Inferência estatística

Aula 03

Paulo Justiniano Ribeiro Jr
Fernando Pol Mayer

Disciplina: CE-227
Inferência Bayesiana
Universidade Federal do Paraná

08 de fevereiro de 2022

Objetivo

- Inferência estatística (*what* ...?)
- Aprender com dados.
- Expressões de incerteza.
- Paradigmas de aprendizado.

Laboratório de Estatística e Geoinformação

O problema do teste de diagnóstico

Informação disponível:

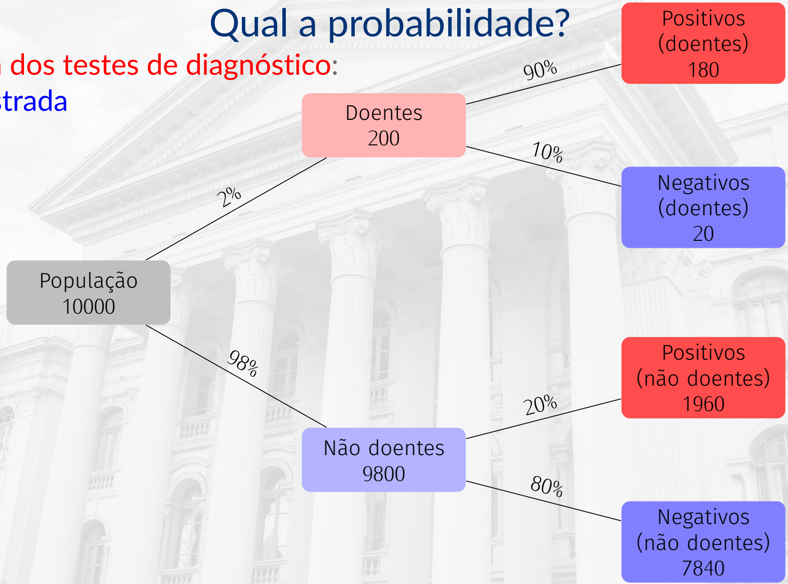
- Teste de varredura (*screening*) para uma determinada doença
- Testes são *imperfeitos*, suponha que:
  acerta 90% dos que tem doença e 80% dos que não tem.
- A doença ocorre em 2% da população

Pergunta de interesse:
Se uma pessoa testou positivo, qual a chance de ter a doença?

Qual a probabilidade?

O problema dos testes de diagnóstico:

Solução ilustrada



Qual a probabilidade? (III)

O problema dos testes de diagnóstico

Terminologia específica:

- Teste de *screening* para uma determinada doença
- Teste *imperfeito*:
  acerta 90% dos que tem doença (*sensibilidade*) e portanto 10% de *falso negativo*
  acerta 80% dos que não tem (*especifidade*) e portanto 20% de *falso positivo*
- A doença ocorre em 2% da população (*prevalência*)

Pergunta de interesse:
Se uma pessoa testou positivo, qual a chance de ter a doença?
(*valor preditivo positivo*)

Laboratório de Estatística e Geoinformação

Testes de diagnóstico - Organizando em tabelas

Características do teste (para diagnósticos conhecidos)

	Positivo	Negativo	
c/ Doença	0,90	0,10	1
s/ Doença	0,20	0,80	1

Temos 2% com a doença na população (e portanto 98% sem)

	Positivo	Negativo	Total
c/ Doença	0,018	0,002	0,02
s/ Doença	0,196	0,784	0,98
Total	0,214	0,786	1

Conhecendo o resultado do teste

	Positivo	Negativo
c/ Doença	0.0841	0.00254
s/ Doença	0.916	0.997
Total	1	1

Testes de diagnóstico

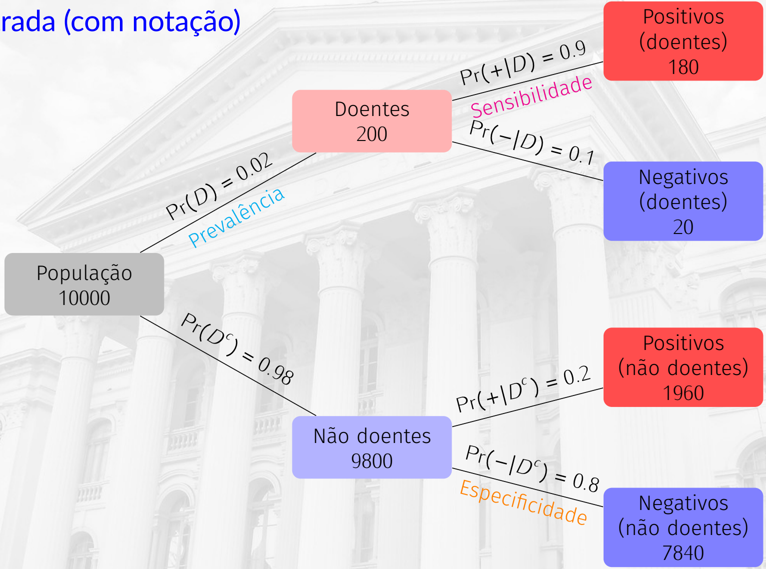
A notação é nossa amiga!

$$
\begin{aligned}
P[+|D] &= 0,90 \longrightarrow P[-|D] = 0,10 \\
P[-|\overline{D}] &= 0,80 \longrightarrow P[+|\overline{D}] = 0,20 \\
P[D] &= 0,02 \\
P[D|+] &= ?
\end{aligned}
$$

E um Teorema resolve o problema!

$$
\begin{aligned}
P[D|+] &= \frac{P[+|D] \cdot P[D]}{P[+|D] \cdot P[D] + P[+|\overline{D}] \cdot P[\overline{D}]} \\
&= \frac{0.90 \cdot 0.02}{0.90 \cdot 0.02 + 0.20 \cdot 0.98} = 0.0841
\end{aligned}
$$

Solução ilustrada (com notação)



Qual a probabilidade?

O problema dos testes de diagnóstico

E se o teste tivesse sido feito por recomendação médica após um exame? Mudaria algo?

Baseado em sua experiência o médico estima que 30% dos pacientes com os sintomas apresentados tem a doença.

Solução? (reproduzir passos acima!)

Neste caso:

$$
\begin{aligned}
P[D|+] &= \frac{P[+|D] \cdot P[D]}{P[+|D] \cdot P[D] + P[+|\overline{D}] \cdot P[\overline{D}]} \\
&= \frac{0.90 \cdot 0.30}{0.90 \cdot 0.30 + 0.20 \cdot 0.70} = 0.659
\end{aligned}
$$

Comparação e *interpretação* (e ... muito cuidado!)

Laboratório de Estatística e Geoinformação

Teorema de Bayes

$$\text{Se eventos } A_i's \text{ são partição de } \Omega, \text{ então } P[A_j|B] = \frac{P[B|A_j] \cdot P[A_j]}{\sum_j P[B|A_j] \cdot P[A_j]}$$

Revisitando o problema do teste de diagnóstico (e tb ex 9):

$$A \text{ :estado do paciente} : A_1(D) : \text{com a doença}, A_2(\overline{D}) : \text{sem a doença}$$

$$B \text{ :resultado do teste} : B(+) : \text{positivo}, \overline{B}(-) : \text{negativo}$$

Dados:

$$P[D] = P[A_1] = 0,02 \qquad P[\overline{D}] = P[A_2] = 0,98$$

$$P[+|D] = P[B|A_1] = 0,90 \qquad P[-|D] = P[\overline{B}|A_1] = 0,10$$

$$P[-|\overline{D}] = P[\overline{B}|A_2] = 0,80 \qquad P[+|\overline{D}] = P[B|A_2] = 0,20$$

Teorema de Bayes:

$$P[D|+] = \frac{P[+|D] \cdot P[D]}{P[+]} = \frac{P[+|D] \cdot P[D]}{P[+|D] \cdot P[D] + P[+|\overline{D}] \cdot P[\overline{D}]}$$

$$= \frac{0,90 \cdot 0,02}{0,90 \cdot 0.02 + 0,20 \cdot 0,98} = \color{red}{0.0841}$$

Teorema de Bayes

Reescrevendo e reinterpretando como problema de classificação

$$P[A_j|B] = \frac{P[B|A_j] \cdot P[A_j]}{\sum_j P[B|A_j] \cdot P[A_j]} \propto P[B|A_j] \cdot P[A_j]$$

Como o paciente deve ser classificado após o teste?

$$P[A_1|B] \propto P[B|A_1] \cdot P[A_1](\text{ou}, P[D|+] \propto P[+|D] \cdot P[D])$$
$$P[A_2|B] \propto P[B|A_2] \cdot P[A_2](\text{ou}, P[\overline{D}|+] \propto P[+|\overline{D}] \cdot P[\overline{D}])$$

Portanto

$$P[A_1|B] \propto 0,9 \cdot 0,02 = 0,018$$
$$P[A_2|B] \propto 0,2 \cdot 0,98 = 0,196$$

A_1 e A_2 são todas as categorias possíveis, as probabilidades devem somar 1:

$$P[A_1|B] = \frac{0,018}{0,018 + 0,196} = 0,084$$
$$P[A_2|B] = \frac{0,196}{0,018 + 0,196} = 0,916$$

Teorema de Bayes

E se o teste for repetido?
Notaçao: B_1 positivo no primeiro teste e B_2 positivo no segundo teste
Supondo independência

$$P[A_1|B_1, B_2] \propto P[B_2|A_1] \cdot P[B_1|A_1] \cdot P[A_1]$$
$$P[A_2|B_1, B_2] \propto P[B_2|A_2] \cdot P[B_1|A_2] \cdot P[A_2]$$

Portanto se for o mesmo teste (mesmas características de sensibilidade e espeficidade)

$$P[A_1|B] \propto 0,9^2 \cdot 0,02 = 0,0162$$
$$P[A_2|B] \propto 0,2^2 \cdot 0,98 = 0,0392$$

Logo,

$$P[A_1|B] = \frac{0,0162}{0,0162 + 0,0392} = 0,292 \quad P[A_2|B] = \frac{0,0392}{0,0162 + 0,0392} = 0,708$$

A classificação ainda é a mesma mas as chances mudaram!
Com três testes positivos a classificação mudaria.

Teorema de Bayes

Voltando a um teste ... mas ... não está estranho?
A probabilidade de doença $P = 0,084$ não está muito baixa?
Qual a população?

- teste de varredura (*screening*) $P(A_1) = 0,02$
- teste por indicação (*auxílio a diagnóstico*) $P(A_1) \gg 0,02$

Consulta levanta suspeita e pede-se o teste.
Opinião especializada após consulta $P(A_1) = 0,50$

$$P[A_1|B] \propto 0,9 \cdot 0,50 = 0.45$$
$$P[A_2|B] \propto 0,2 \cdot 0,98 = 0.196$$

Padronizando para somar 1:

$$P[A_1|B] = \frac{0.45}{0.45 + 0.1} = 0.818$$
$$P[A_2|B] = \frac{0.1}{0.45 + 0.1} = 0.182$$

Laboratório de Estatística e Geoinformação

Testes de diagnóstico - Organizando em tabelas

PRIORI: Temos 2% com a doença na população (e portanto 98% sem)

	Positivo	Negativo	Total
c/ Doença	0,018	0,002	0,02
s/ Doença	0,196	0,784	0,98
Total	0,214	0,786	1

VEROSSIMILHANÇA: Características do teste (para diagnósticos conhecidos)

	Positivo	Negativo	
c/ Doença	0,90	0,10	1
s/ Doença	0,20	0,80	1

POSTERIORI: Conhecendo o resultado do teste

	Positivo	Negativo
c/ Doença	0.0841	0.00254
s/ Doença	0.916	0.997
Total	1	1

Estimando uma proporção

Em uma população (considerada *infinita*) uma proporção θ de indivíduos apresenta determinada característica.
Deseja-se (inferências):

- estimar θ,
- expressar a incerteza sobre esta estimativa,
- verificar se θ (e portanto a população) está fora de normas/referências (proporção max. de 20%), se há evidências de um desvio "relevante" (significativo).

Dados de *uma* amostra (considerada aleatória):
$n = 80$ e $y = 19$
Como proceder?

Paradigmas e métodos de inferência

Objetivos:
Estimativa de θ,
expressão da incerteza,
opinião em relação a valor de interesse $\theta_0 = 0.20$

Abordagens (paradigmas):

- frequentista,
- verossimilhança,
- bayesiana.

Abordagem Frequentista

- Baseia-se em considerar o comportamento das quantidades de interesse medidas na amostra, supondo que diversas amostras fossem tomadas da população.
- Tais quantidades, por serem baseadas em amostras aleatórias são portanto aleatórias, e possuem alguma distribuição de probabilidades.
- Tal distribuição de probabilidades é chamada de `distribuição amostral`.
- As inferências frequentistas são baseadas em probabilidades medidas nestas distribuições.
- Usual nos métodos, técnicas e procedimentos de estatística, especialmente os ligados a cursos e textos básicos e aplicados a diversas áreas.
- As distribuições amostrais podem ser obtidas analiticamente em alguns casos (e.g teste-t), aproximadas por distribuições conhecidas, ou obtidas por procedimentos computacionais intensivos (e.g. testes aleatorizados e bootstrap).

Abordagem frequentista

Inferência se baseia na distribuição amostral

$$\text{Estimativa} : \hat{\theta} = \frac{Y}{n}$$

$$\hat{\theta} \sim \mathrm{N}(\mu = \theta, \sigma^2 = \frac{\theta(1-\theta)}{n})(\text{distribuição amostral})$$

$$\text{IC} : \hat{\theta} \pm z_{1-\alpha/2}\sqrt{\frac{\theta(1-\theta)}{n}}$$

$$\text{TH} : (H_1 : \theta > \theta_0) : \hat{\theta} \sim \mathrm{N}(\theta_0, \frac{\theta_0(1-\theta_0)}{n})$$

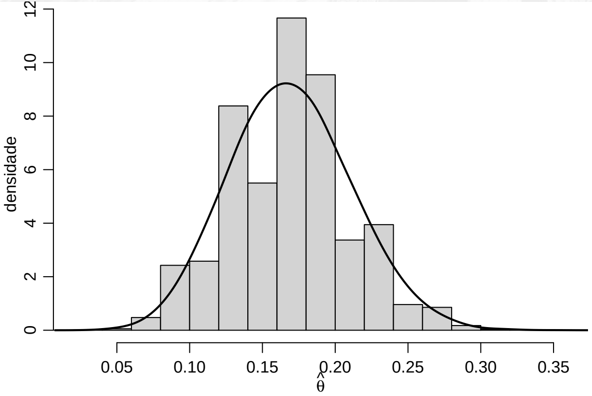
$$\text{equivalentemente } z = \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} \sim \mathrm{N}(0, 1)$$

Usa-se $\theta = \hat{\theta}$ (assintótico) ou $\theta = 0,5$ (conservador)

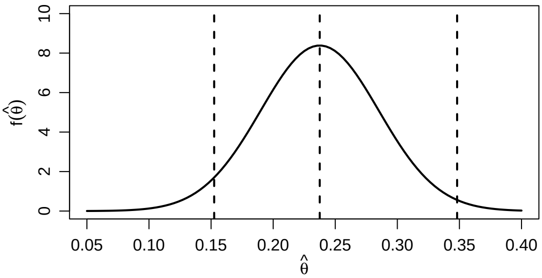
Simulação da distribuição amostral

(código completo em arquivo inf-prop.R) As estimativas variam se tomamos diversas amostras da população:

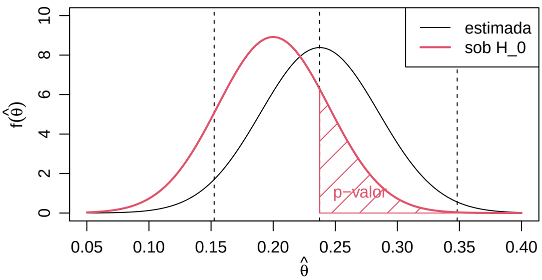
```
summary(ps)
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##   0.0375   0.1375   0.1625   0.1693   0.2000   0.3500
```



Distribuição amostral (estimada)



Distribuição amostral



Inferência para proporção - frequentista

Na prática, com recursos computacionais

```
prop.test(19, 80)$conf
## [1] 0.1524765 0.3481396
## attr(,"conf.level")
## [1] 0.95
prop.test(19, 80, p=0.20, alt="greater")
##
##  1-sample proportions test with continuity correction
##
## data:  19 out of 80, null probability 0.2
## X-squared = 0.48828, df = 1, p-value = 0.2423
## alternative hypothesis: true p is greater than 0.2
## 95 percent confidence interval:
##  0.1632771 1.0000000
## sample estimates:
##      p
## 0.2375
```

Inferência para proporção - frequentista

Resumindo:

- Se baseia no comportamento das *possíveis amostras* que poderiam ser retiradas da população

- Interpretação de intervalo de confiança: *o calculado a partir da amostra é um entre os possíveis, sendo que uma proporção dos possíveis (nível de confiança) conteria o verdadeiro valor*

- Interpretação do Teste de Hipótese e p-valor: *mesmo sob H_0 uma proporção das possíveis amostras produziria valores tão ou mais extremos que o visto na amostra. Se esta proporção (p-valor) é baixa (nível de significância) a amostra é considerada incompatível com a hipótese nula e rejeita-se a hipótese nula.*

Laboratório de Estatística e Geoinformação

Uma alternativa (ainda) frequentista: Teste aleatorizado

Ideia básica:
Reproduzir a essência da ideia frequentista porém obtendo a distribuição amostral por simulação sob H_0

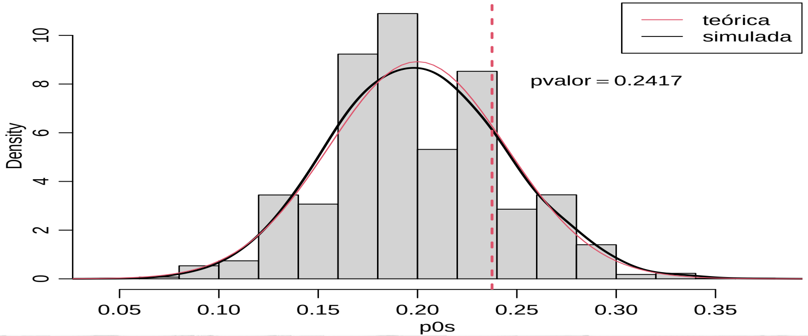
Algorítmo:

- Simular amostras da população sob H_0
- Calcular o valor de interesse ou estatística de teste para cada amostra simulada
- valor-p proporção destes que são mais “extremos” do que o valor observado na amostra

Laboratório de Estatística e Geoinformação

Teste aleatorizado

```
summary(p0s)
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.0500  0.1750  0.2000  0.2005  0.2250  0.3750
(pvalor <- mean(p0s >= 19/80))
## [1] 0.2417
```



Abordagem pela verossimilhança

Inferência é baseada nas características da **função de verossimilhança**

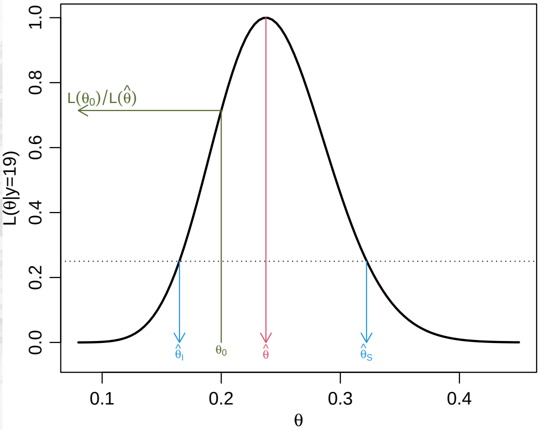


Figura 1. Uma função que expressa a informação dos dados.

Representação alternativa

Função deviance e cortes que definem intervalos (a 90, 95 e 99%)

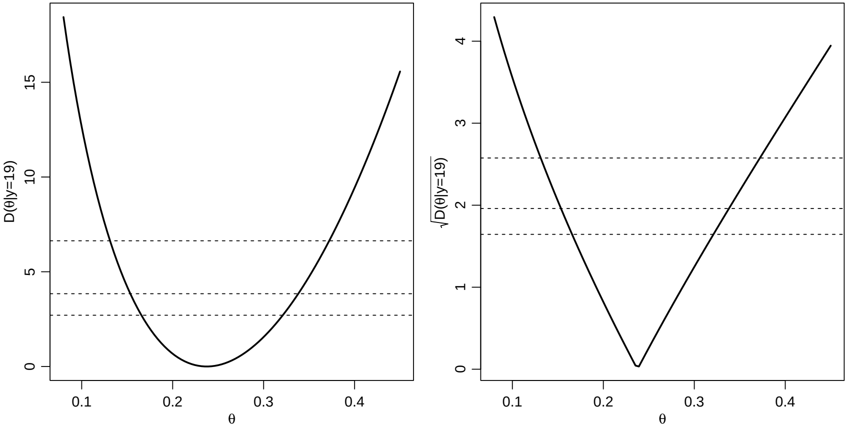


Figura 2. Representações alternativas da função de verossimilhança.

Inferência pela verossimilhança

Necessidade de critérios:

- definir o valor para corte da função para obter intervalos de confiança (IC's) ?
- definir limiar para o valor de verossimilhança (relativa ao máximo) para θ_0?

Possíveis soluções:

- critérios de razoabilidade e comparação (e.g. moedas ou a família do Sr. João!)
- argumento frequentista (comportamento "médio" da verossimilhança) estabelece relações:

r	"Caras"	$P[\|Z\| < \sqrt{c^*}]$
50%	1,00	0,761
26%	1,94	0,899
15%	2,74	0,942
3,6%	4,80	0,990

Laboratório de Estatística e Geoinformação

Comparando

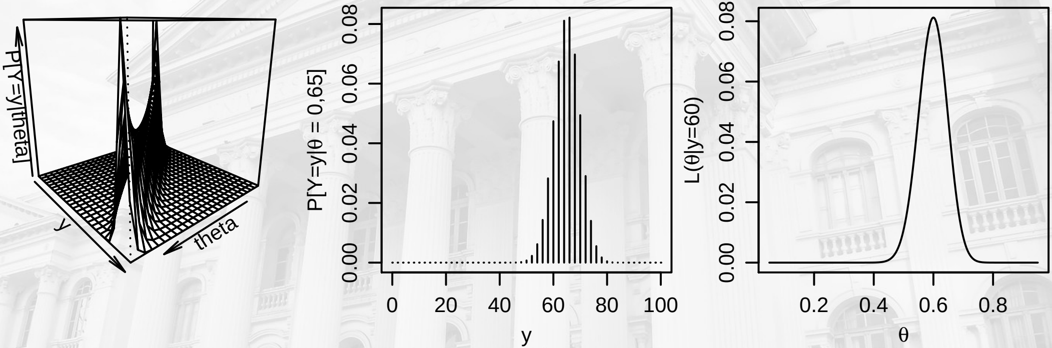
Inferência frequentista

- Estimativa de θ: fornecido por algum método de estimação
- expressão da incerteza: variabilidade da distribuição amostral
- opinião em relação a valor de interesse $\theta_0 = 0,20$: probabilidade na distribuição amostral

Inferência pela verossimilhança

- Estimativa de θ: máximo (supremo) da função
- expressão da incerteza: faixa de valores dentro de um limite de compatibilidade com a amostra, curvatura da função
- opinião em relação a valor de interesse $\theta_0 = 0.20$: comparação da verossimilhança deste valor com a do máximo

O mundo estatístico



Usando informação disponível

Um vídeo vale mais que mil palavras!
Vamos usar na discussão apenas de 0 e 50 segundos do vídeo.
(baseado em (STONE, 2013))
Por que ocorre o mal entendido?
Elementos:

- O pedido é a informação que o vendedor recebe.
- O vendedor tinha alguma opinião anterior sobre o que poderia ter sido pedido?
- O vendedor ao final acha mais provável que o cliente tenha pedido quatro velas (*four candles*) do que o cabo do garfo (*fork handles*).

Um vídeo em notação

- "Adivinhar" o que o cliente quer (estado da natureza):

$$\theta_c \text{ vela ou } \quad \theta_h \text{ cabo.}$$

- Pela experiência o vendedor sabe se vende mais velas ou cabos, tem ideia da chance de alguém comprar um ou outro:

$$P[\theta_c] \text{ vela ou } \quad P[\theta_h] \text{ cabo.}$$

- Informação (dado) y é a fala do comprador e esta fala pode ocorrer para cada possível estado da natureza.

$$P[Y|\theta_c] \text{ vela ou } \quad P[Y|\theta_h] \text{ cabo.}$$

- O vendedor ao final acha mais provável que o cliente tenha pedido quatro velas (*four candles*) do que o cabo do garfo (*fork handles*), ou seja, ele avalia:

$$P[\theta_c|y] \text{ vela ou } \quad P[\theta_h|y] \text{ cabo.}$$

Abordagem Bayesiana

O objeto de inferência é a distribuição à posteriori

- A incerteza inicial sobre θ é expressa na forma de uma distribuição priori para θ
- Com amostra atualizamos opinião θ com a informação contida na verossimilhança
- O conhecimento/incerteza atualizados sobre θ é expresso pela distribuição posteriori

Formalmente:

$$f(\theta|y) = \frac{f(y|\theta) \cdot f(\theta)}{\int f(y|\theta) \cdot f(\theta) \mathrm{d}\theta} \propto f(\theta) \cdot L(\theta|y)$$

ou, usando jargão técnico:

$$\textit{posteriori} \propto \textit{priori} \cdot \textit{verossimilhança}$$

Abordagem Bayesiana

No contexto do exemplo de estimação de proporção θ:

- Extensão da definição do modelo:

$$
\begin{aligned}
[Y|\theta] &\sim B(n, \theta) \\
[\theta] &\sim Pr(a, b) \quad (priori)
\end{aligned}
$$

- permite obter (via teorema de Bayes)

$$
\begin{aligned}
[\theta|y] &\propto [Y|\theta][\theta] \\
[\theta|y] &\sim \pi(a^*, b^*) \quad (posteriori)
\end{aligned}
$$

- Analogias diretas para estimação (pontual e intervalar),
- …mas não diretas ou triviais para testes de hipótese (valores na *posteriori* sem analogias diretas com razão de verossimilhanças).

Laboratório de Estatística e Geoinformação

Abordagem Bayesiana

No contexto do exemplo de estimação de proporção θ:

- Priori: $[\theta] \sim \mathrm{Beta}(a, b)$ (distribuição Beta)

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

- Verossimilhança: $[Y|\theta] \sim \mathrm{Bin}(n, \theta)$ (em θ, é proporcional à distribuição Beta)

$$L[\theta|y] \equiv f(y|\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}$$

- Posteriori:

$$f(\theta|y) \propto f(\theta) \cdot L(\theta|y) \propto \theta^{y+a-1}(1-\theta)^{n-y+b-1}$$

Logo

$$[\theta|y] \sim \mathrm{Beta}(a+y, n-y+b)$$

(distribuição Beta - conjugada)

A essência de Bayes ilustrada (I)

Exemplo I : estimação da proporção de atributo (θ) na população

Priori: Acredita-se que o atributo ocorre em 40% da população com 70% de chance de estar entre 30 e 50%.
Informação expressa como distribuição de probabilidades para θ:

$$[\theta] \sim Beta(9.9, 15)$$

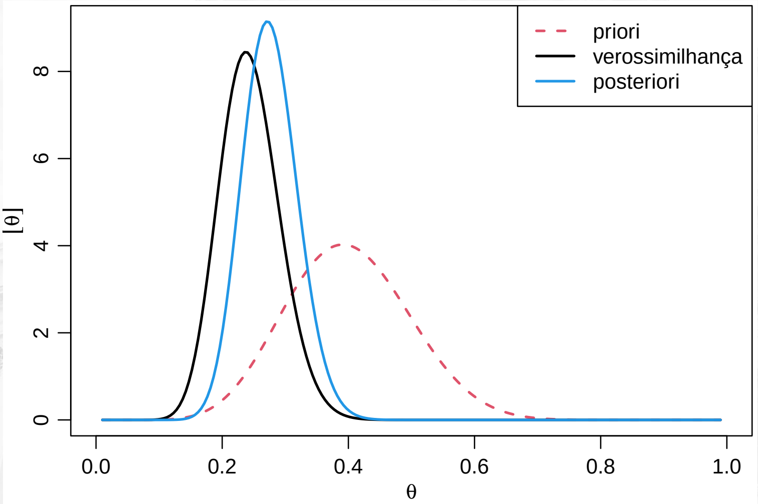
Verossimilhança: Modelo Binomial, amostra n=80 e y=19

$$[y|\theta] = \binom{80}{19}\theta^{19}(1-\theta)^{80-19}$$

Posteriori: a distribuição de probabilidades para θ após observar os dados:

$$[\theta|y] \sim Beta(29, 76)$$

A essência de Bayes ilustrada (I)



A essência de Bayes ilustrada (I)

```
(prI <- prioriBeta(0.4, c(0.30, 0.50), 0.70))
##     alpha      beta
##  9.912277 14.868415
postBinom(19, 80, prI, plot=FALSE)
## $pars
##                alpha     beta
## priori      9.912277 14.86842
## posteriori 28.912277 75.86842
##
## $summary
##                 moda     media   variancia
## priori     0.3912206 0.4000000 0.009309292
## posteriori 0.2715712 0.2759313 0.001888750
##
## $EMV
## [1] 0.2375
```

A essência de Bayes ilustrada (II)

Uma priori bem diferente:

- **Priori**: Acredita-se que o atributo ocorre em 8% da população com 90% de chance de estar entre 3 e 20%.

$$[\theta] \sim Beta(2.1, 24)$$

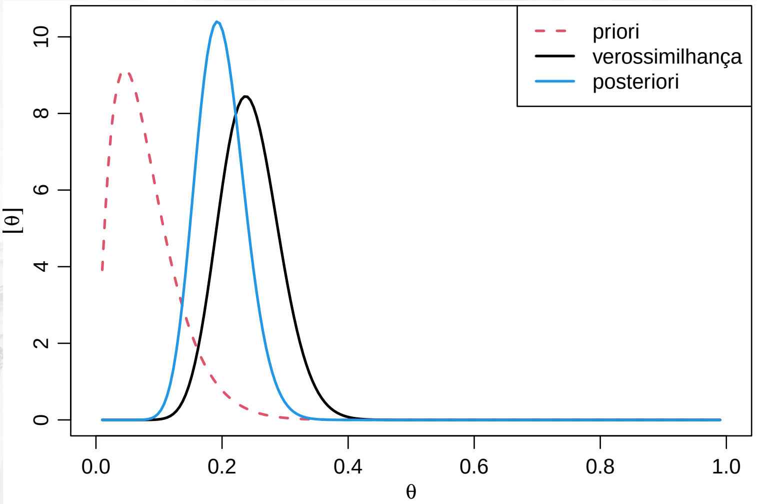
- **Verossimilhança**: Modelo Binomial, amostra n=80 e y=19

$$[y|\theta] = \binom{80}{19} \theta^{19} (1-\theta)^{80-19}$$

- **Posteriori**: após observar os dados:

$$[\theta|y] \sim Beta(21, 85)$$

A essência de Bayes ilustrada (II)



A essência de Bayes ilustrada (II)

```
(prII <- prioriBeta(0.08, c(0.03, 0.20), 0.90))
##     alpha      beta
##  2.124901 24.436358
postBinom(19, 80, prII, plot=FALSE)
## $pars
##                alpha     beta
## priori      2.124901 24.43636
## posteriori 21.124901 85.43636
##
## $summary
##                 moda     media   variancia
## priori     0.0457998 0.0800000 0.002670415
## posteriori 0.1924700 0.1982418 0.001477688
##
## $EMV
## [1] 0.2375
```

A essência de Bayes ilustrada (III)

Uma priori vaga :

- **Priori**: Não se sabe praticamente nada sobre θ. Expressa-se então que o atributo ocorre em 50% da população mas com 90% de chance de estar entre 5 e 95%.

$$[\theta] \sim Beta(1.2, 1.2)$$

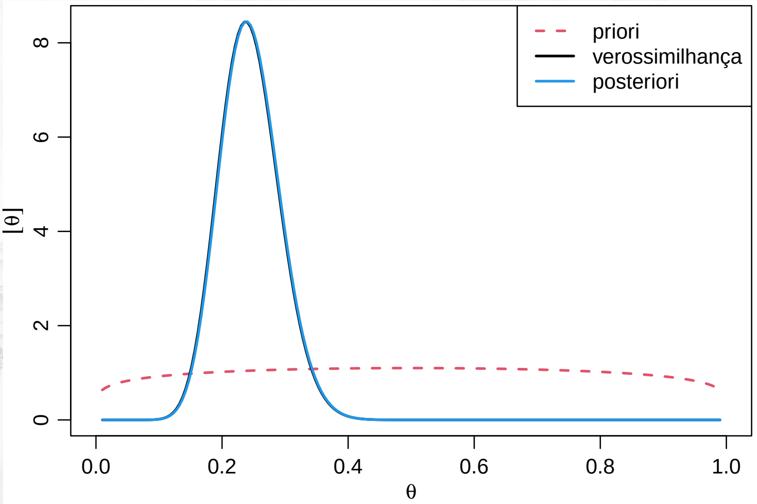
- **Verossimilhança**: Modelo Binomial, amostra n=80 e y=19

$$[y|\theta] = \binom{80}{19} \theta^{19}(1-\theta)^{80-19}$$

- **Posteriori**: após observar os dados:

$$[\theta|y] \sim Beta(20, 62)$$

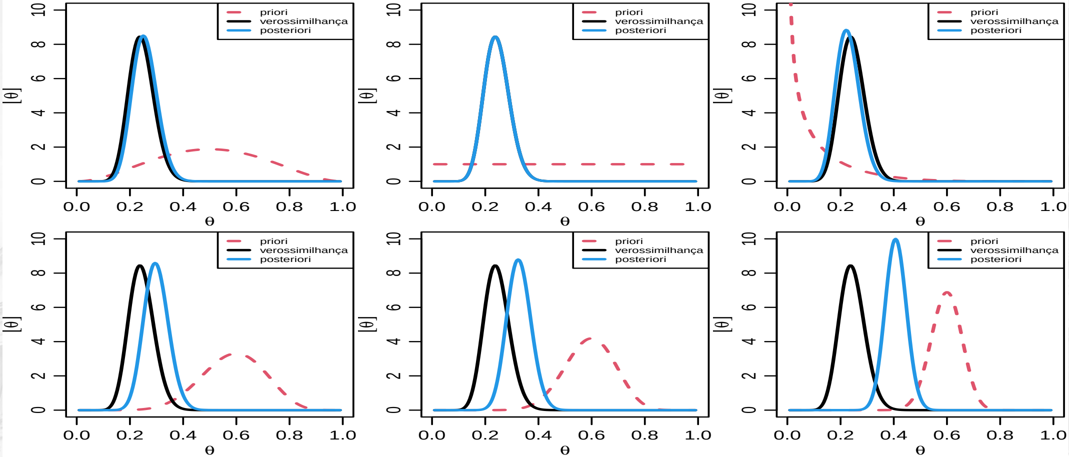
A essência de Bayes ilustrada



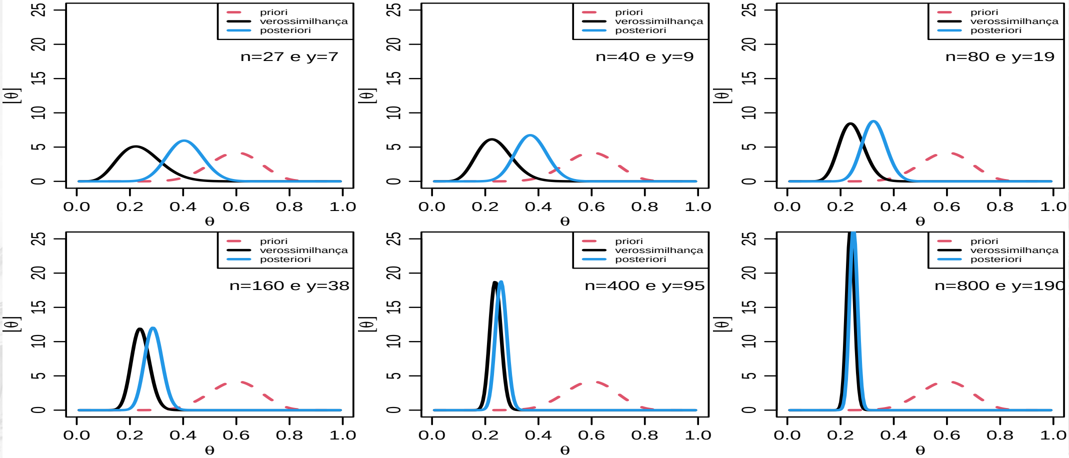
A essência de Bayes ilustrada (III)

```
(prIII <- prioriBeta(0.50, c(0.05, 0.95), 0.90))
##    alpha     beta
## 1.170088 1.170088
postBinom(19, 80, prIII, plot=FALSE)
## $pars
##                alpha      beta
## priori      1.170088  1.170088
## posteriori 20.170088 62.170088
##
## $summary
##                 moda     media   variancia
## priori     0.5000000 0.5000000 0.074846333
## posteriori 0.2386115 0.2449605 0.002219276
##
## $EMV
## [1] 0.2375
```

Efeito da priori (fixando amostra)



Efeito do tamanho da amostra (fixando priori)



Comentários

- Expressão da opinião "a priori" é necessária e sua especificação é um desafio,
- as interpretações de intervalo de confiança são agora probabilísticas, por exemplo pode-se falar em:

$$P[a < \theta < b] = 0.95$$

- bem como, no contexto do exemplo, pode-se falar em

$$P[\theta \geq 0, 20]$$

Comparando paradigmas

Qual o valor de θ? (estimação pontual):

- Frequentista: fornecido por algum método de estimação
- Verossimilhança: máximo (supremo) da função de verossimilhança
- Bayesiana: alguma medida resumo da posteriori (média, moda, mediana, …

Expressão da incerteza sobre θ (estimação intervalar):

- Frequentista: variabilidade na distribuição amostral (intervalo de confiança)
- Verossimilhança: faixa de valores dentro de um limite de compatibilidade com a amostra, curvatura da função
- Bayesiana: variabilidade na distribuição posteriori (intervalo de credibilidade)

Comparando paradigmas

Opinião em relação a valor de interesse $\theta_0 = 0,20$ (teste de hipótese):

- Frequentista: probabilidade na distribuição amostral (p-valor)
- Verossimilhança: comparação da verossimilhança deste valor com a do máximo
- Bayesiana: probabilidade na posteriori

Referência bibliográfica

STONE, J. V. **Bayes' rule: a tutorial introduction to Bayesian analysis**. [S.l.: s.n.], 2013.