

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**):

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

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

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

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

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

- ▶ 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:

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

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**

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

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

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

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

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

Abordagem frequentista

Inferência se baseia na **distribuição amostral**

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

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

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

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

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

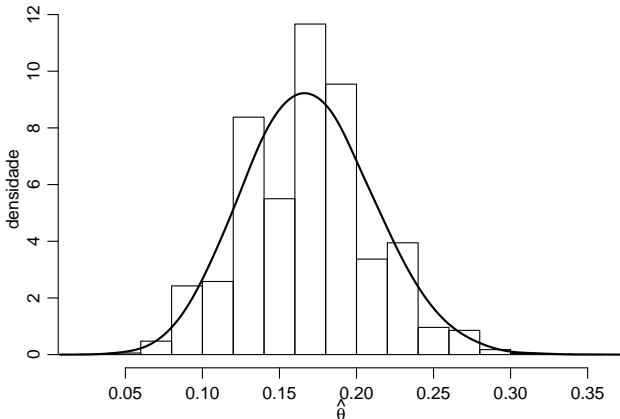
Usa-se $\theta = \hat{\theta}$ (assintótico) ou $\theta = 0,5$ (conservador)

Simulação da distribuição amostral

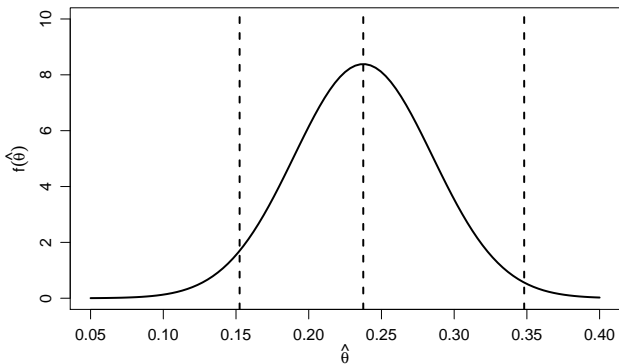
(código completo em arquivo **inf-prop.R**) As estimativas variam se tomamos diversas amostras da população:

```
summary(ps)
```

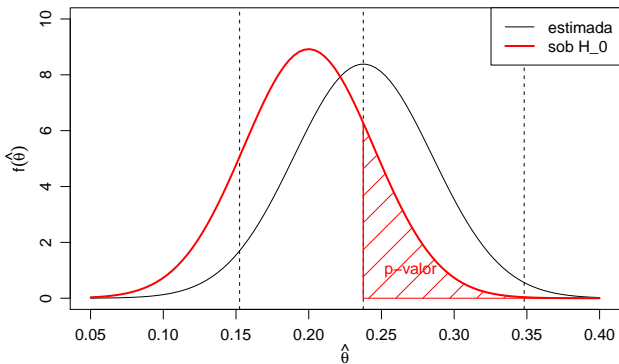
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0375 0.1375 0.1625 0.1693 0.2000 0.3500
```



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

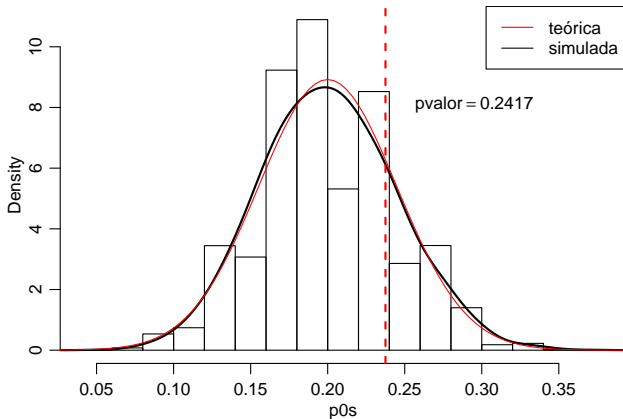
Teste aleatorizado

```
summary(p0s)
```

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

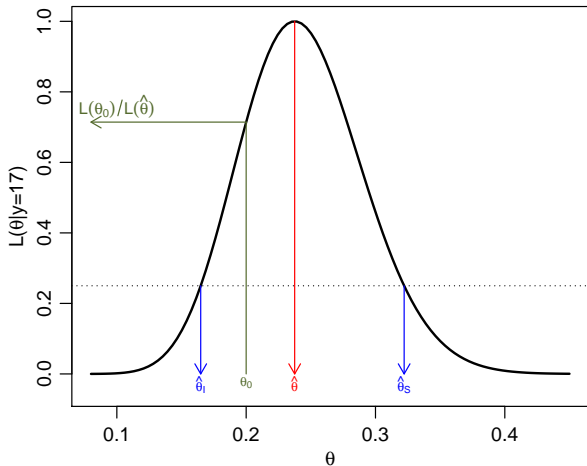
```
(pvalor <- mean(p0s >= 19/80))
```

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



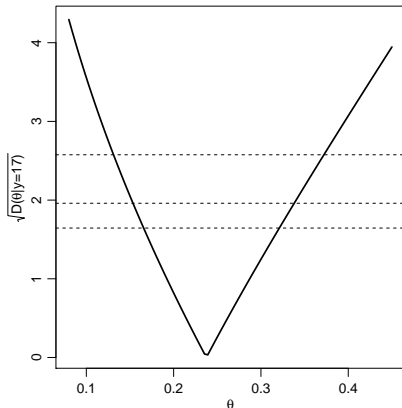
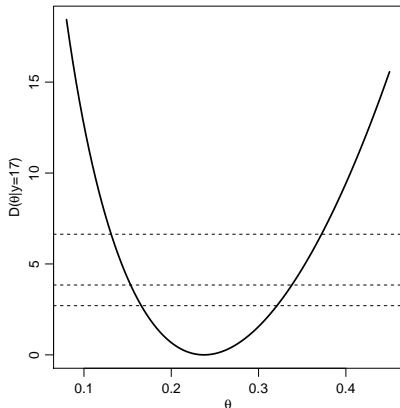
Abordagem pela verossimilhança

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

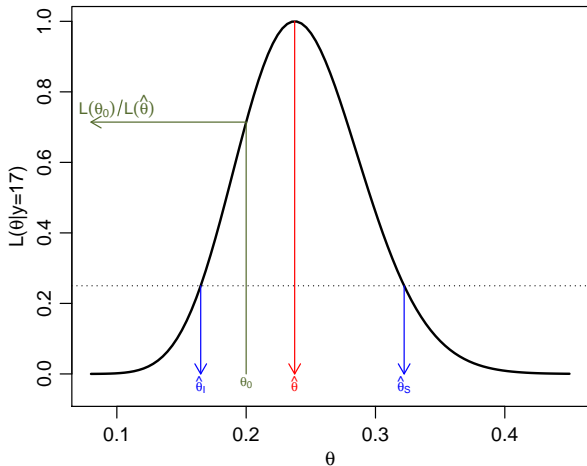
- ▶ **Estimativa de θ** : fornecido por algum método de estimação
- ▶ **expressão da incerteza**: variabilidade da distribuição amostral
- ▶ **opinião em relação a valor de interesse $\theta_0 = 0, 20$** :
probabilidade na distribuição amostral

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 :

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

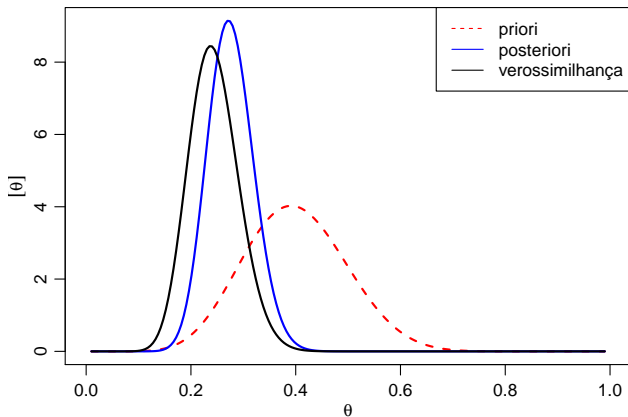
Verossimilhança: Modelo Binomial, amostra $n=80$ e $y=19$

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

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 <- 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 0.271571210 0.400000000  
## posteriori 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%.

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

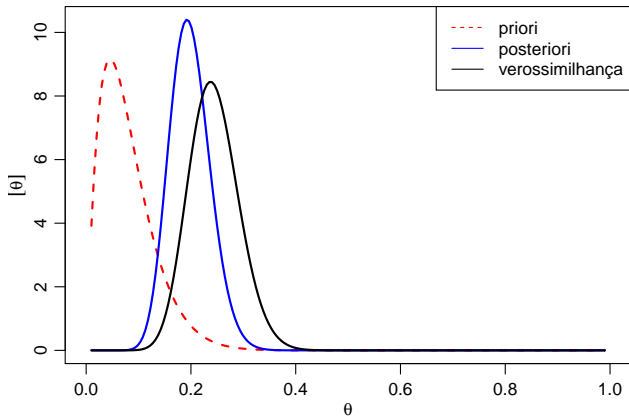
- ▶ **Verossimilhança:** Modelo Binomial, amostra $n=80$ e $y=19$

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

- ▶ **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%.

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

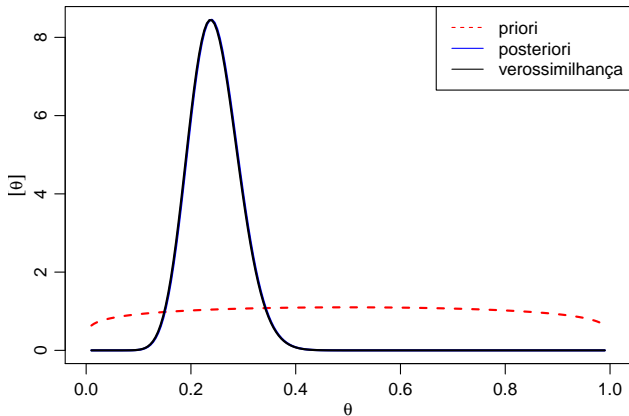
- ▶ **Verossimilhança**: Modelo Binomial, amostra $n=80$ e $y=19$

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

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

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

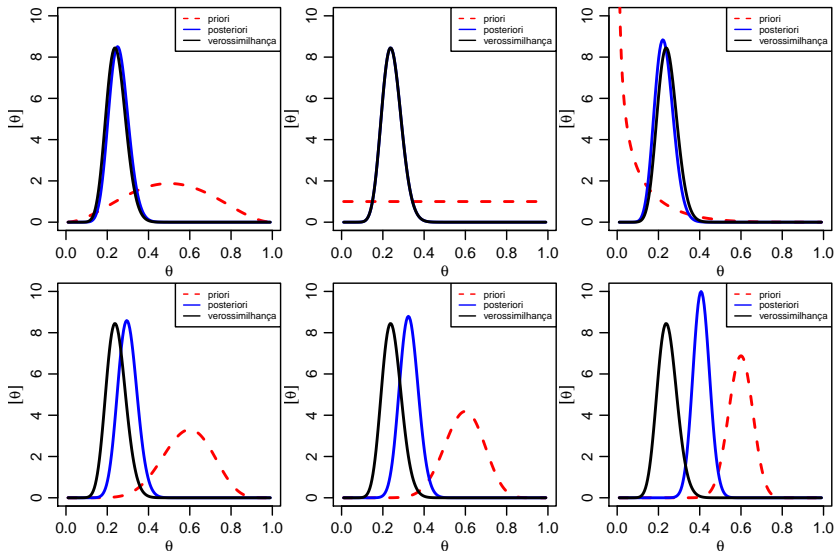
A essência de Bayes ilustrada (III)



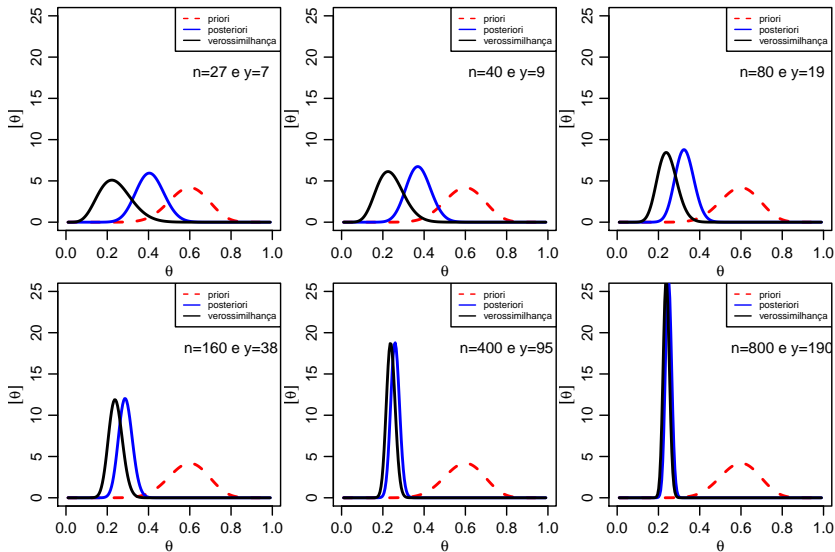
A essência de Bayes ilustrada (III)

```
(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

Exemplo 2: Comparando dois grupos

Questões de inferência:

- ▶ Quais são os **parâmetros** (quantidades de interesse na população) a serem estimados a partir dos dados?
- ▶ Qual a **incerteza** associada ao(aos) parâmetro(s) de interesse?
- ▶ **Pode-se afirmar** que os grupos diferem quanto ao peso e/ou altura?

As questões podem ser endereçadas pelos paradigmas frequentista, de verosimilhança e bayesiano.

OBS: pode-se adotar o ainda algum métodos não paramétrico mas este é, tipicamente, ainda um método frequentista.

Exemplo 2: Comparando dois grupos (cont)

Uma possível formatação do problema:

$$\textit{Grupo1} : Y_1 \sim N(\mu_1, \sigma^2)$$

$$\textit{Grupo2} : Y_2 \sim N(\mu_2, \sigma^2)$$

Exemplo 2: Comparando dois grupos (cont)

Uma possível formatação do problema:

$$\text{Grupo1} : Y_1 \sim N(\mu_1, \sigma^2)$$

$$\text{Grupo2} : Y_2 \sim N(\mu_2, \sigma^2)$$

Reexpressão alternativa (e conveniente)

$$\text{Grupo1} : Y_1 \sim N(\mu, \sigma^2)$$

$$\text{Grupo2} : Y_2 \sim N(\mu + \theta, \sigma^2)$$

Parâmetros: μ , θ e σ^2

- ▶ Quais são os **parâmetros** (quantidades desconhecidas na população) a serem estimados a partir dos dados?
- ▶ Qual a **incerteza** associada ao(aos) parâmetro(s) de interesse?
- ▶ **Pode-se afirmar** que os grupos diferem quanto ao peso e altura?

Exemplo 2: Comparando dois grupos (cont)

Sob os paradigmas discutidos aqui baseia-se a inferência em:

- ▶ frequentista: distribuição amostral da $\hat{\theta}$
- ▶ verossimilhança: na função de verossimilhança (perfilhada) para $\hat{\theta}$
- ▶ bayesiano: na distribuição à posteriori para θ

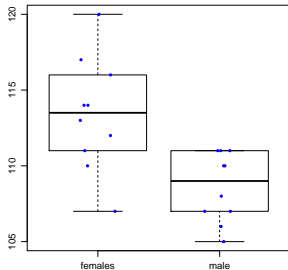
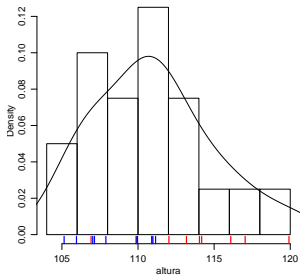
Procedimentos frequentistas incluem testes e procedimentos alternativos tais como não paramétricos, aleatorizados, *bootstrap*.

Um outro exemplo

Comparando dois grupos: comprimento da mandíbula (chacal dourado)

##	females	male
## 1	120	110
## 2	107	111
## 3	110	107
## 4	116	108
## 5	114	110
## 6	111	105
## 7	113	107
## 8	117	106
## 9	114	111
## 10	112	111

Um outro exemplo



Um outro exemplo

- ▶ Frequentista:
 - ▶ analítica: teste- t , opções e limitações (formulário)
 - ▶ computacional: teste aleatorizado – algoritmo
- ▶ Verossimilhança (perfilhada) para quantidade de interesse
- ▶ Bayesiana – distribuição marginal para parâmetro de interesse

Inferência frequentista

$$\text{Estimativa : } \hat{\theta} = \bar{y}_2 - \bar{y}_1$$

$$\hat{\theta} \sim t_{\nu}(\cdot) \text{ (distribuição amostral)}$$

$$\text{IC : } \hat{\theta} \pm t_{1-\alpha/2}(\nu) S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\text{TH : } (H_1 : \theta \neq \theta_0 = 0) : t = \frac{\bar{y}_2 - \bar{y}_1}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(\nu)$$

$$\nu = n_1 + n_2 - 2 \text{ e } S = \hat{\sigma} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

O teste- t (1)

teste- t para duas amostras supondo variâncias desconhecidas iguais entre os grupos.

```
with(mandible, t.test(females, male, var.equal=TRUE))  
##  
## Two Sample t-test  
##  
## data: females and male  
## t = 3.4843, df = 18, p-value = 0.002647  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 1.905773 7.694227  
## sample estimates:  
## mean of x mean of y  
## 113.4 108.6
```

O teste- t - opções

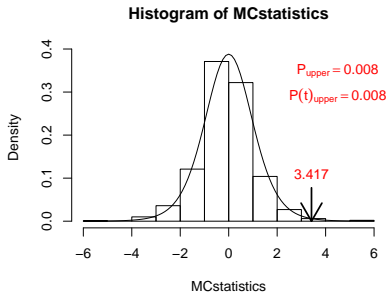
Diferentes “cenários” para o teste- t :

- ▶ **Amostras independentes**
 - ▶ Variâncias iguais
 - ▶ Variâncias diferentes
- ▶ Amostras pareadas (dependentes)

Opções na função para definições do teste- t do R:

```
t.test(x, y = NULL,  
       alternative = c("two.sided", "less", "greater"),  
       mu = 0, paired = FALSE, var.equal = FALSE,  
       conf.level = 0.95, ...)
```

Um outro exemplo



```
## Paired data
## data statistics = 3.416968784709
##
## probabilities based on Monte Carlo simulations:
## upper.tail lower.tail
## 0.007992 0.992008
##
## probabilities based on the "t" distribution:
## upper.tail lower.tail
## 0.003832 0.996168
```

Inferência pela verossimilhança

Princípios e procedimentos *gerais*:

- ▶ Adotar um modelo para os dados
- ▶ Notar os **parâmetros** do modelo
- ▶ Obter a função de verossimilhança para os dados obtidos (função dos **parâmetros**)
- ▶ Maximizar a função de verossimilhança (estimativas são os valores dos **parâmetros** que maximizam a função)
- ▶ Obter inferências de interesse sobre parâmetros de interesse

Procedimentos podem ser **analíticos** (matemáticos) ou **numéricos** (computacional).

Inferência pela verossimilhança

Sob modelo com variância comum: **Função de verossimilhança:**

```
require(stats4)
l11 <- function(mu, theta, lsigma, am1, am2){
  sigma <- exp(lsigma)
  l1 <- sum(dnorm(am1, m=mu, sd=sigma, log=T))
  l2 <- sum(dnorm(am2, m=mu+theta, sd=sigma, log=T))
  return(-(l1+l2))
}
```

Maximização:

```
fit1 <- mle(l11, start=list(mu=110, theta=0, lsigma=log(10)),
           fixed=list(am1 = mandible$fem, am2 = mandible$male))
```

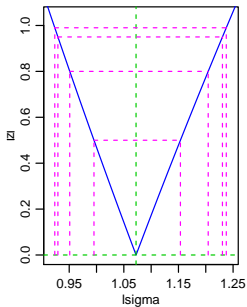
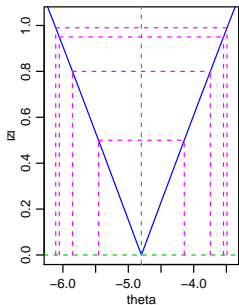
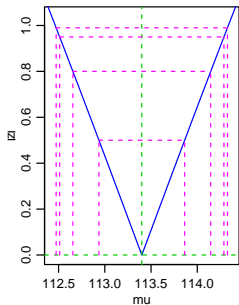
Inferência pela verossimilhança

Inferências sobre parâmetro de interesse:

```
summary(fit1)
## Maximum likelihood estimation
##
## Call:
## mle(minuslogl = ll1, start = list(mu = 110, theta = 0, lsigma = log(10)),
##     fixed = list(am1 = mandible$fem, am2 = mandible$male))
##
## Coefficients:
##           Estimate Std. Error
## mu       113.400003  0.9241214
## theta    -4.800002  1.3069050
## lsigma    1.072381  0.1581138
##
## -2 log L: 99.65276
confint(fit1, level=0.95)
## Profiling...
##           2.5 %      97.5 %
## mu       111.4978494 115.302151
## theta    -7.4900472  -2.109953
## lsigma    0.7914638  1.417925
prof1 <- profile(fit1)
```

Inferência pela verossimilhança

Inferências sobre parâmetro de interesse:



Inferência pela verossimilhança

Sob modelo com variâncias distintas: **Função de verossimilhança:**

```
l12 <- function(mu, theta, lsigma1, lsigma2, am1, am2){  
  sigma1 <- exp(lsigma1)  
  sigma2 <- exp(lsigma2)  
  l1 <- sum(dnorm(am1, m=mu, sd=sigma1, log=T))  
  l2 <- sum(dnorm(am2, m=mu+theta, sd=sigma2, log=T))  
  return(-(l1+l2))  
}
```

Maximização:

```
fit2 <- mle(l12,  
  start=list(mu=110, theta=0, lsigma1=log(10), lsigma2=log(10)),  
  fixed=list(am1 = mandible$fem, am2 = mandible$male))
```

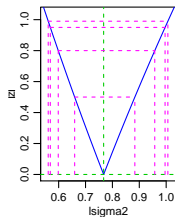
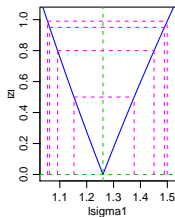
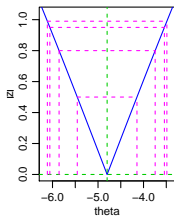
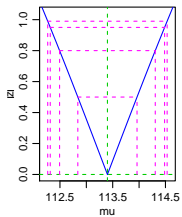
Inferência pela verossimilhança

Inferências sobre parâmetro de interesse:

```
summary(fit2)
## Maximum likelihood estimation
##
## Call:
## mle(minuslogl = ll2, start = list(mu = 110, theta = 0, lsigma1 = log(10),
##   lsigma2 = log(10)), fixed = list(am1 = mandible$fem, am2 = mandible$male))
##
## Coefficients:
##           Estimate Std. Error
## mu          113.399998  1.1153465
## theta        -4.799999  1.3069034
## lsigma1       1.260458  0.2236064
## lsigma2       0.767356  0.2236064
##
## -2 log L: 97.31386
confint(fit2, level=0.95)
## Profiling...
##           2.5 %      97.5 %
## mu          110.9855144 115.814486
## theta        -7.5249732 -2.075027
## lsigma1       0.8779986  1.773255
## lsigma2       0.3848972  1.280154
prof2 <- profile(fit2)
```

Inferência pela verossimilhança

Inferências sobre parâmetro de interesse:



Comparando modelos pelas verossimilhanças

Comparando modelos para os dados

$$Y_{ij} \sim N(\mu, \sigma^2)$$

$$Y_{ij} \sim N(\mu_i, \sigma^2)$$

$$Y_{ij} \sim N(\mu_i, \sigma_i^2)$$

```
fit0 <- lm(mandible~1, data=mand)
m1 <- lm(mandible~sex, data=mand) ## equivalente a fit1
(lLs <- c(logLik(fit0), logLik(fit1), logLik(fit2)))
## [1] -54.98137 -49.82638 -48.65693
```

Comparando modelos pelas verossimilhanças

Testes da Razão de Verossimilhanças

```
nps <- c(2, 3, 4)
(dlls <- -2 * diff(lLs))
## [1] 10.309971  2.338906
(dnps <- diff(nps))
## [1] 1 1
2*pchisq(dlls, df=dnps, lower=FALSE)
## [1] 0.002646268 0.252356341
```


Verossimilhanças penalizadas

$$-2 \log(Lik) + k * \#par$$

AIC's: $k = 2$

```
-2 * logLik(fit0) + 2 * 2
## 'log Lik.' 113.9627 (df=2)
-2 * logLik(fit1) + 2 * 3
## 'log Lik.' 105.6528 (df=3)
-2 * logLik(fit2) + 2 * 4
## 'log Lik.' 105.3139 (df=4)
c(AIC(fit0), AIC(fit1), AIC(fit2))
## [1] 113.9627 105.6528 105.3139
```

BIC's: $k = \log(n)$

```
-2 * logLik(fit0) + log(20) * 2
## 'log Lik.' 115.9542 (df=2)
-2 * logLik(fit1) + log(20) * 3
## 'log Lik.' 108.64 (df=3)
-2 * logLik(fit2) + log(20) * 4
## 'log Lik.' 109.2968 (df=4)
c(AIC(fit0, k=log(20)), AIC(fit1, k=log(20)), AIC(fit2, k=log(20)))
## [1] 115.9542 108.6400 109.2968
```

Inferência Bayesiana

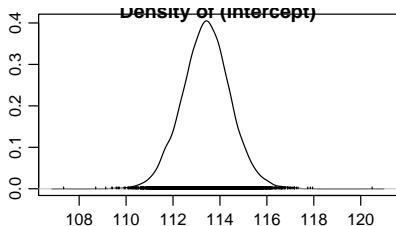
Obtenção de (amostras da) distribuição **posteriori** para a diferença de médias

```
require(MCMCpack)
Btt <- MCMCregress(mandible ~ sex, dat=mand)
summary(Btt)

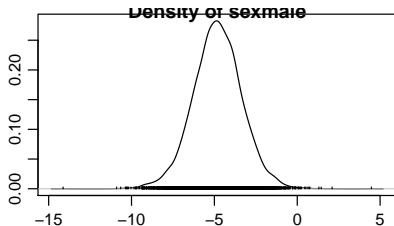
##
## Iterations = 1001:11000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 10000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##              Mean      SD Naive SE Time-series SE
## (Intercept) 113.413 1.040 0.01040      0.01017
## sexmale     -4.817 1.479 0.01479      0.01479
## sigma2      10.728 4.104 0.04104      0.04573
##
## 2. Quantiles for each variable:
##
##              2.5%      25%      50%      75%      97.5%
## (Intercept) 111.374 112.751 113.421 114.078 115.483
## sexmale     -7.807  -5.759  -4.822  -3.867  -1.887
## sigma2       5.454   7.901   9.875  12.523  20.853
```

Inferência Bayesiana

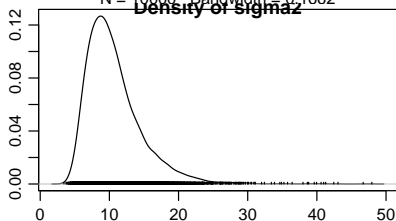
Obtenção de distribuição **posteriori** para a diferença de médias



N = 10000 Bandwidth = 0.1662



N = 10000 Bandwidth = 0.2371



N = 10000 Bandwidth = 0.5795

Abordagem geral e generalizando

Generalizações ... muitas possíveis
Vamos começar reescrevendo

$$Y_{ij}^{(\lambda)} \sim N(\mu_{ij}, \sigma_{ij}^2)$$

$$g(\mu_{ij}) = f(x_{ij}, \beta)$$

$$g(\sigma_{ij}^2) = f(z_{ij}, \varphi)$$

Abordagem geral e generalizando

- ▶ Especificação do modelo associado ao teste- t (amostras independentes, variâncias iguais)

$$Y_{ij} \sim N(\mu_{ij}, \sigma_{ij}^2)$$

$$\mu_{ij} = \beta_0 + \beta_1 x_{sex}$$

$$\sigma_{ij}^2 = \sigma^2$$

- ▶ Mudando a distribuição

$$Y_{ij} \sim G(\mu_{ij}, \phi_{ij})$$

$$\log(\mu_{ij}) = \beta_0 + \beta_1 x_{sex}$$

$$\phi_{ij} = \phi$$