

Inferência Bayesiana

Introdução

Contextualização

Paulo Justiniano Ribeiro Jr

Curso de Graduação em Estatística
Universidade Federal do Paraná

2o semestre de 2019

Inferência

Inferência

► inferência?

Inferência

- ▶ inferência?
- ▶ inferência estatística?

Inferência

- ▶ inferência?
- ▶ inferência estatística?
- ▶ aprender com os dados!

Inferência

- ▶ inferência?
- ▶ inferência estatística?
- ▶ aprender com os dados!
- ▶ como assim ...aprender com os dados!

Inferência

- ▶ inferência?
- ▶ inferência estatística?
- ▶ aprender com os dados!
- ▶ como assim ...aprender com os dados!
Vamos usar alguma notação:

y : dados θ : "quantidades" desconhecidas

Inferência

- ▶ inferência?
- ▶ inferência estatística?
- ▶ aprender com os dados!
- ▶ como assim ...aprender com os dados!
Vamos usar alguma notação:

y : dados θ : "quantidades" desconhecidas

- ▶ inferência é portanto falar sobre $\theta|y$

Inferência

- ▶ inferência?
- ▶ inferência estatística?
- ▶ aprender com os dados!
- ▶ como assim ...aprender com os dados!
Vamos usar alguma notação:

y : dados θ : "quantidades" desconhecidas

- ▶ inferência é portanto falar sobre $\theta|y$
- ▶ mas ...como aprendemos (sobre θ) com os dados (y)?

Inferência

- ▶ inferência?
- ▶ inferência estatística?
- ▶ aprender com os dados!
- ▶ como assim ...aprender com os dados!
Vamos usar alguma notação:

y : dados θ : "quantidades" desconhecidas

- ▶ inferência é portanto falar sobre $\theta|y$
- ▶ mas ...como aprendemos (sobre θ) com os dados (y)?
- ▶ mas ...**só** aprendemos (sobre θ) com os dados (y)?

O que usamos para aprender/decidir?

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).

O que usamos para aprender/decidir?

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?

O que usamos para aprender/decidir?

Um vídeo vale mais que mil palavras!

Vamos usar na discussão apenas de 0 e 50 segundos do vídeo.
(baseado em (??))

O que usamos para aprender/decidir?

Um vídeo vale mais que mil palavras!

Vamos usar na discussão apenas de 0 e 50 segundos do vídeo.
(baseado em (??))

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**):

θ_c vela ou θ_h cabo.

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

$P[\theta_c]$ vela ou $P[\theta_h]$ cabo.

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

$P[Y|\theta_c]$ vela ou $P[Y|\theta_h]$ 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:

$P[\theta_c|y]$ vela ou $P[\theta_h|y]$ cabo.

Revisando o Teorema de Bayes

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

Ex: O problema dos testes de diagnóstico

Teste de *screening* para uma determinada doença.

Teste *imperfeito*: acerta 90% dos que tem doença e 80% dos que não tem.

Sabe-se de antemão que a doença ocorre em 2% da população.

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

Testes de diagnóstico

A notação "usual"!

$$P[+|D] = 0,90 \longrightarrow P[-|D] = 0,10$$

$$P[-|\bar{D}] = 0,80 \longrightarrow P[+|\bar{D}] = 0,20$$

$$P[D] = 0,02$$

$$P[D|+] = ?$$

Testes de diagnóstico

A notação "usual"!

$$P[+|D] = 0,90 \longrightarrow P[-|D] = 0,10$$

$$P[-|\bar{D}] = 0,80 \longrightarrow P[+|\bar{D}] = 0,20$$

$$P[D] = 0,02$$

$$P[D|+] = ?$$

$$P[D|+] = \frac{P[+|D] \cdot P[D]}{P[+|D] \cdot P[D] + P[+|\bar{D}] \cdot P[\bar{D}]} = 0.0841$$

Teorema de Bayes

Alterando notação:

$$P[\theta_j|y] = \frac{P[Y|\theta_j] \cdot P[\theta_j]}{\sum_j P[Y|\theta_j] \cdot P[\theta_j]}$$

Teorema de Bayes

Alterando notação:

$$P[\theta_j|y] = \frac{P[Y|\theta_j] \cdot P[\theta_j]}{\sum_j P[Y|\theta_j] \cdot P[\theta_j]}$$

θ :estado do paciente : θ_1 : com a doença, θ_2 : sem a doença

Y :resultado do teste : y_1 : positivo, y_2 : negativo

Dados:

$$P[D] = P[\theta_1] = 0,02$$

$$P[\bar{D}] = P[\theta_2] = 0,98$$

$$P[+|D] = P[Y_1|\theta_1] = 0,90$$

$$P[-|D] = P[Y_2|\theta_1] = 0,10$$

$$P[-|\bar{D}] = P[Y_2|\theta_2] = 0,80$$

$$P[+|\bar{D}] = P[Y_1|\theta_2] = 0,20$$

Teorema de Bayes:

$$\begin{aligned} P[\theta_1|y_1] &= \frac{P[Y_1|\theta_1] \cdot P[\theta_1]}{P[Y_1]} = \frac{P[Y_1|\theta_1] \cdot P[\theta_1]}{P[Y_1|\theta_1] \cdot P[\theta_1] + P[Y_1|\theta_2] \cdot P[\theta_2]} \\ &= \frac{0,90 \cdot 0,02}{0,90 \cdot 0,02 + 0,20 \cdot 0,98} = 0,0841 \end{aligned}$$

Teorema de Bayes

$$P[\theta_j|y] = \frac{P[Y|\theta_j] \cdot P[\theta_j]}{\sum_j P[Y|\theta_j] \cdot P[\theta_j]}$$

Teorema de Bayes

$$P[\theta_j|y] = \frac{P[Y|\theta_j] \cdot P[\theta_j]}{\sum_j P[Y|\theta_j] \cdot P[\theta_j]}$$

No exemplo só haviam dois possíveis **estados da natureza**:

θ_1 : com a doença, θ_2 : sem a doença

O resultado é mais geral, para várias categorias.

Aplicação em **problemas de classificação**.

Estados da natureza não precisam ser apenas categóricos (ou ainda discretos/enumeráveis)

Os **estados da natureza** podem ocorrer em um domínio contínuo.

O Teorema de Bayes pode então ser reescrito como:

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

Abordagem Frequentista

- ▶ Baseia-se em considerar o comportamento das quantidades de interesse medidas na amostra, supondo que diversas amostras fossem tomadas da população.

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.

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.

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.

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.

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**

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

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

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

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

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

Abordagem frequentista

Inferência se baseia na **distribuição amostral**

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

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

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

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

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

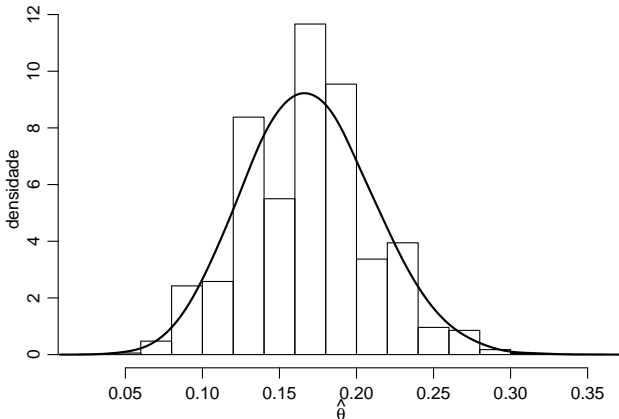
Usa-se $\theta = \hat{\theta}$ (assintótico) ou $\theta = 0,5$ (conservador)

Simulação da distribuição amostral

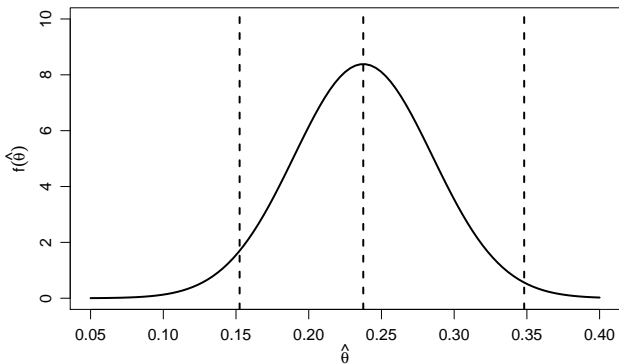
(código completo em arquivo **inf-prop.R**) As estimativas variam se tomamos diversas amostras da população:

summary(ps)

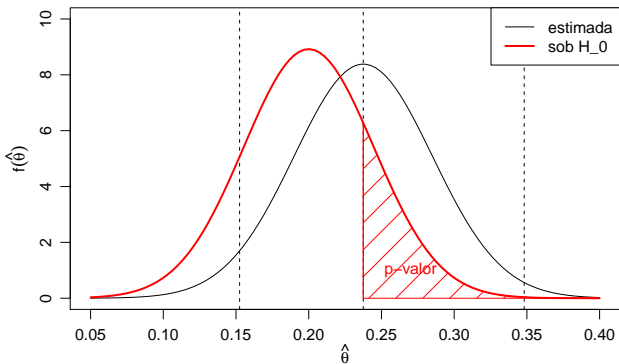
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.0375	0.1375	0.1625	0.1693	0.2000	0.3500



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 **valor-p**: *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.*

Uma alternativa (ainda) frequentista: Teste aleatorizado

Ideia básica:

Reproduzir a essência da ideia frequentista porém obtendo a **distribuição amostral** por simulação sob H_0

Algoritmo:

- ▶ 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

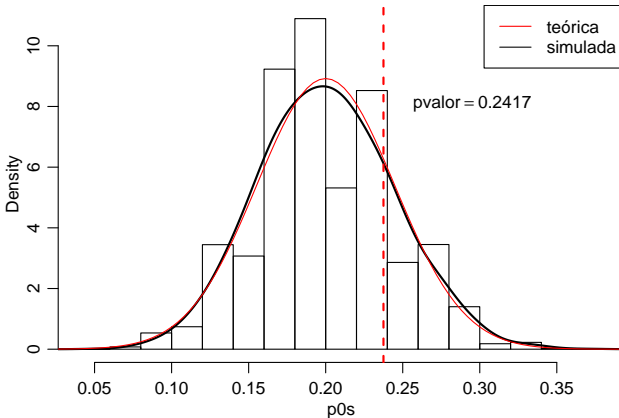
Teste aleatorizado

```
summary(p0s)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0500  0.1750  0.2000  0.2005  0.2250  0.3750
```

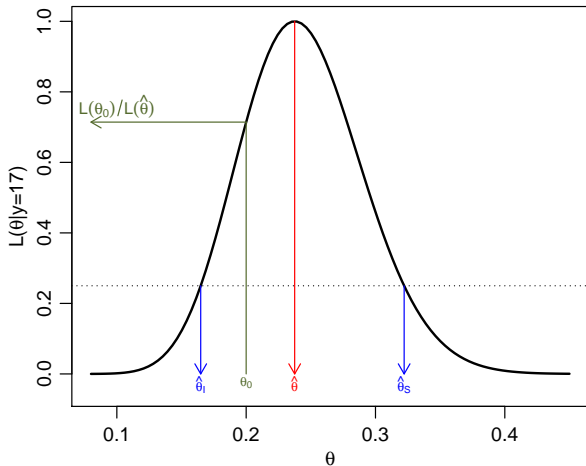
```
(pvalor <- mean(p0s >= 19/80))
```

```
## [1] 0.2417
```



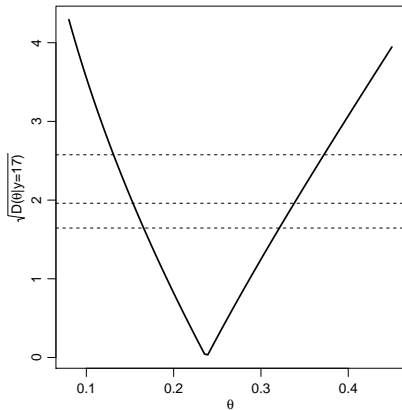
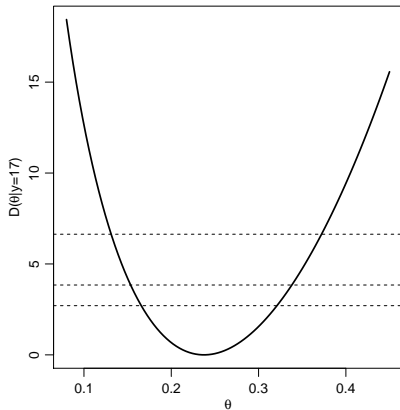
Abordagem pela verossimilhança

Inferência é baseada nas características da
função de verossimilhança



Representação alternativa

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



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. equivalência com caras consecutivas)
- ▶ 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

Comparando

Inferência frequentista

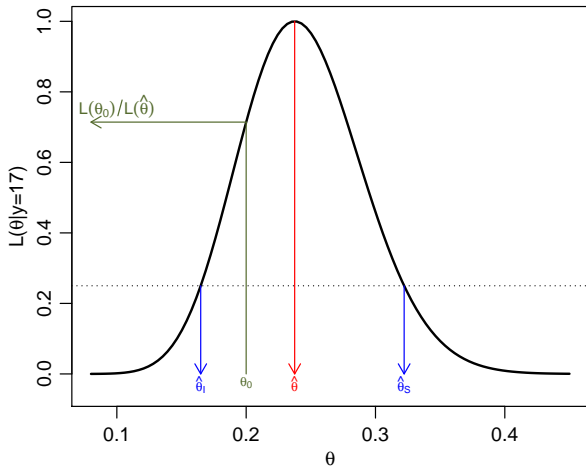
- ▶ **Estimativa de θ** : fornecido por algum método de estimação
- ▶ **expressão da incerteza**: variabilidade da distribuição amostral
- ▶ **opinião em relação a valor de interesse $\theta_0 = 0, 20$** :
probabilidade na distribuição amostral

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

All we need is ...likelihood

Inferência é baseada nas características da
função de verossimilhança



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**

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:

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

ou, usando jargão técnico:

$$posteriori \propto priori \cdot verossimilhança$$

Abordagem Bayesiana

No contexto do exemplo de estimação de proporção θ :

- ▶ Extensão da definição do modelo:

$$[Y|\theta] \sim B(n, \theta)$$

$$[\theta] \sim Pr(\psi) \text{ (priori)}$$

- ▶ permite obter (via teorema de Bayes)

$$[\theta|y] \propto [Y|\theta][\theta]$$

$$[\theta|y] \sim \pi(\psi^*) \text{ (posteriori)}$$

- ▶ 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).

Abordagem Bayesiana

No contexto do exemplo de estimação de proporção θ :

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

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

- Verossimilhança: $[Y|\theta] \sim \text{Bin}(n, \theta)$ (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

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

(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 :

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

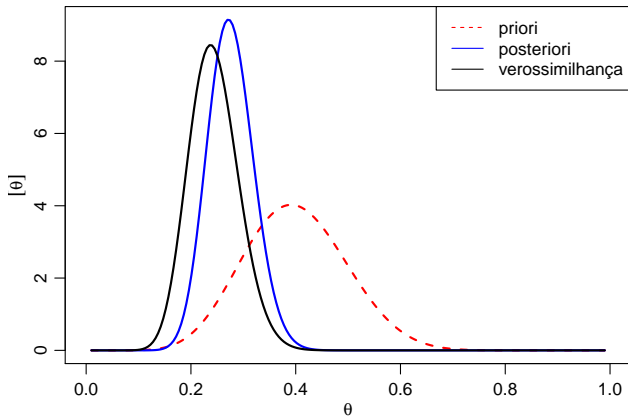
Verossimilhança: Modelo Binomial, amostra $n=80$ e $y=19$

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

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

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

A essência de Bayes ilustrada (I)



A essência de Bayes ilustrada (I)

```
(prI <- priorBeta(0.4, c(0.30, 0.50), 0.70))  
##      alpha      beta  
## 9.912277 14.868415  
postBinom(19, 80, prI, plot=FALSE)  
## $pars  
##           alpha      beta  
## prior      9.912277 14.86842  
## posterior 28.912277 75.86842  
##  
## $summary  
##           moda      media  variancia  
## prior      0.3912206 0.271571210 0.400000000  
## posterior 0.2759313 0.009309292 0.00188875  
##  
## $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%.

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

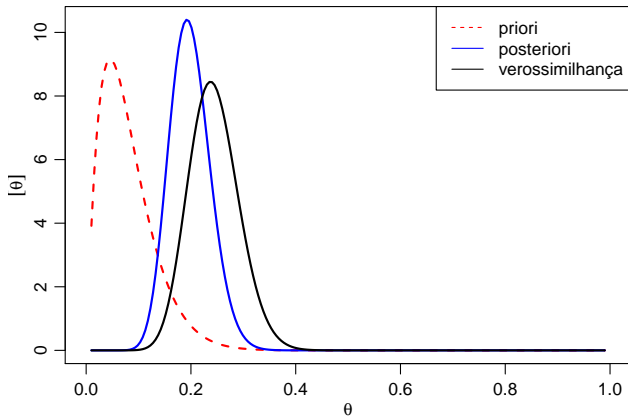
- **Verossimilhança:** Modelo Binomial, amostra $n=80$ e $y=19$

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

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

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

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 0.192469955 0.080000000  
## posteriori 0.1982418 0.002670415 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%.

$$[\theta] \sim \text{Beta}(1.2, 1.2)$$

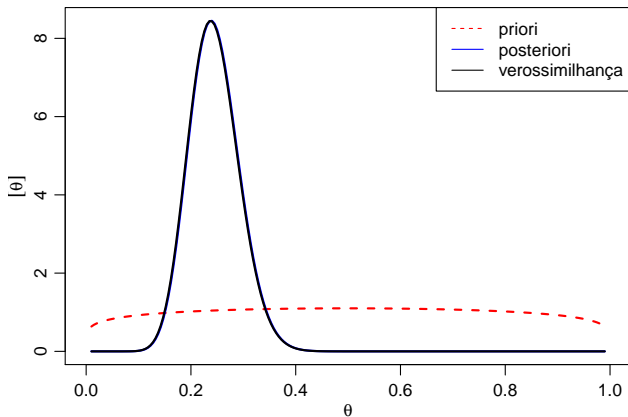
- **Verossimilhança**: Modelo Binomial, amostra $n=80$ e $y=19$

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

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

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

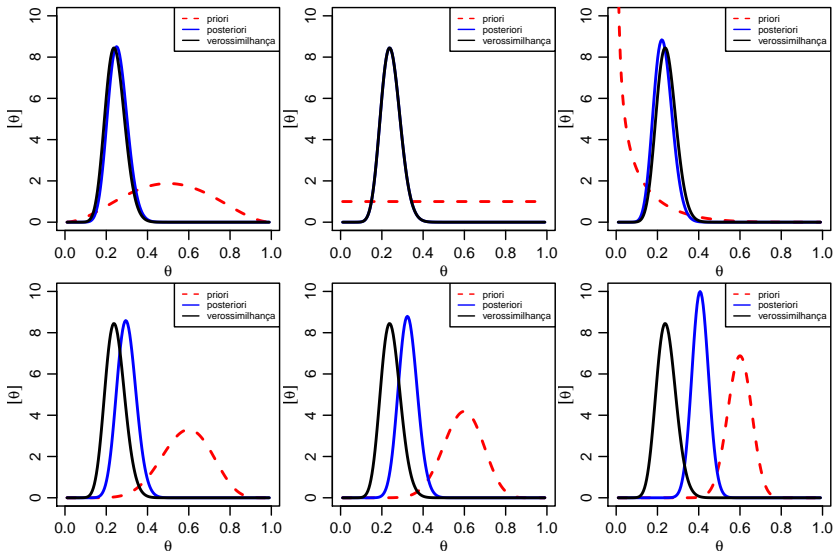
A essência de Bayes ilustrada (III)



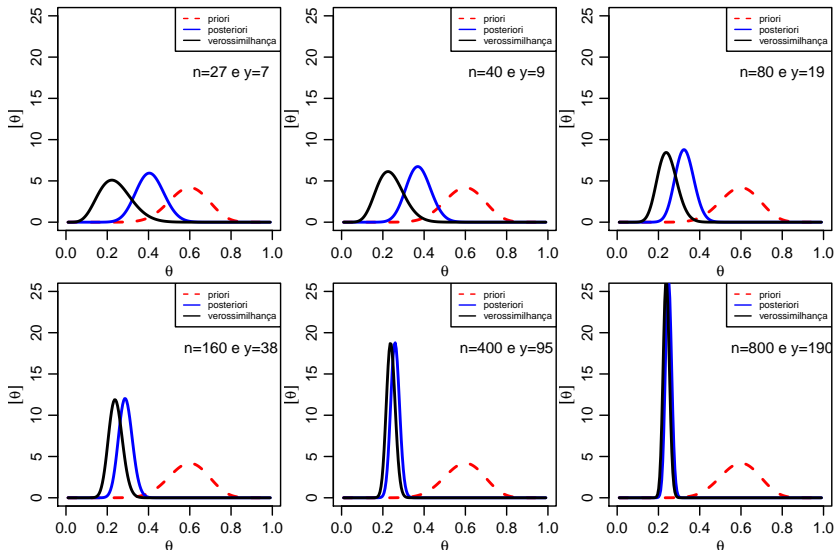
A essência de Bayes ilustrada (III)

```
(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 0.23861148 0.500000000  
## posteriori 0.2449605 0.07484633 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