i - \mu)^2}{2\sigma^2}\right] \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{\sum_{i=1}^n x_i^2 - 2x_i\mu + \mu^2}{2\sigma^2}\right] \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}\right] \end{aligned}$$

### 3. Estimadores de máxima verossimilhança

- Exemplo 3 (cont):

- Log-verossimilhança:

$$\begin{aligned}\ell(\mu|\mathbf{x}, \sigma^2) &= n \log \left( \frac{1}{\sigma\sqrt{2\pi}} \right) - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2} \\ &= k - \frac{q}{2\sigma^2} + \frac{2\mu s}{2\sigma^2} - \frac{n\mu^2}{2\sigma^2},\end{aligned}$$

onde  $k = n \log \left( \frac{1}{\sigma\sqrt{2\pi}} \right)$ ,  $q = \sum_{i=1}^n x_i^2$ ,  $s = \sum_{i=1}^n x_i$

(note que  $k$ ,  $q$  e  $s$  foram definidos apenas para simplificar as expressões)

- Derivada da log-verossimilhança (em relação a  $\mu$ ):

$$\ell'(\mu|\mathbf{x}, \sigma^2) = \frac{2s}{2\sigma^2} - \frac{2n\mu}{2\sigma^2} = \frac{s - n\mu}{\sigma^2}$$

- Igualando  $\ell'(\mu|\mathbf{x}, \sigma^2) = 0$ :

$$\ell'(\mu|\mathbf{x}, \sigma^2) = 0 \implies s - n\mu = 0 \implies \mu = \frac{s}{n} = \bar{x}$$

- Ou seja, a média amostral é o EMV para  $\mu$ .

### 3. Estimadores de máxima verossimilhança

- Nos exemplos anteriores, notamos que, para calcular a função de verossimilhança de um parâmetro  $\theta$  (e, portanto, estimar  $\theta$ ) a partir de uma AAS  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ , não precisaríamos ter conhecimento individual de cada observação da amostra, mas sim uma ou mais estatísticas (sumários) calculadas sobre essa amostra
- Quando não necessitamos conhecer todos os valores individuais de uma AAS  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  para calcular a função de verossimilhança  $L(\theta|\mathbf{x})$ , mas apenas uma estatística  $T(x_1, x_2, \dots, x_n)$ , dizemos que  $T$  é uma **estatística suficiente** para  $\theta$ 
  - Existe uma definição mais formal para estatística suficiente, mas que não será apresentada aqui.

# 3. Estimadores de máxima verossimilhança

- Exemplos:
  - Para calcular a função de verossimilhança  $L(\lambda|\mathbf{x})$  para a distribuição exponencial, basta conhecer  $s$ , a soma das observações (além de  $n$ , naturalmente)
    - Logo,  $s$  é estatística suficiente para  $\lambda$
  - Para calcular a função de verossimilhança  $L(p|\mathbf{x})$  para a distribuição binomial, basta conhecer o número de sucessos nas  $n$  tentativas,  $x$ 
    - Logo,  $x$  é estatística suficiente para  $p$
  - Para calcular a função de verossimilhança  $L(\mu|\mathbf{x}, \sigma^2)$  para a distribuição normal, basta conhecer  $s$  (soma das observações) e  $q$  (soma dos quadrados das observações)
    - Logo,  $(s, q)$  é estatística suficiente para  $\mu$

# 3. Estimadores de máxima verossimilhança

- Exercícios:

1. Seja  $X$  uma variável com distribuição normal com média  $\mu$  **conhecida** e variância  $\sigma^2$  **desconhecida**.

a) Derive o EMV para  $\sigma^2$ , dada uma AAS  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ .

b) Crie uma planilha em Excel para plotar  $L(\sigma^2)$  dentro de um certo intervalo de  $\sigma^2$

2. A fim de planejar a quantidade de atendentes e infraestrutura necessários, um call-center necessita deseja modelar o número de chamadas de clientes no intervalo de 1h, por dia da semana e horário específicos (p.ex. 2<sup>as</sup> feiras às 20:00).

Assume-se que o número de chamadas segue uma distribuição Poisson, com função de probabilidade (f.p.)

$$f(x|\lambda) = \frac{e^{-\lambda} \lambda^x}{x!},$$

onde  $\lambda$  (desconhecido) é a taxa de ocorrência do evento, ou seja, o número médio de chamadas de clientes por hora.

(para detalhes, ver DeGroot, Probability and Statistics, Sec. 5.4).

# 3. Estimadores de máxima verossimilhança

- Exercícios:

- 2. (cont)

- a) Derive o EMV para  $\lambda$ , dada uma AAS  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ .
    - b) Indique se há uma estatística suficiente para  $\lambda$  e, em caso afirmativo, indique qual.
    - c) Plote em Excel a função de verossimilhança de  $L(\lambda|\mathbf{x})$  dentro de um certo intervalo de  $\lambda$

## 6. Intervalos de confiança

- Todos os estimadores apresentados anteriormente foram *pontuais*, ou seja, especificam um único valor para o estimador
- Esse procedimento não permite julgar qual a possível magnitude do erro que estamos cometendo
- Surge daí a necessidade de construir intervalos estatísticos
- Intervalos de confiança: intervalos estatísticos baseados na distribuição amostral do estimador pontual

## 6. Intervalos de confiança

- Exemplo 11.12:
  - Suponha que queiramos estimar a média  $\mu$  de uma população qualquer, e para tanto usamos a média  $\bar{X}$  de uma amostra de tamanho  $n$ . Do TLC,

$$e = (\bar{X} - \mu) \sim N(0, \sigma_{\bar{X}}^2), \quad (11.32)$$

com  $Var(\bar{X}) = \sigma_{\bar{X}}^2 = \sigma/n$ .

- Desse resultado podemos determinar a probabilidade de cometermos erros de determinadas magnitudes

## 6. Intervalos de confiança

- Exemplo 11.12 (cont):

– Por exemplo,

$$P(|e| < 1,96\sigma_{\bar{X}}) = 0,95$$

ou

$$P(|\bar{X} - \mu| < 1,96\sigma_{\bar{X}}) = 0,95,$$

que é equivalente a

$$P(-1,96\sigma_{\bar{X}} < \bar{X} - \mu < 1,96\sigma_{\bar{X}}) = 0,95,$$

e, finalmente,

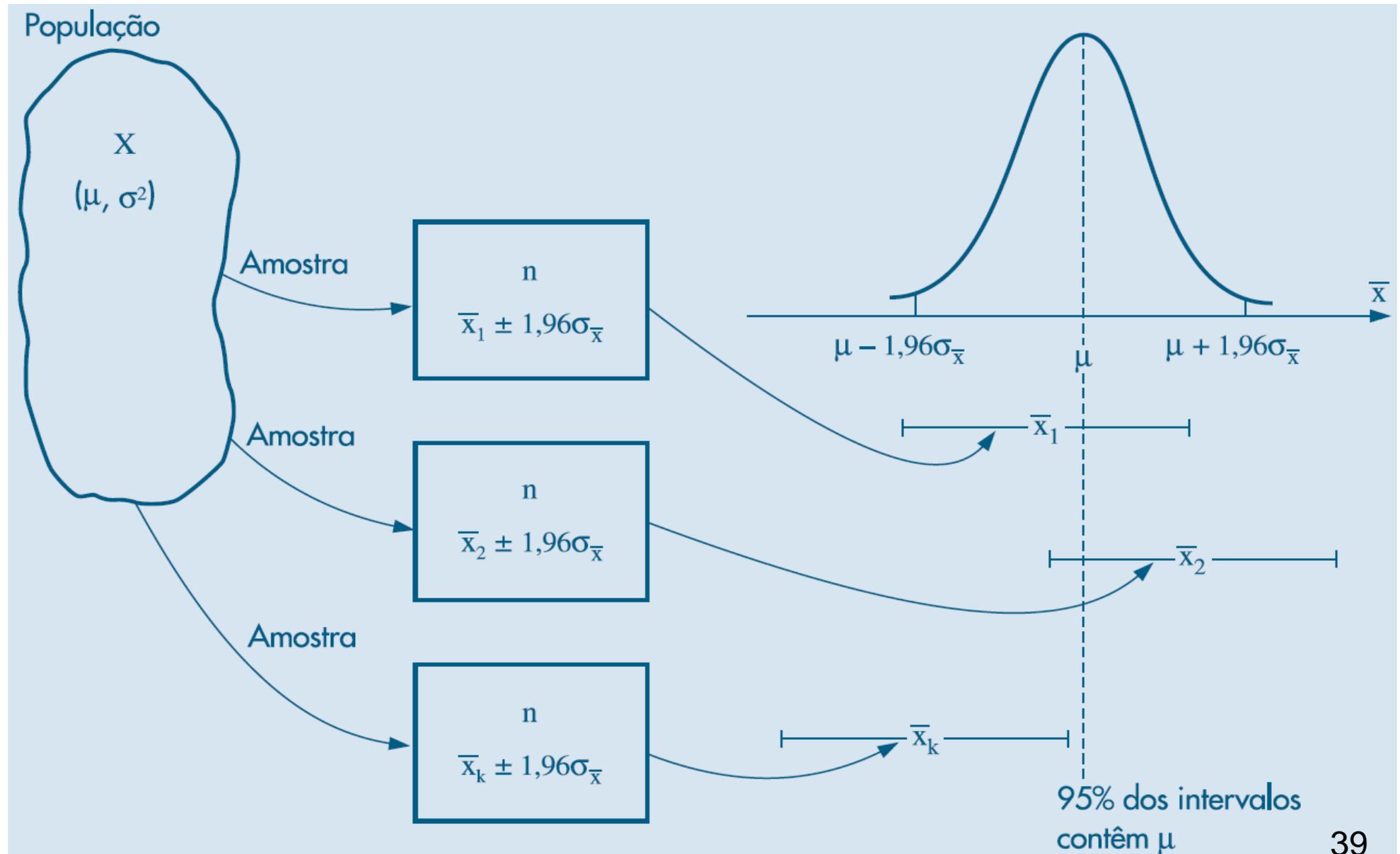
$$P(\bar{X} - 1,96\sigma_{\bar{X}} < \mu < \bar{X} + 1,96\sigma_{\bar{X}}) = 0,95. \quad (11.33)$$

## 6. Intervalos de confiança

- É importante lembrar que, sob o ponto de vista frequentista,  $\mu$  não é uma variável aleatória e sim um parâmetro, e a expressão (11.33) deve ser interpretada da seguinte maneira:
- Se pudéssemos construir uma quantidade grande de intervalos (aleatórios) da forma  $(\bar{X} - 1.96 \sigma_{\bar{X}}, \bar{X} + 1.96 \sigma_{\bar{X}})$ , todos baseados em amostras de tamanho  $n$ , 95% deles conteriam o parâmetro  $\mu$ .
  - Figura 11.13 mostra o funcionamento e o significado de um intervalo de confiança (IC) para  $\mu$ , com  $\gamma = 0.95$  e  $\sigma^2$  conhecido
- Dizemos que  $\gamma$  é o **coeficiente de confiança**
- Escolhida uma amostra e encontrada sua média  $\bar{x}_0$ , e admitindo-se  $\sigma^2$  conhecido, podemos construir o intervalo
$$(\bar{x}_0 - 1.96 \sigma_{\bar{X}}, \bar{x}_0 + 1.96 \sigma_{\bar{X}}) \quad (11.34)$$
- Esse intervalo pode ou não conter o parâmetro  $\mu$ , mas pelo exposto acima temos 95% de confiança de que contenha.

# 6. Intervalos de confiança

- Figura 11.3: significado de um IC para  $\mu$ , com  $\gamma = 0.95$

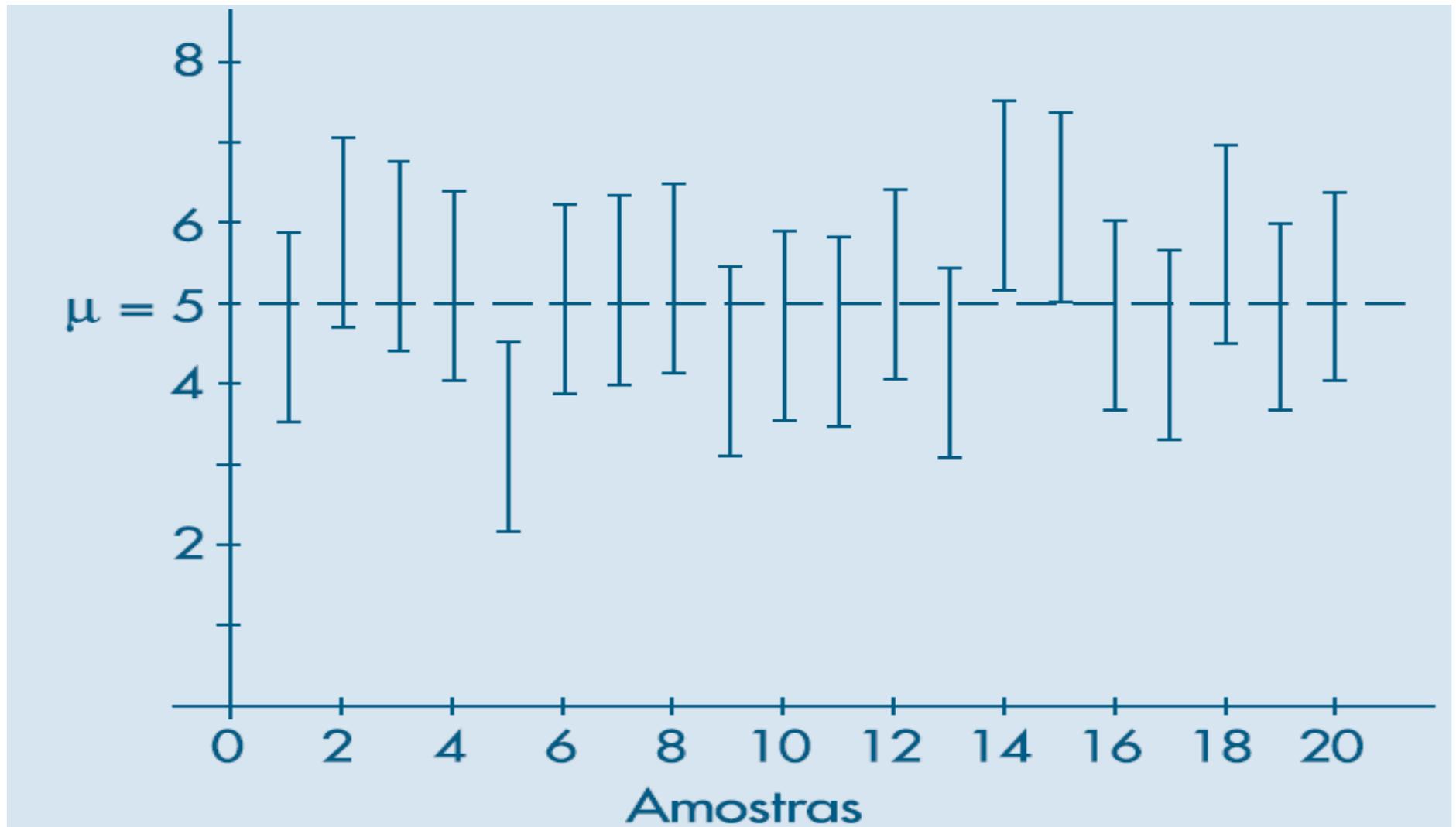


## 6. Intervalos de confiança

- Para ilustrar, consideremos o seguinte experimento de simulação:
  - Foram geradas 20 amostras de tamanho  $n = 25$  de uma distribuição normal com média  $\mu = 5$  e desvio padrão  $\sigma = 3$
  - Para cada amostra foi construído o intervalo de confiança para  $\mu$ , com coeficiente de confiança  $\gamma = 0.95$ :
$$\sigma_{\bar{X}} = \frac{3}{\sqrt{25}} = \frac{3}{5} = 0.6 \Rightarrow 1.96 \sigma_{\bar{X}} = 1.176$$
ou seja, cada intervalo é da forma  $\bar{X} \pm 1.176$
  - Figura 11.4 apresenta os intervalos obtidos, dentre os quais 3 intervalos (amostras 5, 14 e 15) não contêm a média  $\mu = 5$
- Exercício:
  - Repetir a simulação descrita acima, com 1000 amostras de tamanho  $n = 25$  e demais parâmetros similares ao exemplo
  - Calcular a proporção de intervalos contendo a média  $\mu = 5$

## 6. Intervalos de confiança

- Figura 11.4: Intervalos de confiança para a média de uma  $N(5,9)$ , para 20 amostras de tamanho  $n = 25$



## 6. Intervalos de confiança

- Usando a aproximação pela distribuição normal (Eq. 11.34), para um coeficiente de confiança qualquer  $\gamma$ , teremos que usar o valor  $z_\gamma$  tal que  $P(-z_\gamma \leq Z \leq z_\gamma) = \gamma$ , com  $Z \sim N(0,1)$ . O valor de  $z_\gamma$  é calculado por

$$z_\gamma = -\Phi^{-1}\left(\frac{1-\gamma}{2}\right)$$

(ver Seção 10.11 para detalhes)

- Outra forma de calcular  $z_\gamma$ :  $\alpha = (1 - \gamma)$ ;  $z_\gamma = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$
- O intervalo fica então:  
$$\left(\bar{X} - z_\gamma \sigma_{\bar{X}}, \bar{X} + z_\gamma \sigma_{\bar{X}}\right) \quad (11.37)$$
- Observe que a amplitude do intervalo (11.37) é  $L = 2z_\gamma \sigma / \sqrt{n}$ , que é uma constante, independente de  $\bar{X}$ 
  - Se construirmos vários intervalos de confiança para o mesmo valor de  $n, \sigma$  e  $\gamma$ , estes terão extremos aleatórios, mas todos terão a mesma amplitude  $L$

## 6. Intervalos de confiança

- Se a variância populacional  $\sigma^2$  não for conhecida, podemos substituir em (11.37)  $\sigma_{\bar{X}}$  por  $\sqrt{S^2/n}$ , onde  $S^2$  é a variância amostral dada pela Eq. (11.9)
  - Para  $n$  grande, da ordem de 100, o intervalo (11.37), com essa modificação, ainda pode ser usado
  - Para  $n$  não muito grande, a distribuição normal não pode mais ser usada e terá de ser substituída pela distribuição  $t$  de Student (a ser apresentada posteriormente)

## 6. Intervalos de confiança - proporções

- Consideremos o problema de obter um intervalo de confiança para o parâmetro  $p$  de uma distribuição  $Bin(n, p)$ 
  - Sabemos que se  $Y$  é o número de sucessos nas  $n$  provas, então  $\hat{p} = Y/n$  tem distribuição aproximadamente normal, com média  $\mu = p$  e variância  $pq/n$ , onde  $q = 1 - p$
  - Ou ainda,

$$Z = \frac{\hat{p} - p}{\sqrt{pq/n}} \sim N(0,1)$$

- Dado o coeficiente de confiança  $\gamma$ , desejamos encontrar o intervalo  $(-z_\gamma, z_\gamma)$  tal que

$$P(-z_\gamma \leq Z \leq z_\gamma) = \gamma$$

o que implica

$$P\left(\hat{p} - z_\gamma \sqrt{pq/n} \leq p \leq \hat{p} + z_\gamma \sqrt{pq/n}\right) = \gamma$$

## 6. Intervalos de confiança - proporções

- Como não conhecemos  $p$ , pode-se proceder de duas formas:
- A primeira é usar um limitante superior para a variância de  $\hat{p}$ : sabemos que  $pq \leq 1/4$ , o que implica que

$$\sqrt{pq/n} \leq 1/\sqrt{4n}$$

- Assim, substituindo  $\sqrt{pq/n}$  por  $1/\sqrt{4n}$ , o intervalo

$$\hat{p} \pm z_\gamma/\sqrt{4n} \quad (11.40)$$

é um intervalo de confiança para  $p$ , com coeficiente de confiança  $\gamma$

- Intervalo conservador

- Outra forma possível é estimar  $\hat{p}$  a partir de uma amostra preliminar:  $\hat{p} = Y/n$ ; o intervalo de confiança será então dado por

$$\hat{p} \pm z_\gamma \sqrt{\hat{p}\hat{q}/n}, \quad (11.41)$$

onde  $\hat{q} = (1 - \hat{p})$ .

## 6. Intervalos de confiança - proporções

- Exemplo 11.16:
  - Suponha que em  $n = 400$  provas obtemos  $y = 80$  sucessos. Vamos obter um intervalo de confiança para  $p$  com  $\gamma = 0.9$
  - Como  $\hat{p} = 80/400 = 0.2$  e  $\hat{q} = 1 - \hat{p} = 0.8$ , então (11.41) fica
$$0.2 \pm (1.645)\sqrt{(0.2)(0.8)/400} = 0.2 \pm 0.033,$$
ou seja,
$$IC(p; 0.9) = (0,167, 0.233).$$
  - Usando (11.40), o intervalo conservador é
$$IC(p; 0.9) = (0,159, 0.241)$$
- Observações:
  - Intervalo dado por (11.41) tem amplitude menor do que aquele dado por (11.40)
  - Intervalo dado por (11.40) depende apenas de  $n$  e  $\gamma$ , mas não da amostra coletada

## 6. Intervalos de confiança - proporções

- Problemas com a aproximação do intervalo de confiança para proporções pelo TLC:
  - Limites do intervalo fora do domínio  $[0,1]$ 
    - Limite inferior  $< 0$  ou limite superior  $> 1$
  - Intervalo degenerado se  $\hat{p} = 0$  ou  $\hat{p} = 1$ 
    - Não é possível calcular a variância usando a estimativa  $\hat{p}$ ;
    - Nesse caso, pode-se substituir  $\hat{p}$  por  $1/2$
- Algumas variantes:
  - Intervalo de Agresti-Coull
  - Intervalo (Bayesiano) de Jeffreys
  - Intervalo Bayesiano de credibilidade
    - intervalo de máxima densidade de probabilidade (HPD interval)

## 6. Intervalos de confiança - proporções

- Intervalo de Agresti-Coull:

- Outro intervalo de confiança aproximado para distribuições binomiais

- Dados  $Y$  sucessos em  $n$  tentativas, define-se

$$\tilde{n} = n + z_\gamma^2 \text{ e } \tilde{p} = \frac{1}{\tilde{n}} \left( Y + \frac{1}{2} z_\gamma^2 \right).$$

- Então, um intervalo de confiança para  $p$  é dado por

$$\tilde{p} \pm z_\gamma \sqrt{\tilde{p} (1 - \tilde{p}) / \tilde{n}}$$

- Para  $\gamma = 0.95$ , tem-se  $z_\gamma \approx 1.96$  e  $z_\gamma^2 \approx 3.84$

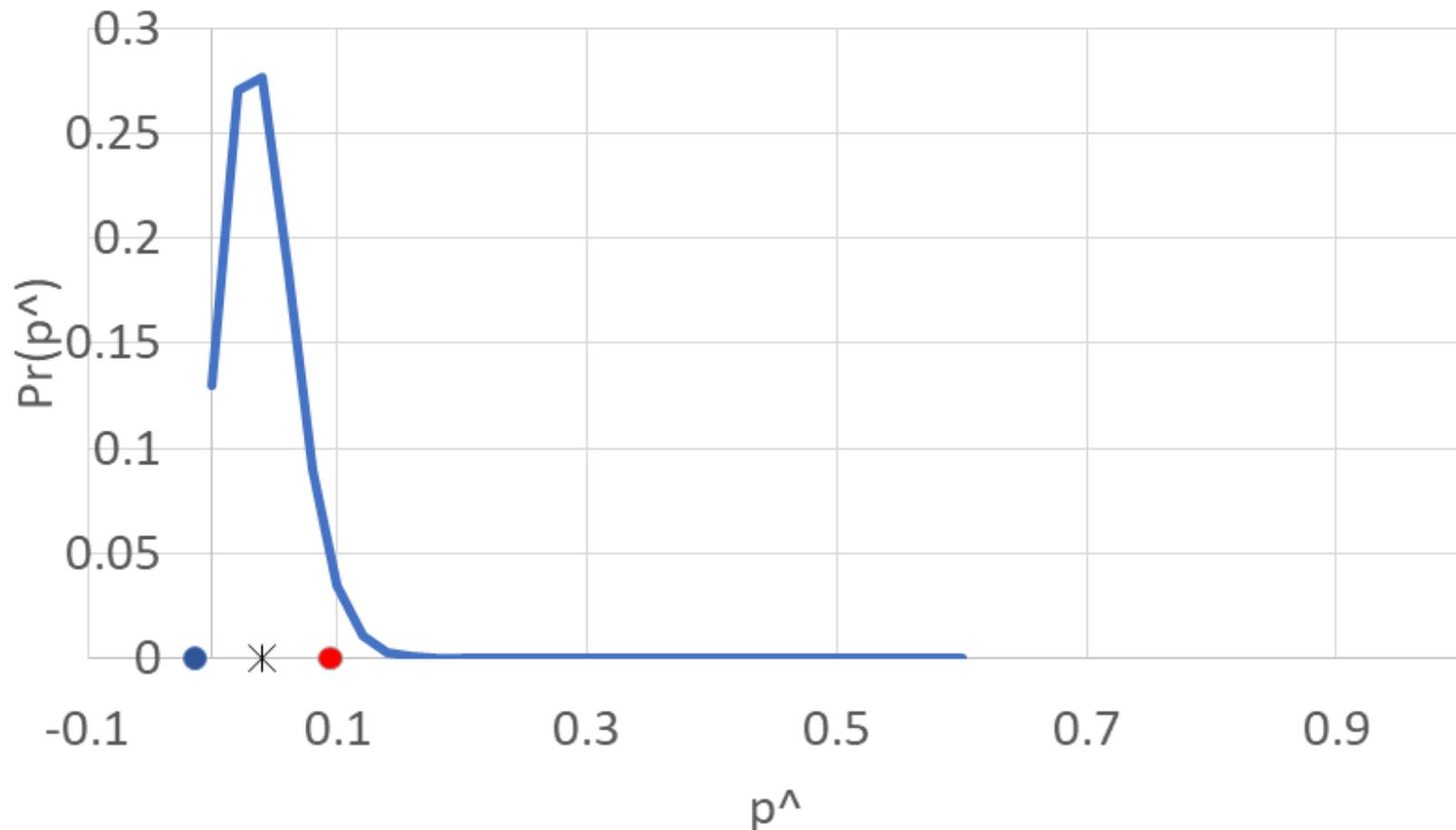
- Se usarmos 2 ao invés de 1.96, isso equivale a “adicionar dois sucessos e dos fracassos” no intervalo dado pela Eq. (11.41)

- Referência:

- [https://en.wikipedia.org/wiki/Binomial\\_proportion\\_confidence\\_interval](https://en.wikipedia.org/wiki/Binomial_proportion_confidence_interval)

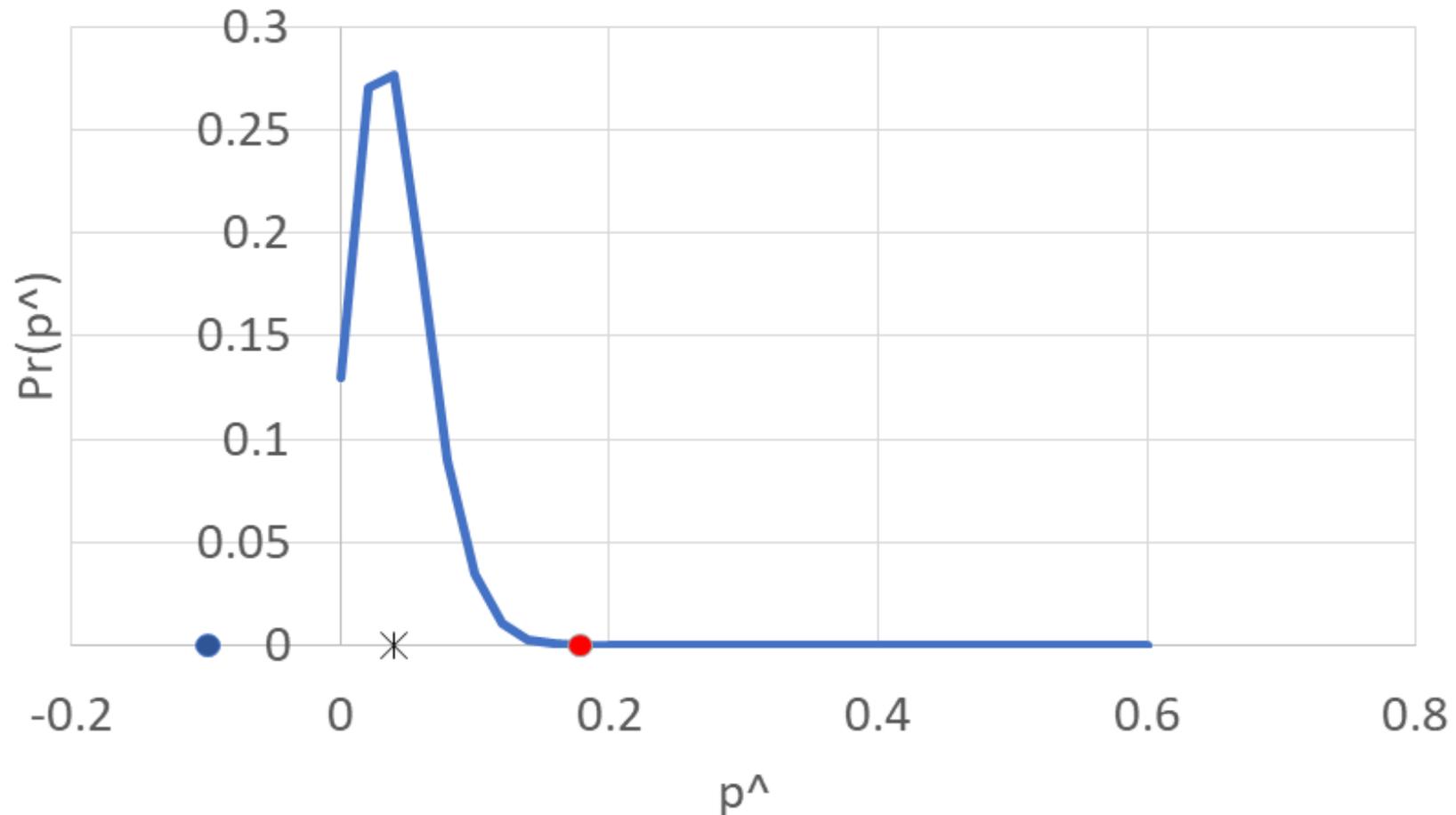
## 6. Intervalos de confiança - proporções

- Exemplo 1 (TLC, usando  $\sqrt{p(1-p)/n}$  como erro padrão):
  - $n = 50, y = 2 \Rightarrow \hat{p} = 0.04, LI = -0.014, LS = 0.094$ 
    - Limite inferior do intervalo (LI)  $< 0$



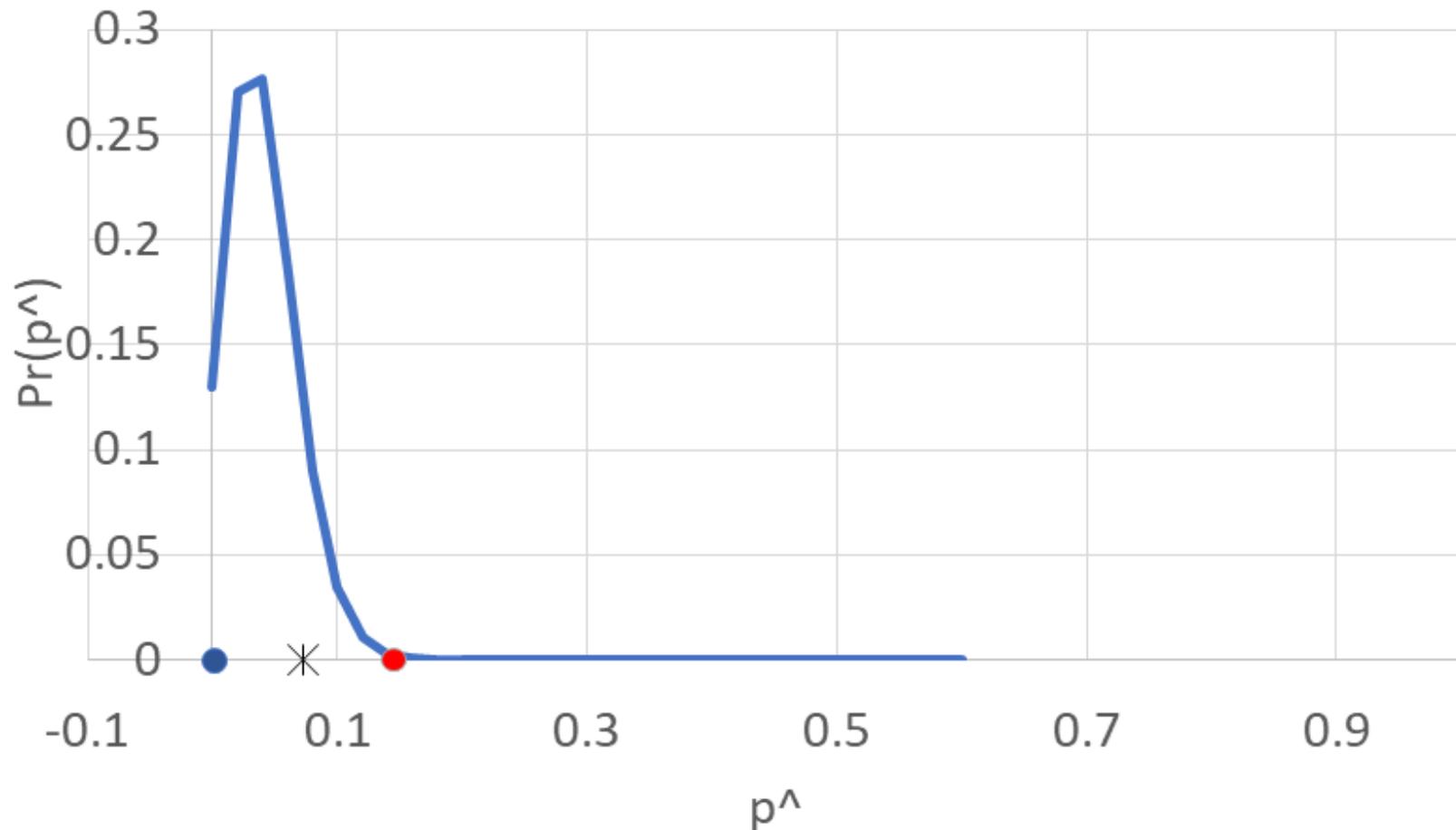
## 6. Intervalos de confiança - proporções

- Exemplo 1 (TLC, usando  $1/\sqrt{4n}$  como erro padrão):
  - $n = 50, y = 2 \Rightarrow \hat{p} = 0.04, LI = -0.099, LS = 0.179$ 
    - Limite inferior do intervalo (LI)  $< 0$ ; intervalo conservador



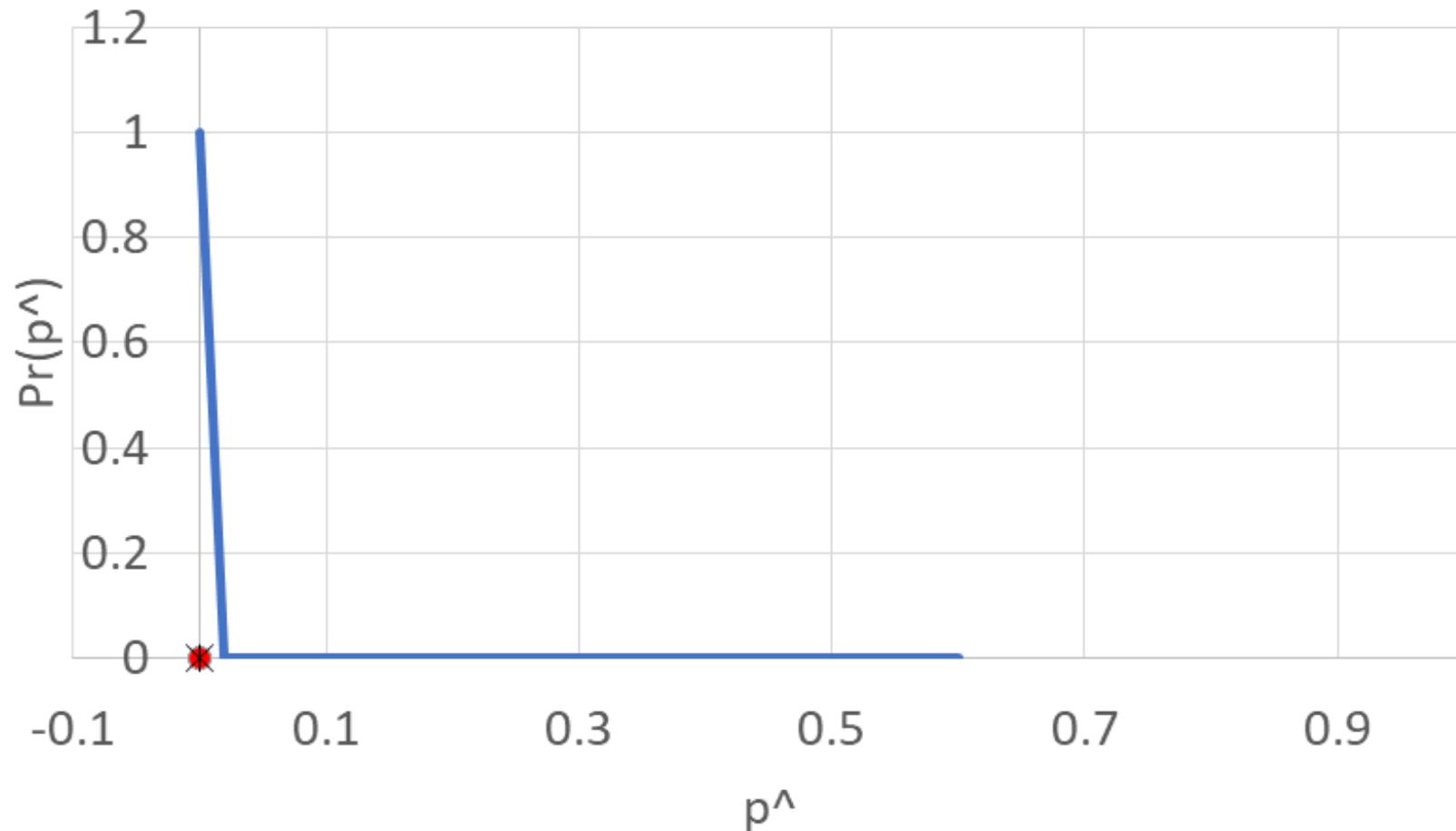
## 6. Intervalos de confiança - proporções

- Exemplo 1 (intervalo de Agresti-Coull):
  - $n = 50, y = 2 \Rightarrow \hat{p} = 0.04, LI = 0.001, LS = 0.145$



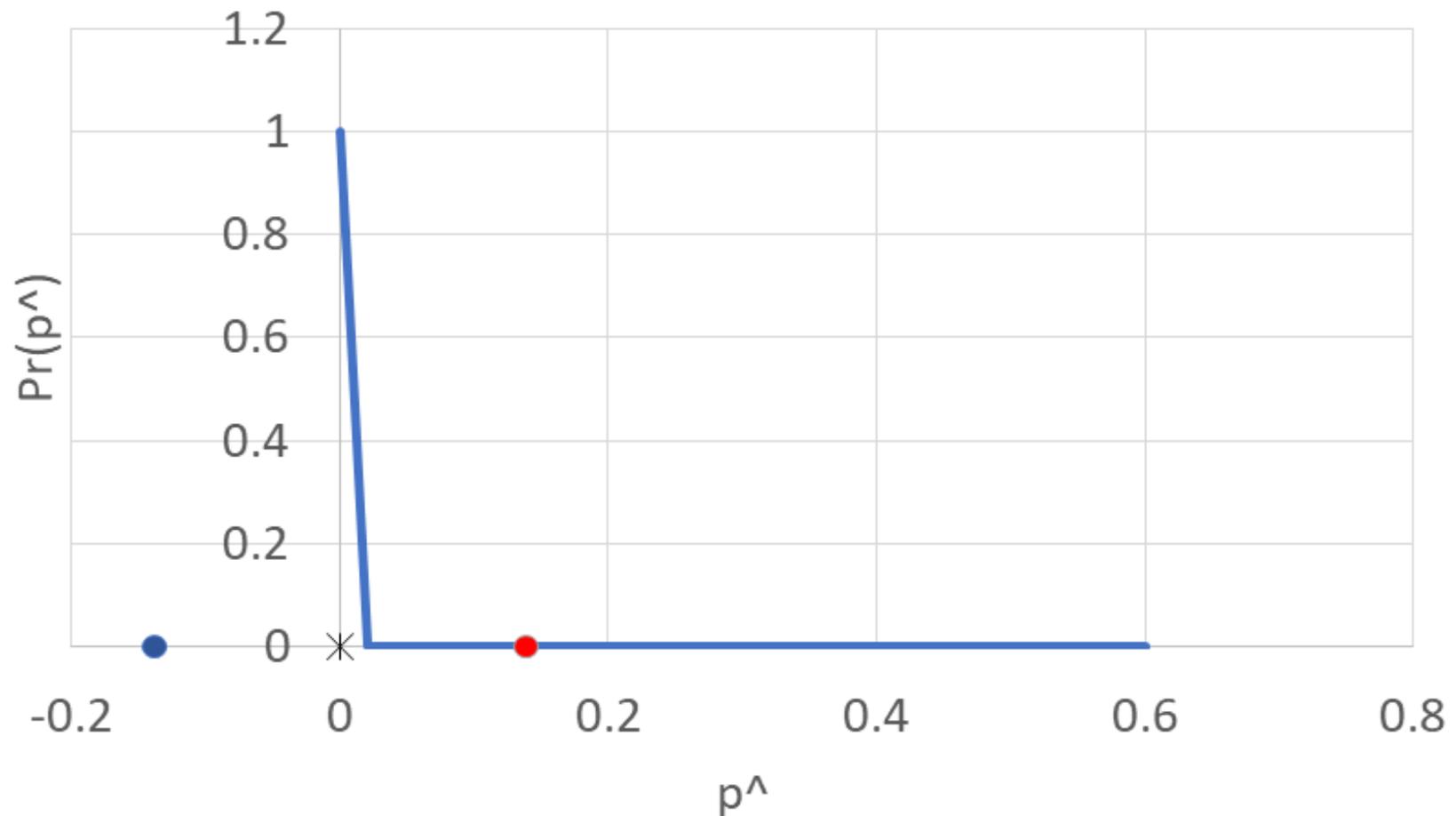
## 6. Intervalos de confiança - proporções

- Exemplo 2 (TLC, usando  $\sqrt{p(1-p)/n}$  como erro padrão):
  - $n = 50, y = 0 \Rightarrow \hat{p} = 0, LI = 0, LS = 0$ 
    - Intervalo degenerado



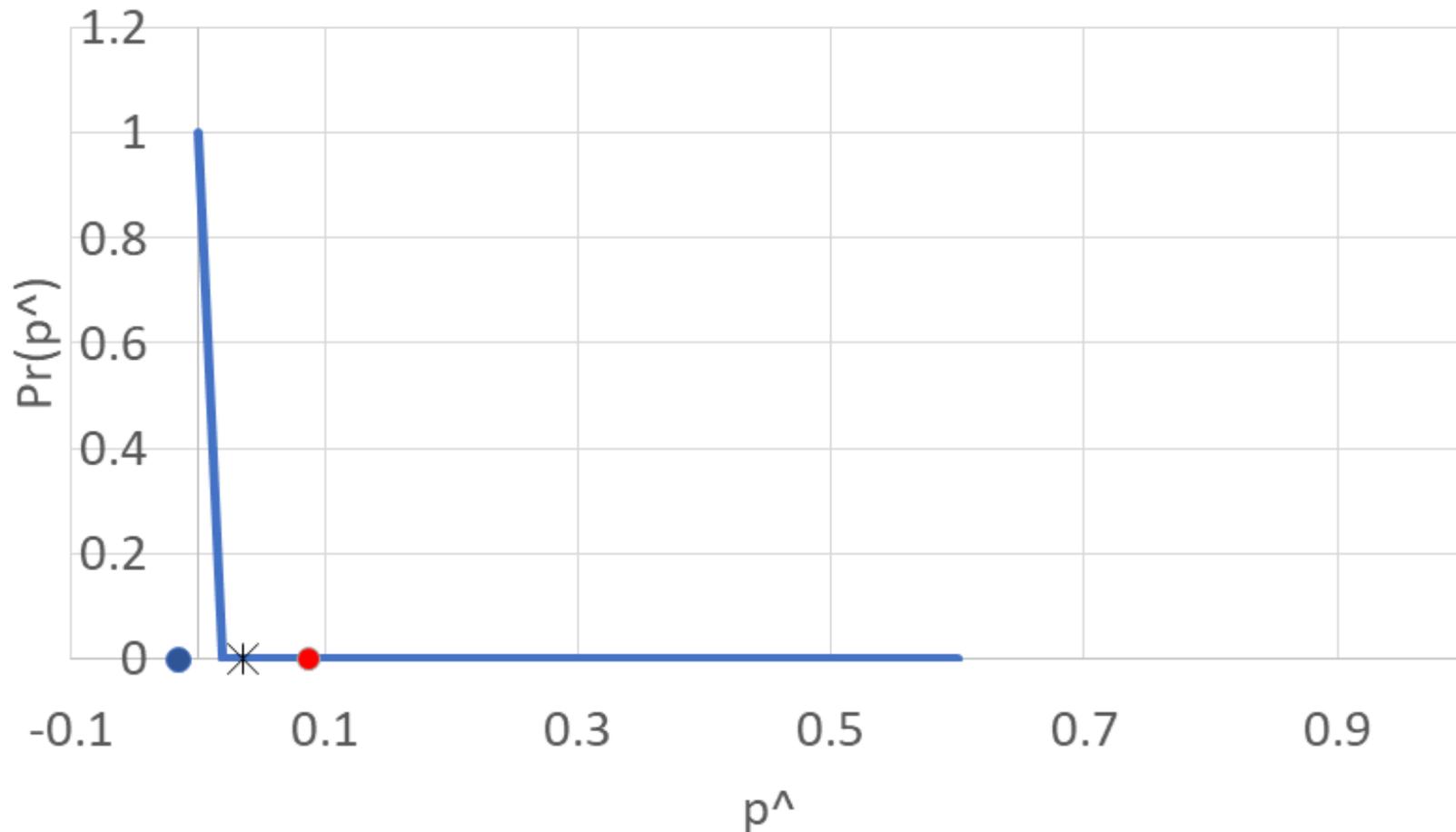
## 6. Intervalos de confiança - proporções

- Exemplo 2 (TLC, usando  $1/\sqrt{4n}$  como erro padrão):
  - $n = 50, y = 0 \Rightarrow \hat{p} = 0, LI = -0.139, LS = 0.139$ 
    - Intervalo não degenerado, mas  $LI < 0$ ; intervalo conservador



# 6. Intervalos de confiança - proporções

- Exemplo 2 (intervalo de Agresti-Coull):
  - $n = 50, y = 0 \Rightarrow \hat{p} = 0, LI = -0.016, LS = 0.087$
  - Intervalo não degenerado, mas  $LI < 0$



# 6. Intervalos de confiança - proporções

- Problemas:

14. Calcule o intervalo de confiança para a média de uma  $N(\mu, \sigma^2)$  em cada um dos casos abaixo.

Média Amostral	Tamanho da Amostra	Desvio Padrão da População	Coefficiente de Confiança
170 cm	100	15 cm	95%
165 cm	184	30 cm	85%
180 cm	225	30 cm	70%

15. De 50.000 válvulas fabricadas por uma companhia retira-se uma amostra de 400 válvulas, e obtém-se a vida média de 800 horas e o desvio padrão de 100 horas.

(a) Qual o intervalo de confiança de 99% para a vida média da população?

(b) Com que confiança dir-se-ia que a vida média é  $800 \pm 0,98$ ?

(c) Que tamanho deve ter a amostra para que seja de 95% a confiança na estimativa  $800 \pm 7,84$ ?

(Que suposições você fez para responder às questões acima?)



# 6. Intervalos de confiança - proporções

- Problemas:

20. Antes de uma eleição, um determinado partido está interessado em estimar a proporção  $p$  de eleitores favoráveis ao seu candidato. Uma amostra piloto de tamanho 100 revelou que 60% dos eleitores eram favoráveis ao candidato em questão.
- (a) Determine o tamanho da amostra necessário para que o erro cometido na estimação seja de, no máximo, 0,01 com probabilidade de 80%.
  - (b) Se na amostra final, com tamanho igual ao obtido em (a), observou-se que 55% dos eleitores eram favoráveis ao candidato em questão, construa um intervalo de confiança para a proporção  $p$ . Utilize  $\gamma = 0,95$ .
21. Suponha que estejamos interessados em estimar a proporção de consumidores de um certo produto. Se a amostra de tamanho 300 forneceu 100 indivíduos que consomem o dado produto, determine:
- (a) o intervalo de confiança para  $p$ , com coeficiente de confiança de 95% (interprete o resultado);
  - (b) o tamanho da amostra para que o erro da estimativa não exceda a 0,02 unidades com probabilidade de 95% (interprete o resultado).

Em todos os problemas anteriores, recalcule os intervalos de confiança para proporções usando o método de Agresti-Coull

# Intervalos Bayesianos de credibilidade

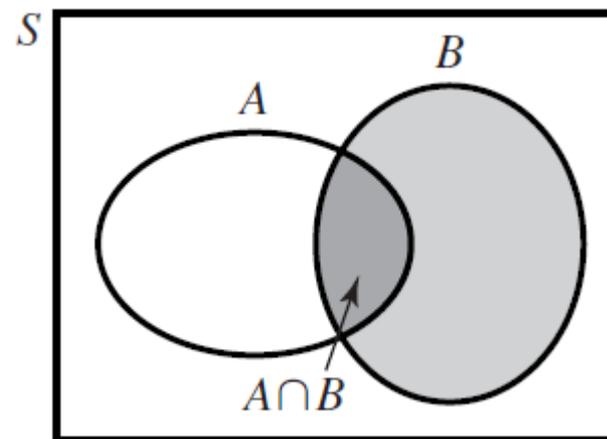
- Vimos que os intervalos de confiança frequentistas tratam o valor verdadeiro do parâmetro  $\theta$  como um valor fixo, e a variável aleatória é o próprio intervalo
  - É por essa razão que não se pode afirmar que a probabilidade do valor verdadeiro de  $\theta$  pertencer ao intervalo calculado é  $\gamma$  – o que se afirma é que a confiança do intervalo conter o valor verdadeiro de  $\theta$  é  $\gamma$
- Na estatística Bayesiana, o próprio parâmetro é tratado como uma variável aleatória, o que torna natural associar probabilidades do valor verdadeiro do parâmetro estar em certas regiões ou intervalos
- Veremos a seguir como calcular intervalos de probabilidade (ou de credibilidade) para o parâmetro  $\theta$
- Para isso, recapitularemos o conceito de probabilidade condicional e Teorema de Bayes, bem como veremos a definição de distribuições *a priori* e *a posteriori*

# Intervalos Bayesianos de credibilidade

- Probabilidade condicional – definição:
  - Suponha que aprendemos que um evento  $B$  ocorreu e queremos calcular a probabilidade de outro evento  $A$  levando em conta que sabemos da ocorrência de  $B$ .
  - A nova probabilidade de  $A$  é denominada a **probabilidade condicional de que o evento  $A$  ocorreu dado que o evento  $B$  tenha ocorrido**. Notação:  $\Pr(A|B)$ .
  - Se  $\Pr(B) > 0$ , calculamos essa probabilidade como

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

- Pela definição acima, é imediato que  $\Pr(A \cap B) = \Pr(B) \Pr(A|B)$



# Intervalos Bayesianos de credibilidade

- **Partição – definição:**

- Seja  $S$  o espaço amostral de um experimento, e considere  $k$  eventos disjuntos  $B_1, B_2, \dots, B_k$  em  $S$  tais que  $\bigcup_{i=1}^k B_i = S$ . Dizemos que esses eventos formam uma **partição** de  $S$ .

- **Lei da probabilidade total:**

- Suponha que os eventos  $B_1, B_2, \dots, B_k$  formam uma partição do espaço  $S$  e que  $\Pr(B_j) > 0$  para  $j=1, \dots, k$ . Então, para qualquer evento  $A$  em  $S$ ,

$$\Pr(A) = \sum_{j=1}^k \Pr(B_j) \Pr(A|B_j).$$

# Intervalos Bayesianos de credibilidade

- **Lei da probabilidade total (cont):**

- Dem: Os eventos  $B_1 \cap A$ ,  $B_2 \cap A$ , ...,  $B_k \cap A$  formam uma partição de  $A$ , como ilustrado na figura abaixo. Assim, podemos escrever

$$A = (B_1 \cap A) \cup (B_2 \cap A) \cup \dots \cup (B_k \cap A).$$

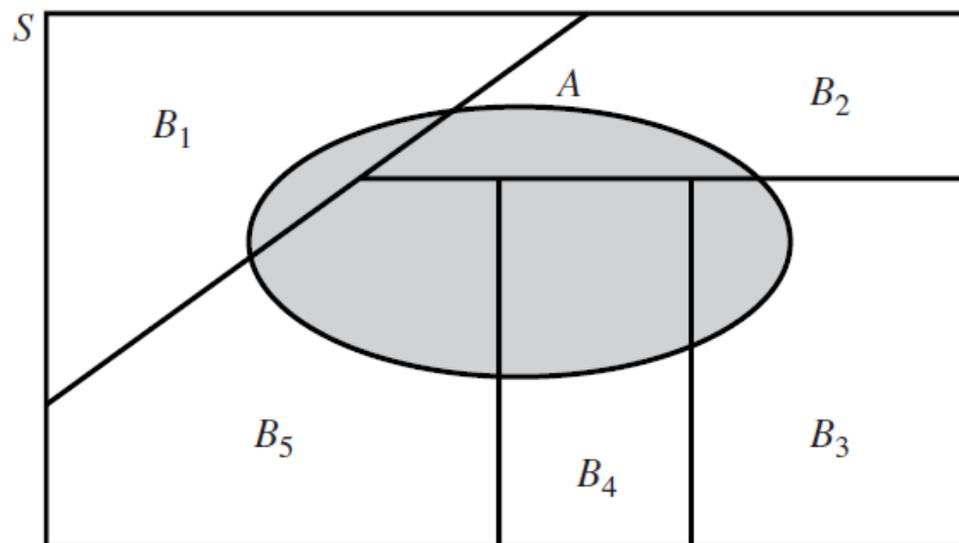
Além disso, como os  $k$  eventos do lado direito da equação são disjuntos, temos:

$$\Pr(A) = \sum_{j=1}^k \Pr(B_j \cap A).$$

Como  $\Pr(B_j) > 0$  para todo  $j$ , vale

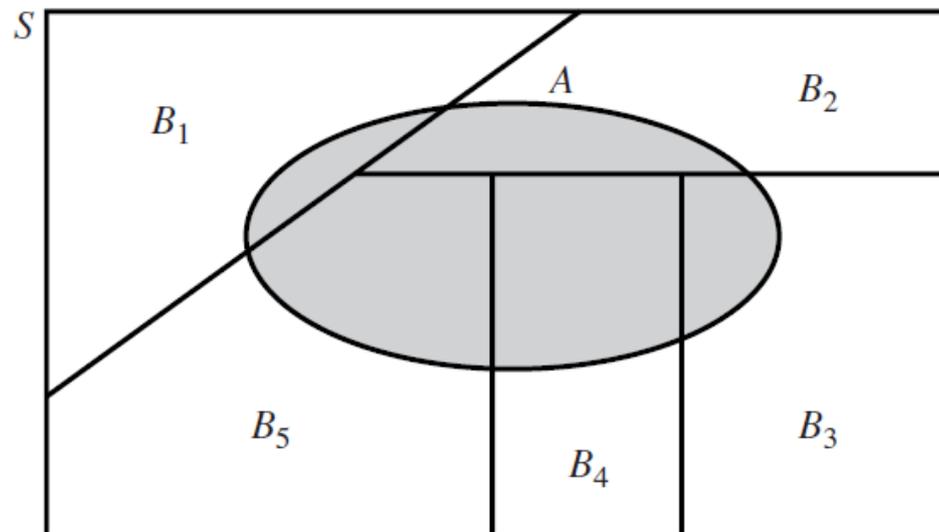
$$\Pr(B_j \cap A) = \Pr(B_j) \Pr(A|B_j)$$

e o resultado é imediato.



# Intervalos Bayesianos de credibilidade

- Intuição (e demonstração) do teorema de Bayes:
  - Suponha que você esteja interessado em obter as probabilidades de  $k$  eventos  $B_1, B_2, \dots, B_k$  após a realização de um experimento
    - Esses eventos formam um partição de  $S \rightarrow$  um deles deve ocorrer
  - Você possui estimativas iniciais de  $\Pr(B_i) > 0, i = 1 \dots k$ .
  - Seu interesse é calcular as probabilidades condicionais  $\Pr(B_i | A)$ ,  $i = 1, \dots, k$ , dado que evento  $A$  ocorreu no experimento.



# Intervalos Bayesianos de credibilidade

- Intuição (e demonstração) do teorema de Bayes:
  - Seu interesse é calcular as probabilidades condicionais  $\Pr(B_i|A)$ ,  $i = 1, \dots, k$ , dado que evento  $A$  ocorreu no experimento.

– **Resposta:**

– Da definição de probabilidade condicional:

$$\Pr(B_i|A) = \frac{\Pr(B_i \cap A)}{\Pr(A)} = \frac{\Pr(B_i) \Pr(A|B_i)}{\Pr(A)}, \quad i = 1, \dots, k$$

– Agora, pela lei da probabilidade total:

$$\Pr(A) = \sum_{j=1}^k \Pr(B_j \cap A) = \sum_{j=1}^k \Pr(B_j) \Pr(A|B_j)$$

– Substituindo  $\Pr(A)$  no denominador da 1ª equação, temos:

$$\Pr(B_i|A) = \frac{\Pr(B_i) \Pr(A|B_i)}{\sum_{j=1}^k \Pr(B_j) \Pr(A|B_j)}.$$

# Intervalos Bayesianos de credibilidade

- **Teorema de Bayes**

Sejam  $B_1, B_2, \dots, B_k$  eventos que formam uma partição do espaço  $S$  tais que  $\Pr(B_i) > 0, i = 1 \dots k$ , e seja  $A$  um evento tal que  $\Pr(A) > 0$ . Então, para  $i = 1 \dots k$ ,

$$\Pr(B_i|A) = \frac{\Pr(B_i) \Pr(A|B_i)}{\sum_{i=1}^k \Pr(B_i) \Pr(A|B_i)} .$$

- $\Pr(B_i|A)$  é calculado exclusivamente a partir das probabilidades a priori  $\Pr(B_i)$  e da verossimilhança  $\Pr(A|B_i)$  de  $A$  ocorrer, condicionado a cada um dos eventos alternativos  $B_1, B_2, \dots, B_k$

# Intervalos Bayesianos de credibilidade

- Teorema de Bayes – Exemplo: Teste de uma doença
  - Você está caminhando na rua e nota que o posto de saúde está fornecendo um teste gratuito para uma certa doença.
  - O teste tem a seguinte confiabilidade:
    - *Sensibilidade*: Se uma pessoa tem a doença, o teste tem 90% de probabilidade de dar um resultado positivo
    - *Especificidade*: Se uma pessoa não tem a doença, o teste tem 90% de probabilidade de dar um resultado negativo. (Portanto, só 10% de probabilidade de dar resultado positivo)
  - Dados epidemiológicos indicam que prevalência da doença é de apenas 1 em 10.000, mas como o teste é gratuito e não invasivo, você decide fazer.
  - Alguns dias depois você recebe uma carta informando que seu teste deu positivo.
  - Com base nessa informação, qual é a probabilidade de você realmente ter a doença?

# Intervalos Bayesianos de credibilidade

- Teorema de Bayes – Exemplo: Teste de uma doença (cont)

- Resp: Usaremos a seguinte notação:

- $D$ : você possui a doença;  $D^C$ : você não possui a doença
- $Pos$ : Seu teste deu positivo
- Prevalência da doença:  $\Pr(D) = 1/10000$
- $\Pr(Pos|D) = 0.9$ ;  $\Pr(Pos|D^C) = 0.1$

$$\begin{aligned}\Pr(D|Pos) &= \frac{\Pr(D) \Pr(Pos|D)}{\Pr(D) \Pr(Pos|D) + \Pr(D^C) \Pr(Pos|D^C)} \\ &= \frac{0.0001 \times 0.9}{0.0001 \times 0.9 + 0.9999 \times 0.1} = 0.00090.\end{aligned}$$

- A probabilidade condicional de você ter a doença dado o resultado do teste aumentou 9x, mas continua baixa: aprox. 1 em 1000

# Intervalos Bayesianos de credibilidade

- Teorema de Bayes – Exemplo: Teste de uma doença (cont)
    - A probabilidade condicional de você ter a doença dado o resultado do teste aumentou 9x, mas continua baixa: aprox. 1 em 1000
    - Explicação:
      - Doença é relativamente rara (1 em 10.000)
      - O teste tem uma alta taxa de resultados positivos (1 em 10), ou seja, o número de positivos pelo teste é 1000 vezes o número de pessoas que realmente têm a doença.
      - Em outras palavras, de cada 1.000 pessoas para as quais o teste dá um resultado positivo, apenas uma pessoa tem a doença.
    - Questões:
      1. E se o teste fosse mais robusto?
        - » P.ex. com menor número de falsos positivos?
      2. E se o teste tivesse dado negativo?
- Exercícios no próximo slide

# Intervalos Bayesianos de credibilidade

- Teorema de Bayes – Exemplo: Teste de uma doença (cont)
  - Neste exemplo, a evidência de você ter a doença aumentaria significativamente se a taxa de falsos positivos do teste fosse menor
  - Exercício 1:
    - Implemente este exemplo em Excel, deixando parametrizáveis:
      - $\Pr(D)$ ,  $\Pr(Pos|D)$ ,  $\Pr(Pos|D^C)$
    - Teste diferentes valores de  $\Pr(Pos|D)$  e  $\Pr(Pos|D^C)$  :
      - $\Pr(Pos|D) = 0.9, 0.99, 0.999, \text{ etc}$
      - $\Pr(Pos|D^C) = 0.1, 0.01, 0.001, 0.0001, \text{ etc.}$
    - Observe que  $\Pr(D|Pos)$  é mais sensível à taxa de falsos positivos –  $\Pr(Pos|D^C)$  – do que à taxa de verdadeiros positivos –  $\Pr(Pos|D)$
  - Exercício 2:
    - Suponha que seu teste deu negativo (chamemos de evento Neg)
    - Calcule a probabilidade de você não ter a doença, dado esse resultado:
      - $\Pr(D|Neg)$

# Intervalos Bayesianos de credibilidad

- Teorema de Bayes - Exemplo 2.3.4:

Identifying the Source of a Defective Item. Three different machines  $M_1$ ,  $M_2$ , and  $M_3$  were used for producing a large batch of similar manufactured items. Suppose that 20 percent of the items were produced by machine  $M_1$ , 30 percent by machine  $M_2$ , and 50 percent by machine  $M_3$ . Suppose further that 1 percent of the items produced by machine  $M_1$  are defective, that 2 percent of the items produced by machine  $M_2$  are defective, and that 3 percent of the items produced by machine  $M_3$  are defective. Finally, suppose that one item is selected at random from the entire batch and it is found to be defective. We shall determine the probability that this item was produced by machine  $M_2$ .

# Intervalos Bayesianos de credibilidade

- Teorema de Bayes - Exemplo 2.3.4 (solução):

Let  $B_i$  be the event that the selected item was produced by machine  $M_i$  ( $i = 1, 2, 3$ ), and let  $A$  be the event that the selected item is defective. We must evaluate the conditional probability  $\Pr(B_2|A)$ .

The probability  $\Pr(B_i)$  that an item selected at random from the entire batch was produced by machine  $M_i$  is as follows, for  $i = 1, 2, 3$ :

$$\Pr(B_1) = 0.2, \quad \Pr(B_2) = 0.3, \quad \Pr(B_3) = 0.5.$$

Furthermore, the probability  $\Pr(A|B_i)$  that an item produced by machine  $M_i$  will be defective is

$$\Pr(A|B_1) = 0.01, \quad \Pr(A|B_2) = 0.02, \quad \Pr(A|B_3) = 0.03.$$

It now follows from Bayes' theorem that

$$\begin{aligned} \Pr(B_2|A) &= \frac{\Pr(B_2) \Pr(A|B_2)}{\sum_{j=1}^3 \Pr(B_j) \Pr(A|B_j)} \\ &= \frac{(0.3)(0.02)}{(0.2)(0.01) + (0.3)(0.02) + (0.5)(0.03)} = 0.26. \end{aligned}$$

# Intervalos Bayesianos de credibilidade

- Probabilidades *a priori* e *a posteriori*:
  - No exemplo 2.3.4, uma probabilidade como  $\Pr(B_2)$  é comumente denominada a probabilidade *a priori* de que um item selecionado ao acaso tenha sido produzido pela máquina  $M_2$ 
    - $\Pr(B_2)$  é a probabilidade do evento  $B_2$  antes do item ser selecionado e antes de sabermos se o item selecionado é defeituoso ou não
  - As probabilidades  $\Pr(A|B_j)$  são as verossimilhanças
  - Uma probabilidade como  $\Pr(B_2|A)$  é usualmente denominada probabilidade *a posteriori* de que o item selecionado tenha sido produzido pela máquina  $M_2$ 
    - $\Pr(B_2|A)$  é a probabilidade do evento  $B_2$  depois de sabermos que o item selecionado é defeituoso
  - Analogamente, no exemplo do teste de uma doença,  $\Pr(D)$  é a probabilidade a priori de um indivíduo ter a doença (antes da realização do exame), e  $\Pr(D|Pos)$  é a probabilidade a posteriori, após termos conhecimento de que o resultado do teste foi positivo

# Intervalos Bayesianos de credibilidade

- Distribuição beta – breve introdução
  - Usualmente utilizada para a modelagem de proporções
  - Função de densidade de probabilidade:

$$f(\theta) = B(a, b)^{-1} \theta^{a-1} (1 - \theta)^{b-1}, \quad a, b > 0, 0 < \theta < 1,$$

onde  $B(a, b)^{-1}$  é a constante de normalização que depende apenas de  $a$  e  $b$ :

$$B(a, b) = \int_0^1 \theta^{a-1} (1 - \theta)^{b-1} d\theta = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}$$

e  $\Gamma()$  é a função gama (por simplicidade, sua definição não será apresentada)

- Substituindo  $B(a, b)$  na equação de  $f(\theta)$ , tem-se:

$$f(\theta) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1 - \theta)^{b-1}$$

- Note que o suporte da distribuição é o intervalo real  $(0, 1)$

# Intervalos Bayesianos de credibilidade

- Distribuição beta – breve introdução
  - Alguns parâmetros de interesse:

- Média:

$$E(\theta) = \frac{a}{a + b}$$

- Moda:

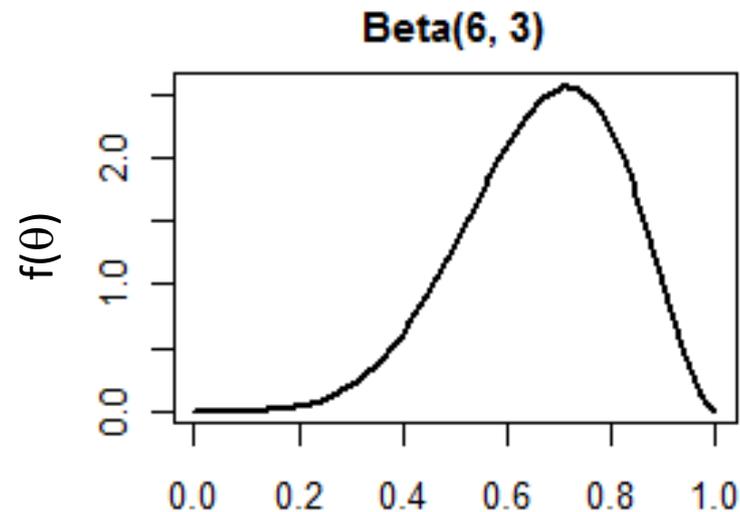
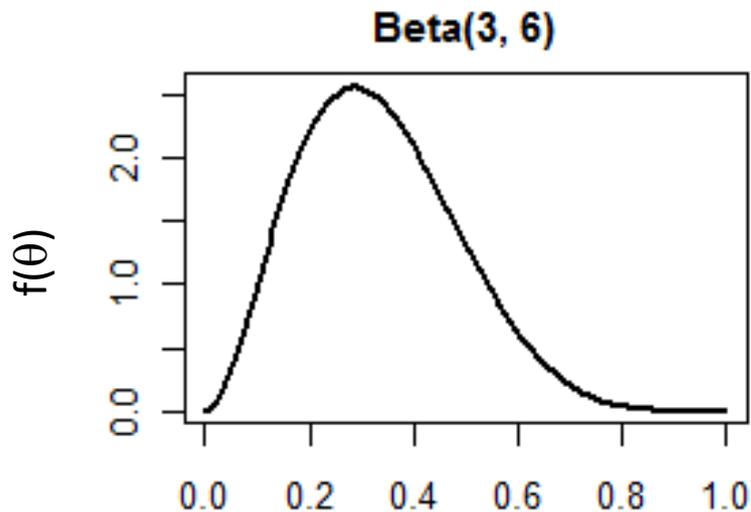
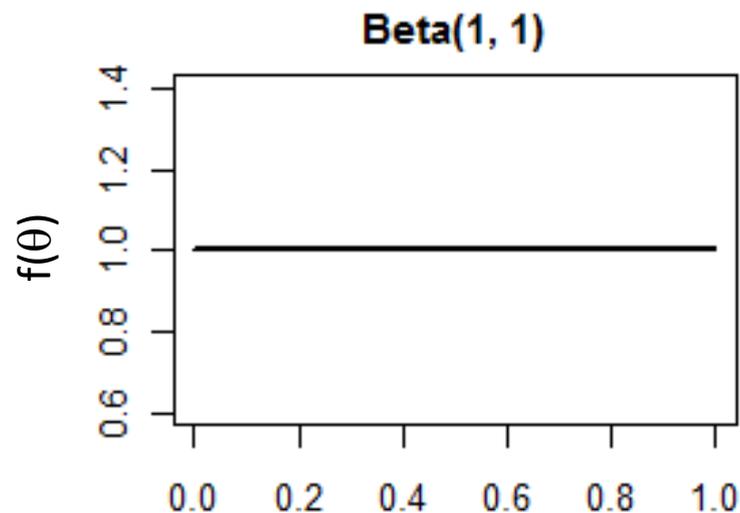
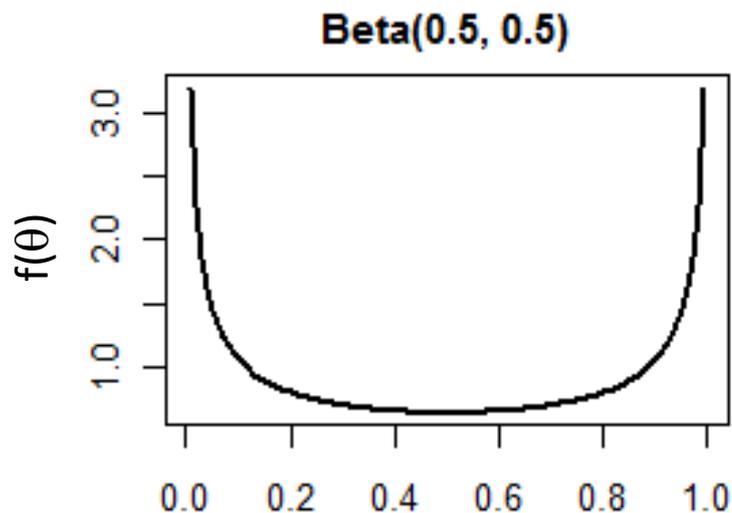
$$\theta^* = \arg \max_{\theta} f(\theta) = \frac{a - 1}{(a - 1) + (b - 1)} = \frac{a - 1}{a + b - 2}$$

Dem:

Basta tomar a derivada de  $\log f(\theta)$ , denotada por  $lf'(\theta)$ , igualar a 0 e isolar o valor de  $\theta^*$  (que corresponderá exatamente à equação acima). notar que  $\theta^*$  corresponde ao ponto no intervalo  $(0,1)$  tal que  $lf'(\theta) = 0$ .

# Intervalos Bayesianos de credibilidade

- Distribuição beta – breve introdução
  - Exemplos de distribuições beta



# Intervalos Bayesianos de credibilidade

- Prioris conjugadas – beta-binomial:
  - Exemplo – proporção de votos em um candidato:
    - Dois candidatos A e D participam do 2º turno de uma eleição
    - Um analista está interessado em inferir a respeito da proporção,  $\theta$ , de eleitores que votarão em D.
    - Por conveniência, o analista atribui a distribuição a priori  $\theta \sim \text{Beta}(a, b)$ , com  $a, b > 0$ :
$$f(\theta) = B(a, b)^{-1} \theta^{a-1} (1 - \theta)^{b-1}$$
    - O analista amostra  $n$  indivíduos da população e observa o número  $X$  de entrevistados que declaram o voto em D
    - Como a amostra é uma AAS, o analista considera que  $X|\theta \sim \text{Bin}(n, \theta)$ , onde  $\theta$ , naturalmente, é desconhecido:
$$f(X|\theta) = \binom{n}{X} \theta^X (1 - \theta)^{n-X}$$
    - Qual a distribuição a posteriori  $f(\theta|X)$ ?

# Intervalos Bayesianos de credibilidade

- Prioris conjugadas – beta-binomial:
  - Exemplo – proporção de votos em um candidato (cont):
    - Aplicando o Teorema de Bayes, temos:

$$\begin{aligned}
 f(\theta|X) &= \frac{f(\theta) f(X|\theta)}{\int_0^1 f(v) f(X|v) dv} = \frac{B^{-1} \theta^{a-1} (1-\theta)^{b-1} \binom{n}{X} \theta^X (1-\theta)^{n-X}}{\int_0^1 B^{-1} v^{a-1} (1-v)^{b-1} \binom{n}{X} v^X (1-v)^{n-X} dv} \\
 &= \frac{\theta^{a+X-1} (1-\theta)^{b+n-X-1}}{\int_0^1 v^{a+X-1} (1-v)^{b+n-X-1} dv} \quad (1)
 \end{aligned}$$

Note que, pelo mesmo raciocínio que

$$B(a, b) = \int_0^1 v^{a-1} (1-v)^{b-1} dv = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

vale que o denominador da Eq. (1) acima resulta em  $\frac{\Gamma(a+X)\Gamma(b+n-X)}{\Gamma(a+X+b+n-X)}$ ,

e portanto

$$f(\theta|X) = \frac{\Gamma(a+b+n)}{\Gamma(a+X)\Gamma(b+n-X)} \theta^{a+X-1} (1-\theta)^{b+n-X-1},$$

o que equivale à definição da distribuição  $Beta(a+X, b+n-X)$

# Intervalos Bayesianos de credibilidade

- Prioris conjugadas – beta-binomial:
  - Exemplo – proporção de votos em um candidato (cont):
    - Ou seja,  $\theta|X$  também segue uma distribuição beta, com parâmetros  $(a + x, b + n - x)$ .
    - Por essa razão, dizemos que a família de distribuições beta é uma **família de prioris conjugadas** para a família de distribuições binomiais
      - Ou simplifcadamente, dizemos que a distribuição beta é **uma priori conjugada** para a distribuição binomial
  - Definição um pouco mais geral:
    - Sempre que a função de probabilidade/função de densidade de probabilidade a posteriori  $f(\theta|X)$  pertence à mesma família  $\Psi$  da f.p./f.d.p. a priori  $f(\theta)$  (para quaisquer valores dos parâmetros assumidos para  $f(\theta)$  e quaisquer valores observados  $x_1, x_2, \dots, x_n$ ), dizemos que a família  $\Psi$  é uma *família conjugada de distribuições a priori* para amostras das distribuições  $f(X|\theta)$ .

# Intervalos Bayesianos de credibilidade

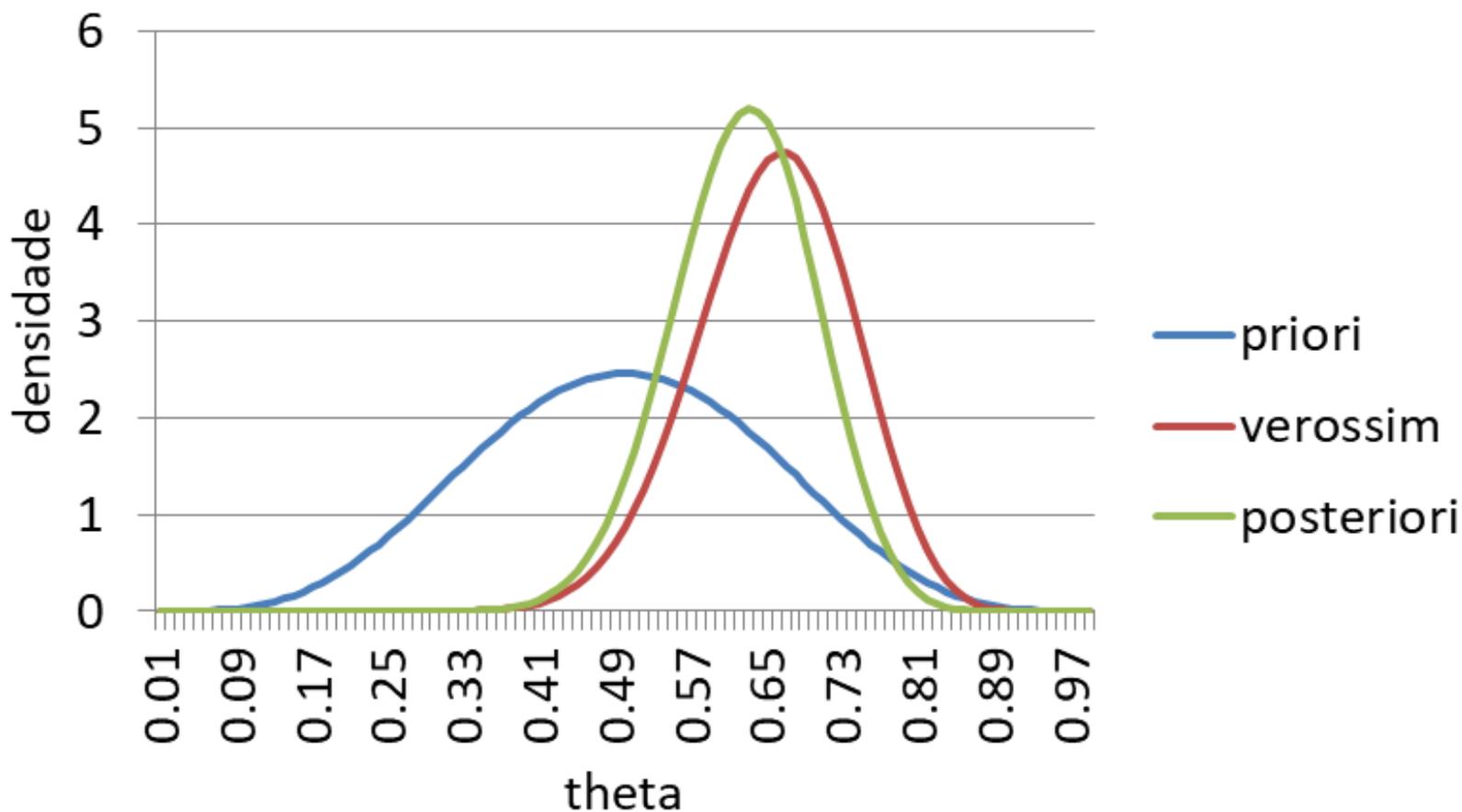
- Prioris conjugadas – beta-binomial:
  - Exercícios:
    1. Continuando com nosso exemplo da proporção de votos em um candidato, suponha que:
      - O analista assumiu inicialmente que os candidatos estão empatados e atribuiu uma distribuição a priori  $\theta \sim \text{Beta}(5,5)$
      - O analista amostrou  $n=30$  entrevistados, e o número de entrevistados que declarou o voto no candidato D foi  $x=20$
      - Qual é a distribuição a posteriori para  $\theta$ ?
    2. Implemente um script em R para calcular e plotar a distribuição a priori, a verossimilhança e a distribuição a posteriori para  $\theta$ , mantendo  $a, b, n, x$  parametrizáveis
      - Teste a priori  $\theta \sim \text{Beta}(1,1)$
      - Observe que a forma da distribuição a posteriori é similar à forma da função de verossimilhança

# Intervalos Bayesianos de credibilidade

- Prioris conjugadas – beta-binomial:

– Exercícios (cont):

Exercício 1: Priori:  $\theta \sim \text{Beta}(5,5)$ ;  $n=30$ ;  $x=20$



# Intervalos Bayesianos de credibilidade

- Prioris conjugadas – beta-binomial:

- Priori  $\theta \sim \text{Beta}(1,1)$ :

No exercício 2 do slide anterior, em que o interesse era extrair a distribuição da proporção  $\theta$  de votos em um candidato, pediu-se para verificar o que ocorre quando assume-se a priori  $\theta \sim \text{Beta}(1,1)$ . Note que:

- $\text{Beta}(1,1)$  corresponde à distribuição Uniforme(0,1), pois

$$f(\theta|a, b) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1 - \theta)^{b-1} = \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} \theta^0 (1 - \theta)^0 = 1$$

Por essa razão, a distribuição *a priori*  $\text{Beta}(1,1)$  é usualmente denominada *priori uniforme* no intervalo (0,1)

# Intervalos Bayesianos de credibilidade

- Prioris conjugadas – beta-binomial:

– Priori  $\theta \sim \text{Beta}(1,1)$  (cont):

- A distribuição a posteriori  $f(\theta|X)$  é uma distribuição beta com parâmetros  $(1 + x, 1 + n - x)$ , ou seja,

$$f(\theta|X) = K \theta^{1+x-1} (1 - \theta)^{1+n-x-1} = K \theta^x (1 - \theta)^{n-x},$$

onde  $K = \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)} = \frac{(n+1)!}{x!(n-x)!}$ ; logo,

$$f(\theta|X) = (n + 1) \frac{n!}{x!(n-x)!} \theta^x (1 - \theta)^{n-x} = (n + 1) f(X|\theta)$$

Ou seja,  $f(\theta|X)$  é proporcional a  $f(X|\theta)$  por um fator  $(n + 1)$ . Em outras palavras, as formas de  $f(\theta|X)$  são idênticas, o que significa que a priori  $\text{Beta}(1,1)$  não adiciona nenhuma informação à verossimilhança  $f(X|\theta)$  obtida a partir dos dados.

Quando uma distribuição a priori possui a propriedade  $f(\theta|X) \propto f(X|\theta)$ , dizemos que a priori  $f(\theta)$  é *não informativa*

- Assim, pode-se dizer que a distribuição a priori  $\text{Beta}(1,1)$ , que é uniforme no intervalo  $(0,1)$ , é uma priori conjugada não informativa para a distribuição binomial

# Intervalos Bayesianos de credibilidade

- Prioris conjugadas – gama-Poisson:
  - Exemplo: chegadas de clientes
    - A fim de planejar sua estrutura de atendimento, um varejista deseja modelar o parâmetro  $\theta$  que representa o número médio esperado de clientes que chegam no intervalo de 1 hora (em um dia da semana e horário)
    - A chegada de clientes é assumida como um processo de Poisson com esse parâmetro desconhecido  $\theta$
    - O varejista atribui a  $\theta$  uma distribuição a priori gama com parâmetros  $(a, b)$ , onde  $a$  é o parâmetro forma e  $b$  é o parâmetro taxa da distribuição
    - Suponha que o varejista tenha observado amostras independentes  $x_1, x_2, \dots, x_n$ , onde  $x_i$  denota o número de clientes que chegaram no  $i$ -ésimo intervalo de 1h (dentro do mesmo dia da semana e horário).
    - Problema: determinar a distribuição a posteriori de  $\theta$ , dados  $x_1, x_2, \dots, x_n$

# Intervalos Bayesianos de credibilidade

- Prioris conjugadas – gama-Poisson:

– Exemplo: chegadas de clientes (cont)

- A f.d.p. da distribuição a priori para  $\theta$ , dados os parâmetros  $a > 0, b > 0$ , é dada por

$$f(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}, \quad \theta > 0$$

- Como se assume que as observações seguem um processo de Poisson, cada observação  $x$  segue a distribuição de Poisson com parâmetro  $\theta$ :

$$f(x|\theta) = \frac{\theta^x e^{-\theta}}{x!}, \quad x \geq 0$$

- Logo, a função de verossimilhança  $f(\mathbf{x}|\theta)$ ,  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ , é:

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n \frac{\theta^{x_i} e^{-x_i}}{x_i!} = \frac{\theta^s e^{-n\theta}}{z},$$

onde  $s = \sum_{i=1}^n x_i, z = \prod_{i=1}^n (x_i!)$

# Intervalos Bayesianos de credibilidade

- Prioris conjugadas – gama-Poisson:
  - Exemplo: chegadas de clientes (cont)

- Pelo teorema de Bayes,

$$f(\theta|\mathbf{x}) = \frac{f(\theta)f(\mathbf{x}|\theta)}{\int f(v)f(\mathbf{x}|v)dv}$$

- Note que o denominador da equação acima não envolve  $\theta$ , e portanto é proporcional a  $f(\theta)f(\mathbf{x}|\theta)$   
Notação:  $f(\theta|\mathbf{x}) \propto f(\theta)f(\mathbf{x}|\theta)$

- Assim, podemos escrever a f.d.p. a posteriori  $f(\theta|\mathbf{x})$  como:

$$f(\theta|\mathbf{x}) \propto f(\theta)f(\mathbf{x}|\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \frac{\theta^s e^{-n\theta}}{z}$$

- Eliminando novamente os termos que não dependem de  $\theta$ , temos

$$f(\theta|\mathbf{x}) \propto \theta^{a-1} e^{-b\theta} \theta^s e^{-n\theta} = \theta^{a+s-1} e^{-(b+n)\theta}$$

- O lado direito da equação acima é, exceto por um fator constante, a f.d.p. de uma distribuição gama com parâmetros  $(a + \sum_{i=1}^n x_i, b + n)$

# Intervalos Bayesianos de credibilidade

- Prioris conjugadas – gama-Poisson:
  - No exemplo acima, mostramos que a distribuição gama é conjugada para a distribuição Poisson:
    - Ou seja, se  $\theta$  é assumida a priori com uma distribuição gama( $a, b$ ) e cada observação  $x|\theta$  segue uma distribuição Poisson( $\theta$ ), então  $\theta|x$  segue uma distribuição gama( $a + \sum_{i=1}^n x_i, b + n$ ).

# Intervalos Bayesianos de credibilidade

- Prioris conjugadas – gama-Poisson:
  - Que valores assumir para os parâmetros da distribuição a priori gama?
  - Se temos pouca informação prévia sobre  $\theta$ , é recomendável que  $f(\theta)$  seja “fracamente informativa”, o que é possível fazendo com que  $f(\theta)$  tenha uma variância alta
    - É possível verificar na literatura que, se  $\theta \sim \text{gama}(a, b)$ , então  $E(\theta) = \frac{a}{b}$ ,  $Var(\theta) = \frac{a}{b^2}$
    - Assim, se nossa experiência prévia nos indica que o número médio de ocorrências/hora é de 10 e se quisermos uma variância de 100 vezes a média, teríamos:
$$Var(\theta) = \frac{a}{b^2} = 100E(\theta) = 100 \frac{a}{b} \Rightarrow b = \frac{1}{100} = 0.01$$
$$E(\theta) = \frac{a}{b} = 10 \Rightarrow a = 10b = \frac{10}{100} = 0.1$$
    - Logo,  $\theta \sim \text{gama}(0.1, 0.01)$  possui a média e a variância desejadas.

# Intervalos Bayesianos de credibilidade

- Intervalo de credibilidade:
  - Um intervalo de credibilidade é uma faixa de valores dentro da qual o parâmetro  $\theta$  deve pertencer com uma certa probabilidade.
  - É calculado a partir da distribuição a posteriori
  - Dentre as formas de calcular intervalos de credibilidade, temos:
    - Intervalo por quantis da posteriori: consiste no intervalo  $[\theta_1, \theta_2]$  tal que  $\theta_1 = F^{-1}\left(\frac{1-\gamma}{2}\right)$  e  $\theta_2 = F^{-1}\left(1 - \frac{1-\gamma}{2}\right)$ , onde  $F^{-1}$  denota a função quantil da distribuição a posteriori
      - P.ex, para  $\gamma = 0.95$ ,  $\theta_1 = F^{-1}(0.025)$  e  $\theta_2 = F^{-1}(0.975)$
    - Intervalo de máxima densidade de probabilidade a posteriori (HPD – highest posterior density interval): consiste no menor intervalo possível  $[\theta_1, \theta_2]$  tal que  $\Pr(\theta_1 \leq \theta \leq \theta_2) = \gamma$ , onde  $\gamma$  é o nível de probabilidade desejado

# Intervalos Bayesianos de credibilidade

- Intervalo de credibilidade:
  - Intervalo por quantis da posteriori:
    - Procedimento mais simples e direto (não necessita de simulações)
    - consiste no intervalo  $[\theta_1, \theta_2]$  tal que  $\theta_1 = F^{-1}\left(\frac{1-\gamma}{2}\right)$  e  $\theta_2 = F^{-1}\left(1 - \frac{1-\gamma}{2}\right)$ , onde  $F^{-1}$  denota a função quantil da distribuição a posteriori
      - P.ex, para  $\gamma = 0.95$ ,  $\theta_1 = F^{-1}(0.025)$  e  $\theta_2 = F^{-1}(0.975)$
    - Intervalo de máxima densidade de probabilidade a posteriori (HPD – highest posterior density interval): consiste no menor intervalo possível  $[\theta_1, \theta_2]$  tal que  $\Pr(\theta_1 \leq \theta \leq \theta_2) = \gamma$ , onde  $\gamma$  é o nível de probabilidade desejado

# Intervalos Bayesianos de credibilidade

- Intervalo de credibilidade:
  - Intervalo de máxima densidade de probabilidade a posteriori:
    - Procedimento:
      - Gere um número grande  $B \geq 5000$  de valores de  $\theta$  de acordo com a distribuição a posteriori  $f(\theta|X)$
      - Ordene os valores de  $\theta$  em ordem decrescente de  $f(\theta|X)$ ; denotemos por  $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(B)}$  os valores ordenados em ordem decrescente
      - Identifique o valor  $\theta^{(B\gamma)}$  na lista ordenada acima; denote esse valor por  $\theta^*$ 
        - » P.ex. para  $B = 10000$ ,  $\gamma = 0.95$ , tomaríamos como  $\theta^*$  o valor correspondente à posição 9500 na lista ordenada acima
      - Defina o intervalo  $[\theta_1, \theta_2]$  tal que:
        - »  $\theta_1 = \min \theta | f(\theta|X) \geq f(\theta^*|X)$
        - »  $\theta_2 = \max \theta | f(\theta|X) \geq f(\theta^*|X)$
      - (Note que  $\theta_1, \theta_2 \leq \theta^*$ )