MCIE:
Métodos Computacionais para Inferência Estatística

com ênfase na verossimilhança

Paulo Justiniano Ribeiro Jr.
Wagner Hugo Bonat
(Elias Teixeira Krainski)
(Walmes Marques Zeviani)

LEG: Laboratório de Estatística e Geoinformação
Universidade Federal do Paraná

MCIE, 2o semestre 2019

Comentários

- Texto de referência:
  - MCIE (PJ, WB, EK & WZ)
  - Disponível na Página: `http://www.leg.ufpr.br/mcie`
  - Capítulos 1 e 2
- Motivações e propósitos do texto e das aulas
  - Visão unificada e intuitiva dos princípios de inferência estatística
  - Facilidade de recursos computacionais e linguagens
  - Uso de rotinas *versus* implementação/teste/ilustração/aprendizado
  - Uso crítico e avaliação e apreciação das limitações de rotinas

Exemplo introdutório

Entendendo e explicando **verossimilhança**.
Visitando um exemplo simples:

- População: $X \sim \mathrm{Ber}(\theta)$
- Amostra: $x_1, \ldots, x_n$
- $Y = x_1 + x_2 \ldots + x_n \sim \mathrm{B}(n, \theta)$
- O que podemos falar sobre θ?
  - Qual a informação contida na amostra?
  - Consideram-se outras fontes de informação?
- informação na amostra resumida por $(n, y = \sum_{i=1}^{n} x_i)$?

Exemplo: $n = 100$ e $y = 87$

Construindo a verossimilhança

- temos dados (amostra), ou seja, $y = 87$
- pode-se calcular a probabilidade de observar $y = 87$ para um valor de θ, por exemplo para $\theta = 0{,}80$

$$P[Y = 87|n = 100, \theta = 0{,}80] = \binom{100}{87} 0{,}80^{87}(1 - 0{,}80)^{100-87}$$

- e para qualquer outro valor de θ

$$P[Y = 87|n = 100, \theta] = \binom{100}{87} \theta^{87}(1 - \theta)^{100-87}$$

- variando θ temos a **função de verossimilhança**

$$L(\theta) \equiv P[Y = 87|n = 100, \theta] = \binom{100}{87} \theta^{87}(1 - \theta)^{100-87}$$

Espaço do Modelo

- Supondo $Y \sim \mathrm{B}(n = 100, \theta)$ e $y = 87$
- O espaço definido pelo modelo (3-D)

$$\binom{n}{y}\theta^y(1-\theta)^{n-y}$$

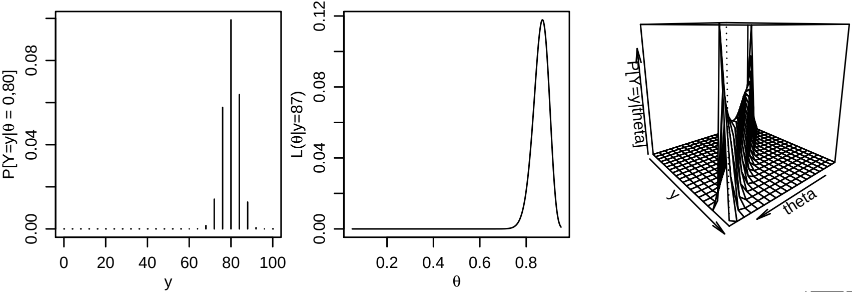


Figura 1: Distribuição de Y, Verossimilhança e espaço do modelo

Verossimilhança

- Verossimilhança: $L(\theta) \equiv P_\theta[y = 87] = \binom{100}{87}\theta^{87}(1-\theta)^{100-87}$
- Verossimilhança Relativa:
  $LR(\theta) = \frac{L(\theta)}{\max L(\theta)} = \frac{L(\theta)}{L(\hat{\theta})} = (\frac{\theta}{\hat{\theta}})^{87}(\frac{1-\theta}{1-\hat{\theta}})^{100-87}$

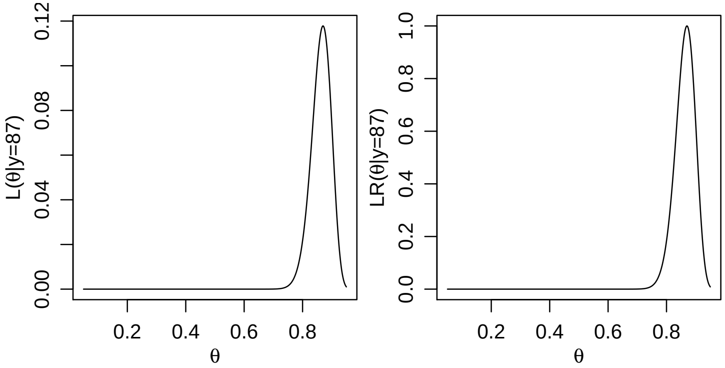


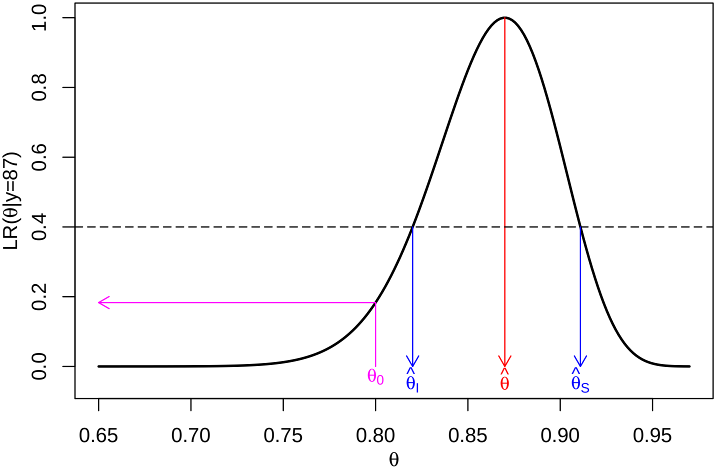
Figura 2: Funções de verossimilhança e verossimilhança relativa

Objetivos de inferência . . .

- função de verossimilhança
  probabilidades da amostra obtida para diferentes valores de θ
- melhor estimador e estimativa
- incerteza associada à estimativa obtida
- conjunto de valores razoavelmente compatíveis com a amostra
- decidir entre dois valores o mais compatível com a amostra
- decidir se a amostra é compatível com certo valor θ_0 de interesse?

... na função de verossimilhança

Objetivos de inferência representados na função de verossimilhança



Comentários

- Suposições/pressupostos.
- Não é o único paradigma para inferência.
- Relações e contrastes com outros paradigmas.
- Meio do caminho entres aborgadens frequentista e bayesiana (?)
- Várias propostas para aproximações, modificações, etc.
- Mas a intuição permanece válida para a lógica do pensamento estatístico.
- Problemas irregulares podem levar a formas "desafiadoras" da verossimilhança.

Formas alternativas

- Verossimilhança:

$$L(\theta)$$

- Verossimilhança Relativa:

$$LR(\theta) = \frac{L(\theta)}{L(\hat{\theta})}$$

- log-Verossimilhança:

$$l(\theta) = \log\{L(\theta)\}$$

- *Deviance*:

$$D(\theta) = -2\log\{LR[\theta]\} = -2\{l[\theta] - l[\hat{\theta}]\}$$

Formas alternativas (exemplo binomial)

- Verossimilhança:

$$L(\theta) = \binom{100}{87} \theta^{87} (1-\theta)^{100-87}$$

- Verossimilhança Relativa:

$$LR(\theta) = \frac{L(\theta)}{L(\hat{\theta})} = \left(\frac{\theta}{\hat{\theta}}\right)^{87} \left(\frac{1-\theta}{1-\hat{\theta}}\right)^{100-87}$$

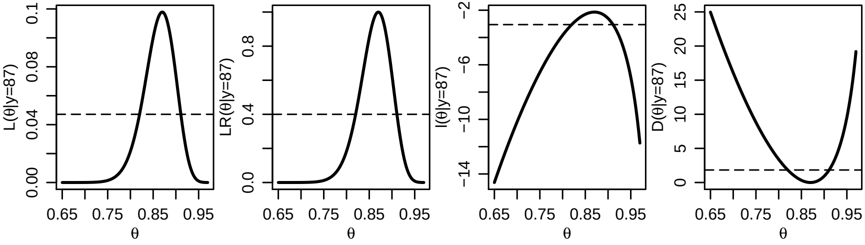
- log-Verossimilhança:

$$l(\theta) = \log\{L(\theta)\} = \log \binom{100}{87} + 87 \log(\theta) + (100-87) \log(1-\theta)$$

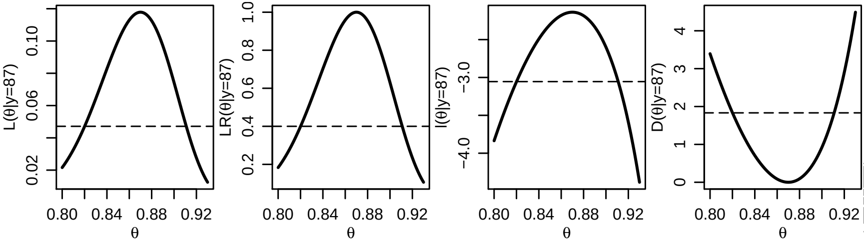
- *Deviance*:

$$D(\theta) = -2\{l[\theta] - l[\hat{\theta}]\} = -2 \left\{ 87 \log\left(\frac{\theta}{\hat{\theta}}\right) + (100-87) \log\left(\frac{1-\theta}{1-\hat{\theta}}\right) \right\}$$

Formas alternativas da função de verossimilhança



“Zoom” na região próxima do máximo



Expressão da Verossimilhança I

V.A. observável discreta (não há ambiguidade)

$$L(\theta) \equiv P_\theta[\underline{Y} = \underline{y}] = P_\theta[Y_1 = y_1, \ldots, Y_n = y_n]$$

Sob independência

$$L(\theta) \equiv \prod_{i=1}^{n} P_\theta[Y_i = y_i]$$

Exemplo:

$$Y \sim P(\theta)$$

Dados: $(y_1, \ldots, y_n)$, amostra aleatória

$$L(\theta) = \prod_{i=1}^{n} \frac{\exp\{-\theta\}\theta^{y_i}}{y_i!} = \frac{\exp\{-n\theta\}\theta^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i!} \propto \exp\{-n\theta\}\theta^{\sum_{i=1}^{n} y_i}$$

Exemplo: distribuição Poisson

$$
\begin{aligned}
L[\theta] &= \frac{\exp\{-n\theta\}\theta^{n\overline{Y}}}{\prod_{i=1}^{n} Y_i!} \\
LR[\theta] &= \exp\{-n(\theta - \hat{\theta})\}(\theta/\hat{\theta})^{n\overline{Y}} \\
l[\theta] &= -n\{\theta + \overline{Y}\log(\theta) - \overline{\log(Y_i!)}\} \\
D(\theta) &= -2n\{(\theta - \hat{\theta}) - \overline{Y}\log(\theta/\hat{\theta})\}
\end{aligned}
$$

Para uma a.a. de observações pontuais:

$$\hat{\theta} = \overline{Y}$$

Exemplo: Distribuição Poisson

1. Código 1

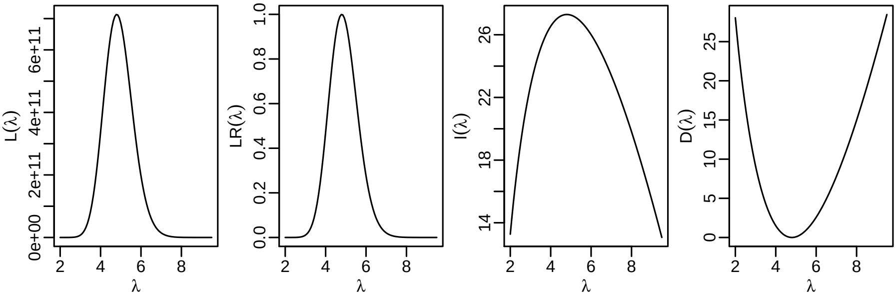
```
veroPois <- function(par, dados, tipo, maxlogL){
   tipo = match.arg(tipo, choices=c("L","LR","logL","dev"))
   ll <- sapply(par, function(p) sum(dpois(dados, lambda=p,
                                             log=TRUE)))
   return(switch(tipo, "L" = exp(ll),
                       "LR" = exp(ll-maxlogL),
                       "logL" = ll,
                       "dev" = 2*(maxlogL-ll)))}
```

2. Código 2

```
veroPois <- function(par, amostra, tipo="logL", maxlogL){
   tipo = match.arg(tipo, choices=c("L","LR","logL","dev"))
   ll <- with(amostra, -n*par + soma * log(par))
   return(switch(tipo, "L" = exp(ll),
                       "LR" = exp(ll-maxlogL),
                       "logL" = ll,
                       "dev" = 2*(maxlogL-ll)))}
```

Exemplo: Distribuição Poisson (cont)

```
set.seed(2012)
(dPo <- rpois(10, lambda=3.5))
amPo <- list(n=length(dPo), soma=sum(dPo))
emvPo <- mean(dPo)
maxll <- veroPois(emvPo, amostra=amPo, tipo="logL")
cat("...")
curve(veroPois(x, amostra=amPo, tipo="dev", maxlogL=maxll), 2, 9.5,
      ylab=expression(D(lambda)), xlab=expression(lambda))
```



Expressão da Verossimilhança II

V.A. contínua: medição à certa precisão $(y_{iI} \leq y_i \leq y_{iS})$

- Forma mais geral (densidade multivariada)

$$L(\theta) = P_\theta[y_{1I} \leq Y_1 \leq y_{1S}, \ldots, y_{nI} \leq Y_n \leq y_{nS}]$$

- Sob independência

$$\begin{aligned} L(\theta) &= P_\theta[y_{1I} \leq Y_1 \leq y_{1S}] \cdot P_\theta[y_{2I} \leq Y_2 \leq y_{2S}] \ldots P_\theta[y_{nI} \leq Y_n \leq y_{nS}] \\ &= \prod_{i=1}^{n} P_\theta[y_{iI} \leq Y_i \leq y_{iS}] \end{aligned}$$

- Se grau de precisão comum, $(y_i - \delta/2 \leq Y_i \leq y_i + \delta/2)$;

$$L(\theta) = \prod_{i=1}^{n} P_\theta[y_i - \delta/2 \leq Y_i \leq y_i + \delta/2] = \prod_{i=1}^{n} \int_{y_i - \delta/2}^{y_i + \delta/2} f(y_i, \underline{\theta}) d(y_i).$$

Expressão da Verossimilhança II (cont)

- alto grau de precisão (δ é pequeno em relação a variabilidade dos dados)

$$L(\theta) \approx \left(\prod_{i=1}^{n} f(y_i, \underline{\theta}) \right) \delta^n,$$

- e se δ não depende dos valores dos parâmetros

$$L(\theta) \approx \prod_{i=1}^{n} f(y_i, \underline{\theta})$$

- observações não independentes - densidade multivariada:

$$L(\theta) \approx f(\underline{y}, \underline{\theta})$$

Verossimilhança e Informação

Considere $Y \sim N(\theta, 1)$ e as as seguintes observações.

1. $x = 2.45$
2. $0.9 < x < 4$
3. somente o máximo de uma amostra de tamanho cinco é fornecido $x_{(5)} = 3.5$

Verossimilhança:

$$L(\theta; x) = \phi(x - \theta) \equiv \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}(x - \theta)^2\};$$

$$L_1 = L(\theta; x = 2.45) = \phi(x - \theta) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}(2.45 - \theta)^2\};$$

$$L_2 = L(\theta; 0{,}9 < x < 4) = \Phi(4 - \theta) - \Phi(0{,}9 - \theta);$$

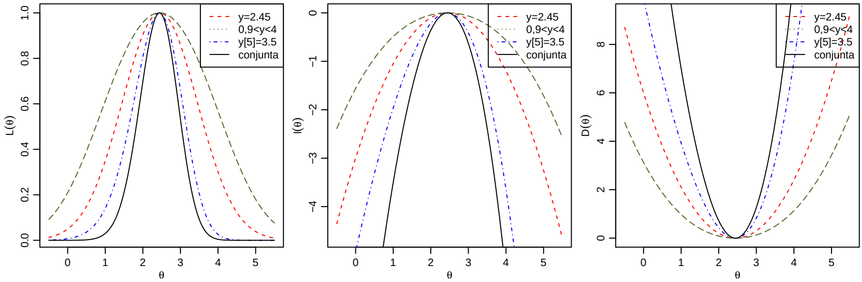
$$L_3 = L(\theta; x_{(5)} = 3.5) = n\{\Phi(x_{(n)} - \theta)\}^{n-1}\phi(x_{(n)} - \theta).$$

Para última - argumento multinomial e com

$$F(y) = P(X_{\{n\}} \leq y) = P[X_{\{i\}} < y \; \forall i \neq n \text{ e } X_{\{n\}} = y]$$

Verossimilhança e Informação (cont)

```
L1 <- function(theta) dnorm(2.45, m=theta, sd=1)
L2 <- function(theta)
    pnorm(4,mean=theta,sd=1)-pnorm(0.9,mean=theta,sd=1)
L3 <- function(theta)
    5*pnorm(3.5,m=theta,s=1)^4 * dnorm(3.5,m=theta,s=1)
```



Funções de interesse para inferência

- Função escore: $U(\theta) = l'(\theta)$
- Hessiano e Informação observada: $I_O(\theta) = -H(\theta) = -l''(\theta)$
- Informação Esperada: $I_E(\theta) = E_Y[I_O(\theta)]$
- Estimadas: $I_O(\hat{\theta})$ e $I_E(\hat{\theta})$

Propriedades assintóticas:

- $\hat{\theta} \sim NM_d(\underline{\theta}, I_E(\underline{\theta})^{-1})$
- Assintoticamente equivalentes:

$$\hat{\theta} \sim NM_d(\theta, I_E(\hat{\theta})^{-1})$$
$$\hat{\theta} \sim NM_d(\theta, I_O(\theta)^{-1})$$
$$\hat{\theta} \sim NM_d(\theta, I_O(\hat{\theta})^{-1}).$$

- $D(\theta) = -2[l(\theta) - l(\hat{\theta})] \sim \chi^2_d$

Exemplo: Distribuição Poisson (cont)

$$L(\theta) = \prod_{i=1}^{n} \frac{\exp\{-\theta\}\theta^{Y_i}}{Y_i!} = \frac{\exp\{-n\theta\}\theta^{\sum_{i=1}^{n} Y_i}}{\prod_{i=1}^{n} Y_i!} = \frac{\exp\{-n\theta\}\theta^{n\bar{Y}}}{\prod_{i=1}^{n} Y_i!}$$

$$l(\theta) = -n\theta + (\sum_{i=1}^{n} Y_i)\log(\theta) - \sum_{i=1}^{n} \log Y_i! = -n(\theta - \bar{Y}\log(\theta) - \overline{\log Y_i!})$$

$$U(\theta) = -n + \frac{\sum_{i=1}^{n} Y_i}{\theta} = -n(1 - \frac{\bar{Y}}{\theta})$$

$$U(\hat{\theta}) = 0 \longrightarrow \hat{\theta} = \frac{\sum_{i=1}^{n} Y_i}{n} = \overline{Y}$$

$$I_O(\theta) = \frac{\sum_{i=1}^{n} Y_i}{\theta^2} = \frac{n\bar{Y}}{\theta^2} \; ; \; I_E(\theta) = \frac{n}{\theta} \; ; \; I_O(\hat{\theta}) = I_E(\hat{\theta}) = \frac{n^2}{\sum_{i=1}^{n} Y_i} = \frac{n}{\overline{Y}}$$

$$V(\hat{\theta}) = I_E^{-1}(\theta) \approx I_O^{-1}(\theta) \approx I_O^{-1}(\hat{\theta}) = I_E^{-1}(\hat{\theta})$$

Exemplo: Distribuição Poisson (cont)

Função escore:

```
UPois <- function(lambda, amostra){
  return(with(amostra, n - soma/lambda))
}
```

Hessiano (negativo da Informação observada):

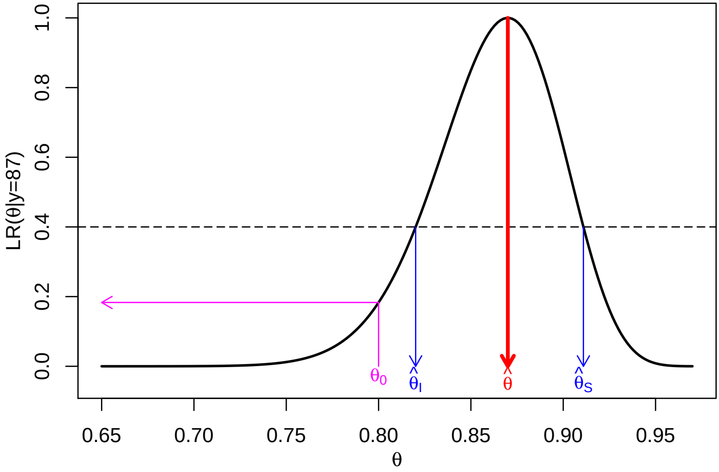
```
HPois <- function(lambda, amostra){
  return(with(amostra, -soma/lambda^2))
}
```

Informação esperada:

```
IePois <- function(lambda, amostra){
  return(with(amostra, n/lambda))
}
```

Estimação

Objetivos de inferência e a função de verossimilhança



Obtendo o EMV

Máximo da função de verossimilhança: $\hat{\theta} = \mathrm{argmax}_{\theta} l(\theta)$
(ou, melhor ainda colocando, o supremo da função)

- analiticamente:
  estudando comportamento de $l(\theta)$ ou resolvendo $U(\theta) = 0$
- numericamente (otimização/aproximações numéricas)
  - Solução da(s) equação(ões) de estimação (função escore)
    - com uso de derivadas (ex: Newton-Raphson)
    - sem uso de derivadas (ex: Brent)
  - Maximização da função de (log)-verossimilhança
- Outros (ex: EM)
- Simulação (ex: verossimilhança Monte Carlo, *data-cloning*, ...)
- Aproximações da verossimilhança (pseudo-verossimilhanças)

Estimadores e Inferência

- Análogos para distribuições *posteriori* em Inferência Bayesiana
- maximização numérica: mais comum em EMV
- simulação: mais usual em Inferência Bayesiana

EMV

Newton Raphson: expansão (Taylor) de 1ª ordem de $U(\theta)$:

$$\theta^{r+1} = \theta^r - \frac{U(\theta)}{H(\theta)} = \theta^r + \frac{U(\theta)}{I_o(\theta)}$$

```
maxit <- 100; lambdaNR <- 5; iter <- 0; d <- 1
while(d > 1e-12 & iter <= maxit){
    lambdaNR.new <-
      lambdaNR - UPois(lambdaNR, am)/HPois(lambdaNR, am)
    d <- abs(lambdaNR - lambdaNR.new)
    lambdaNR <- lambdaNR.new ; iter <- iter + 1
}
c(lambdaNR, iter)
```

Variante - *Fisher scoring*: substituir $I_O(\theta) \longrightarrow I_E(\theta)$

Potencial problema neste caso: $\theta > 0$

Exemplo: Distribuição Poisson (cont)

- Solução de equação $U(\theta) = 0$:

```
uniroot(UPois, interval=range(y), amostra=am)$root
```

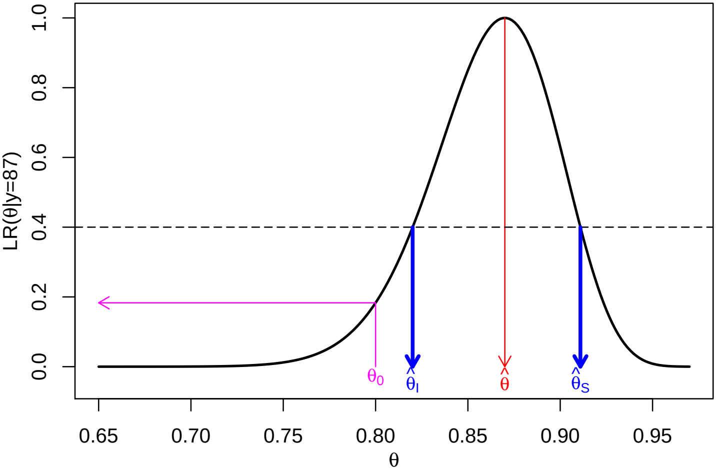
- Maximização da verossimilhança

```
optimize(veroPois, interval=range(y), maximum=TRUE, amostra=am)
optim(par = median(y), fn=veroPois, control=list(fnscale=-1),
      amostra=am, hessian = TRUE)
```

  - uso do gradiente: argumento `gr = Upois`
  - pode retornar hessiano estimado ($I_O(\hat{\theta}) = -H(\hat{\theta})$ obtido numericamente)

Estimação por Intervalo

Objetivos de inferência e a função de verossimilhança



Estimação por Intervalo

Definição informal:

Região de valores do parâmetro com compatibilidade aceitável com os dados .

Definição do "ponto de corte" que define a região por:

a) evidência relativa em $LR(\theta)$:

b) comportamento assintótico $D(\theta) \sim \chi^2_p$:

c) interpretação probabilística direta em Inferência Bayesiana (quantis ou HPD)

Ou seja, evidência avaliada por:

- analogia sobre diferenças em $LR(\theta), l(\theta)$ ou $D(\theta)$
- referência probabilística

OBS: $LR(\theta)$ e $D(\theta)$ são adimensionais.

Relações entre critérios de corte

- $LR(\theta) \geq r$ na função de verossimilhança relativa
- $l(\hat{\theta}) - l(\theta) \geq c$ na função de log-verossimilhança
- $D(\theta) \leq c^*$ na função deviance

$$
\begin{aligned}
LR(\theta) &\geq r \\
l(\hat{\theta}) - l(\theta) &\leq -\log(r) = c \\
D(\theta) = -2[l(\theta) - l(\hat{\theta})] &\leq -2\log(r) = c^* \\
c^* = 2c = -2\log(r) &\longrightarrow r = \mathrm{e}^{-c} = \mathrm{e}^{-c/2}
\end{aligned}
$$

$$
\begin{aligned}
D(\theta) &\approx I_o(\hat{\theta})(\theta - \hat{\theta})^2 \sim \chi^2_{(1)} \\
\sqrt{D(\theta)} &\approx I_o^{1/2}(\hat{\theta})(\theta - \hat{\theta}) \sim \mathrm{N}(0,1)
\end{aligned}
$$

c^* é um quantil da $\chi^2_{(1)}$ ou, equivalentemente,
$\sqrt{c^*}$ é um quantil da $\mathrm{N}(0,1)$.

Relações entre critérios de corte

$$c^* = 2c = -2\log(r) \longrightarrow r = \mathrm{e}^{-c} = \mathrm{e}^{-c/2}$$

Relações entre intervalos baseados no corte na $LR[\theta]$ e por limites de probabilidade.

r	c	c^*	$z = \sqrt{c^*}$	$P[\lvert Z\rvert < \sqrt{c^*}]$
50%	0,693	1,386	1,177	0,761
26%	1.347	2.694	1,641	0,899
15%	1,897	3,794	1,948	0,949
14,6%	1,924	3,848	1,961	0,950
3,6%	3,324	6,648	2,578	0,990

analogia...

Limites do Intervalo $(\tilde{\theta}_I; \tilde{\theta}_S)$

1. Solução de equação (analítica ou numérica) de
   - $LR(\theta) = r$ ou,
   - $l(\hat{\theta}) - l(\theta) = c$ ou,
   - $D(\theta) = c^*$
2. Aproximação quadrática (Taylor)

$$D(\theta) = -2[l(\theta) - l(\hat{\theta})]$$

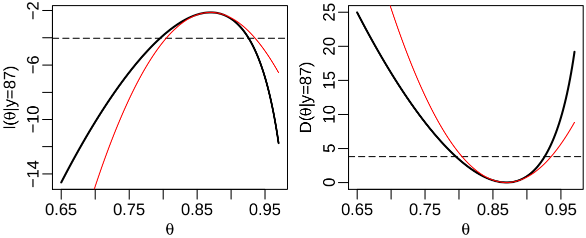
$$\tilde{D}(\theta) = -2\left\{[l(\hat{\theta}) + (\theta - \hat{\theta})l'(\hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})^2 l''(\hat{\theta})] - l(\hat{\theta})\right\}$$

$$= (\theta - \hat{\theta})^2 I_o(\hat{\theta})$$

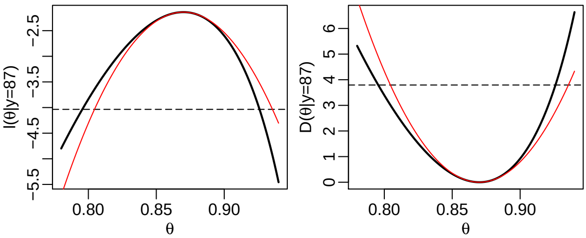
$$\tilde{D}(\theta) = c^* \longrightarrow \hat{\theta} \pm \sqrt{\frac{c^*}{I_o(\hat{\theta})}}$$

3. equivalente à distribuição assintótica: $\underline{\hat{\theta}} \sim NM_d(\underline{\theta}, I_E(\underline{\theta})^{-1})$

Aproximação quadrática (exemplo Binomial)



"Zoom" na região próxima do máximo



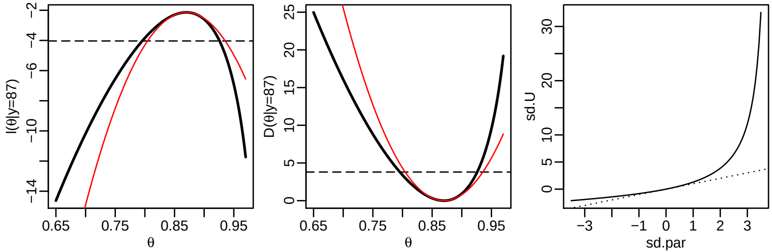
Avaliando a aproximação quadrática

Retomando a aproximação de Taylor (de 2ª ordem) ao redor the $\hat{\theta}$:

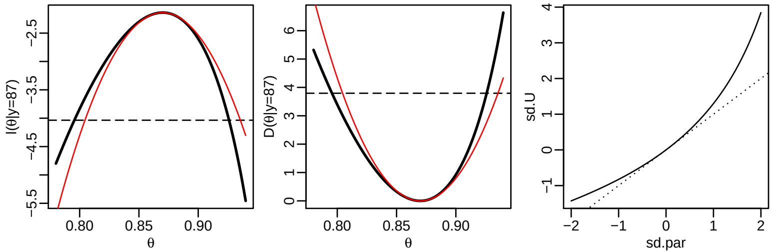
$$
\begin{aligned}
l(\theta) &= l(\hat{\theta}) - \frac{1}{2}(\theta - \hat{\theta})^2 I_o(\hat{\theta}) \\
\frac{\mathrm{dl}(\hat{\theta})}{\mathrm{d}\theta} &= U(\theta) = -I_o(\hat{\theta})(\theta - \hat{\theta}) \\
-I_o^{-1/2}(\hat{\theta})U(\theta) &= I_o^{-1/2}(\hat{\theta})(\theta - \hat{\theta})
\end{aligned}
$$

Quantidades na última expressão são adimensionais!!
Diagnóstico alternativo para aproximação quadrática.

Avaliação da aproximação quadrática



“Zoom” na região próxima do máximo



Exemplo: Exponencial (i.i.d.)

$$f(y_i, \theta) = \theta \exp\{-\theta y_i\} \;\; y > 0 \; ; \; \theta > 0$$

$$F(y_i, \theta) = 1 - \exp\{-\theta y_i\} \;\; y > 0 \; ; \; \theta > 0$$

$$L(\theta) = \theta^n \exp\{-\theta n \bar{y}\}$$

$$l(\theta) = n \log(\theta) - \theta n \bar{y}$$

$$U(\theta) = \frac{n}{\theta} - n\bar{y}$$

$$H(\theta) = -\frac{n}{\theta^2} \text{ (depende do valor de } \theta\text{!!)}$$

$$\hat{\theta} = 1/\bar{y}$$

Código R

Exemplo: Exponencial (cont)

Obtenção da Estimação Intervalar

1. Corte na deviance: (solução apenas numérica)

$$D(\theta) = 2n[\log\left(\hat{\theta}/\theta\right) + \overline{y}(\theta - \hat{\theta})] \leq c^*$$

2. Aproximação quadrática:

$$\tilde{D}(\theta) = n\left(\frac{\theta - \hat{\theta}}{\hat{\theta}}\right)^2$$

$$\left(\tilde{\theta}_I \approx \hat{\theta}(1 - \sqrt{c^*/n}) \ , \ \tilde{\theta}_S \approx \hat{\theta}(1 + \sqrt{c^*/n})\right)$$

3. Distribuição assintótica: $I_E^{-1}(\theta) \approx I_O^{-1}(\hat{\theta}) = \hat{\theta}^2/n$

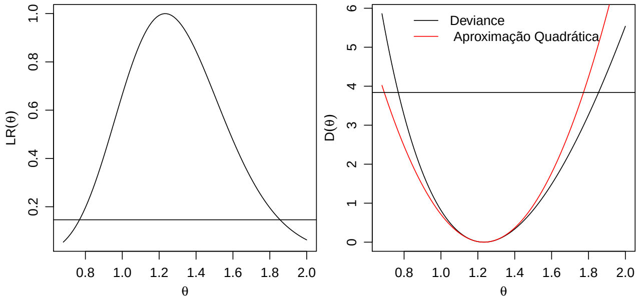
$$\hat{\theta} \pm z_{\frac{\alpha}{2}}\sqrt{V(\hat{\theta})}$$

$$\left(\tilde{\theta}_I = \hat{\theta} - z_{\frac{\alpha}{2}}\hat{\theta}/\sqrt{n} \ , \ \tilde{\theta}_S = \hat{\theta} + z_{\frac{\alpha}{2}}\hat{\theta}/\sqrt{n}\right)$$

Exemplo: Distribuição Exponencial (cont)

```
ICdevExp <- function(theta, theta.hat, y, nivel=0.95){
  n <- length(y)
  dv <- 2*n*( log( theta.hat/theta) + mean(y)*(theta- theta.hat))
  return(dv - qchisq(nivel,df=1))
}
```

```
rootSolve::uniroot.all(ICdevExp, interval=c(0,10),
                        theta.hat=1/mean(y), y=y)
```



Reparametrização

Caso univariado (ou reparametrização 1-1)

$$\phi = g(\theta)$$

Como fazer inferências sobre ϕ?

- estimação pontual
- estimação por intervalo
- testes de hipótese

Um resultado fundamental: invariância da verossimilhança:

$$l(\phi_i) = l(g(\theta_i)) = l(\theta_i)$$

Reparametrização

Duas alternativas iniciais:

1. Reescrever a função de verossimilhança e "recomeçar" do zero:
   - Pontual: $\hat{\phi} = \text{argmax}\{l(\phi)\}$
   - Intervalar:

$$(\hat{\phi}_I, \hat{\phi}_S) \text{ baseado em } l(\phi) \tag{1}$$

$$(\tilde{\phi}_I, \tilde{\phi}_S) \text{ baseado em } \tilde{l}(\phi) \tag{2}$$

2. "Aproveitar" resultados da inferência já obtida para θ:
   - Pontual: $\hat{\phi} = g(\hat{\theta})$ (invariância)
   - Intervalar:

$$(g(\hat{\theta}_I), g(\hat{\theta}_S)) \text{ transformação dos limites em } l(\theta) (\text{invariância}) \tag{3}$$

$$(g(\tilde{\theta}_I), g(\tilde{\theta}_S)) \text{ transformação dos limites em } \tilde{l}(\theta) \tag{4}$$

Estimativas pontuais: iguais em ambos casos

Estimativas intervalares: (1) = (3) e tem-se então três possibilidades

Reparametrização

- Resultados exatos baseados na verossimilhança:
  - $\hat{\phi} = g(\hat{\theta})$
  - IC por corte: $(\hat{\phi}_I, \hat{\phi}_S) = (g(\hat{\theta}_I), g(\hat{\theta}_S))$
- Aproximações quadráticas da verossimilhança
  - aproximação $\tilde{l}(\theta)$: $(g(\tilde{\theta}_I), g(\tilde{\theta}_S))$
  - aproximação $\tilde{l}(\phi)$: $(\tilde{\phi}_I, \tilde{\phi}_S)$
- Distribuição assintótica do estimador pelo Método *delta*:

$$\mathrm{Var}(\hat{\phi}) = [g'(\theta)]^2[I_E(\theta)]^{-1} \longrightarrow \left[\mathrm{se}(\hat{\phi}) = |g'(\theta)|[I_E(\theta)]^{-1/2}\right]$$

Assintoticamente: $\hat{\phi} = g(\hat{\theta}) \sim N(\phi, [g'(\theta)]^2[I_E(\theta)]^{-1})$

Método delta

$$\phi = g(\theta)$$

Teorema

(Método delta). Seja $\hat{\theta}$ um estimador de θ para uma amostra de tamanho n tal que

$$\sqrt{n}(\hat{\theta} - \theta) \to \mathrm{N}(0, \sigma^2_\theta).$$

Então, para qualquer função $g(\cdot)$ que é diferenciável ao redor de θ e $h'(\theta) \neq 0$, tem-se que

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \to \mathrm{N}(0, \sigma^2_\theta |h'(\theta)|^2).$$

Em temos do TCL, aplica-se a funções da média amostral.

Produz uma aproximação quadrática da vesossimilhança de ϕ.

Reparametrização

- ideal: Resultados exatos baseados na verossimilhança:
- muito utilizado: Resultados baseados no método delta
- conveniente: aproximações baseadas nas aproximações quadráticas das verossimilhanças $\tilde{l}(\theta)$ ou $\tilde{l}(\phi)$
- Se transformação $g(\cdot)$ é não linear, a invariância não é válida para a aproximação quadrática, ou seja:

$$\{g(\tilde{\theta}_I), g(\tilde{\theta}_S)\} = \{g(\hat{\theta} - z_{\alpha/2}[I_E(\hat{\theta})]^{-1/2}), g(\hat{\theta} + z_{\alpha/2}[I_E(\hat{\theta})]^{-1/2})\} \neq \{g(\hat{\theta}) - z_{\alpha/2}|g'(\theta)|[I_E(\hat{\theta})]^{-1/2}, g(\hat{\theta}) + z_{\alpha/2}|g'(\theta)|[I_E(\hat{\theta})]^{-1/2}\} = (\tilde{\phi}_I, \tilde{\phi}_S)$$

- se $l(\theta)$ é menos assimétrica: usar $(g(\tilde{\theta}_I), g(\tilde{\theta}_S))$
  se $l(\phi)$ é menos assimétrica: usar $(\tilde{\phi}_I, \tilde{\phi}_S)$

Recomendações

- Melhor abordagem: (mais geral e acurácia)
  IC's baseados verossimilhança/deviance (muitas vezes só obtidos numericamente)
- Intervalos assintóticos (utilizam $se(\hat{\theta})$, obtenção a partir da aproximação quadrática, formas fechadas)
- Escolher parametrização da função que forneça uma boa aproximação quadrática
- IC's para funções dos parâmetros: obtenção pelo método delta ou direta se aproximadamente quadrática

Reparametrização

Por que reparametrizar (ou considerar resultados ra reparametrização)?

- Interpretabilidade de ϕ
- "Melhor"formato da função de verossimilhança com melhor comportamento de métodos numéricos.
- Predição!!
  uma predição de um modelo pode ser vista como uma reparametrização!

Exemplo: Exponencial (cont)

Reparametrização

$$\phi = P[Y \leq u] = 1 - \exp\{-\theta u\}$$

- Obter $se(\hat{\phi})$
- Três intervalos possíveis:

$$(\hat{\phi}_I, \hat{\phi}_S) : (g(\hat{\theta}_I), g(\hat{\theta}_S))$$

$$(\tilde{\phi}_I, \tilde{\phi}_S) : \hat{\phi} \pm z_{\alpha/2} se(\hat{\phi})$$

$$(1 - \exp\{-\tilde{\theta}_S u\}, 1 - \exp\{-\tilde{\theta}_S u\}) : (g(\tilde{\theta}_I), g(\tilde{\theta}_S))$$

- Comparação gráfica das funções e das taxas de cobertura (simulação)

Exemplo: Distribuição Exponencial (cont)

Redefinindo

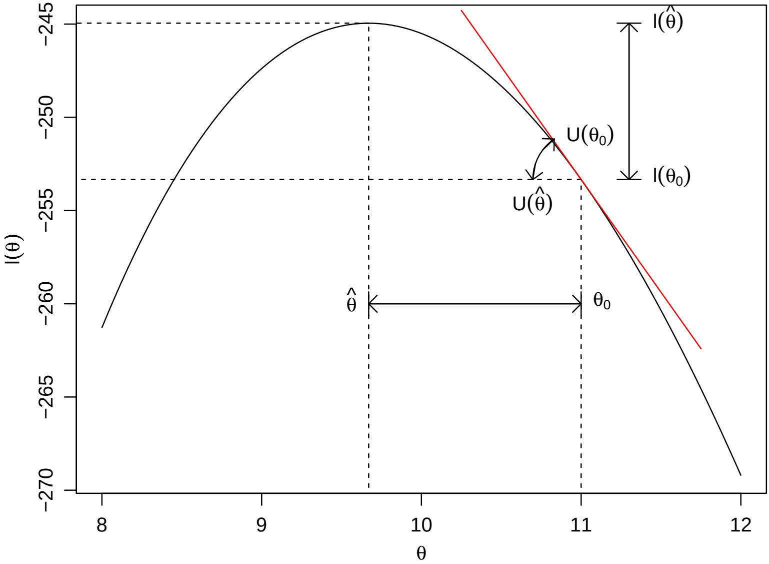
```
ICdevExp <- function(theta, theta.hat, y, nivel=0.95){
  n <- length(y)
  dv <- 2*n*( log( theta.hat/theta) + mean(y)*(theta- theta.hat))
  return(dv - qchisq(nivel,df=1))
}
```

```
require(rootSolve)
uniroot.all(ICdevExp, interval=c(0,10), theta.hat=1/mean(y), y=y)
```

```
ICdevExp <- function(theta, amostra, nivel=0.95){
  ## amostra é um vetor com elementos n e mean(y), nesta ordem
  n <- amostra[1]
  med <- amostra[2]
  dv <- 2*n*(-log(med*theta) + med*theta - 1)
  return(dv - qchisq(nivel, df=1))
}
```

```
am <- c(length(y), mean(y))
uniroot.all(ICdevExp, interval=c(0,10), amostra=am)
```

Teste de Hipótese



Teste de Hipótese

- Teste razão de verossimilhança

```
trv <- function(Est, H0, alpha, ...){
  critico <- qchisq(1-alpha, df=1)
  est.calc <- Est(H0, ...)
  print(ifelse(est.calc < critico, "Aceita H0", "Rejeita H0"))
  return(c(est.calc,critico))}
```

- Teste Wald

```
wald <- function(H0, EMV, V.EMV, alpha){
  critico <- qnorm(1-alpha/2)
  Tw <- (EMV - H0)/sqrt(V.EMV)
  print(ifelse(Tw < critico, "Aceita H0", "Rejeita H0"))
  return(c(Tw,critico))
}
```

- Teste Escore

```
escore <- function(H0, U, Ie, alpha, ...){
  critico <- qnorm(1-alpha/2)
  Te <- U(H0,...)/sqrt(Ie(H0,...))
  print(ifelse(Te < critico, "Aceita H0", "Rejeita H0"))
  return(c(Te,critico))
```

Exemplo: Poisson

TRV

```
Est <- function(H0, x){
   n <- length(x)
   EMV <- mean(x)
   lv <- 2*n*(( H0 - EMV) + EMV*log(EMV/H0))
 return(lv)
}
trv(Est = Est, H0=8, alpha = 0.05, x=x)
```

Wald

```
 wald(H0=8, EMV = mean(x), V.EMV = mean(x)/length(x),alpha=0.05)
```

Escore

```
fc.escore <- function(lambda,x){
   n <- length(x)
   esco <- -n + sum(x)/lambda
   return(esco)}
Ie <- function(lambda,x){
   n <- length(x)
   I <- n/lambda
   return(I)}
escore(H0 = 8, U = fc.escore, Ie = Ie, alpha=0.05, x=x)
```

Condições de regularidade

- Θ é finito dimensional e θ é interior a Θ
- primeiras três derivadas de $l(\theta)$ na vizinhança de θ
- amplitude não depende de θ
- $l(\theta) \approx$ quadrática para $n \to \infty$, passando a depender apenas da posição e curvatura no EMV
- $\ldots I_E(\theta)$ precisa ser inversível

Exemplo: Distribuição Normal

log-Verossimilhança para $\theta = (\mu, \sigma)$

$$l(\mu,\sigma) = -\frac{n}{2}\log 2\pi - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2.$$

Escore

$$U(\mu) = \frac{\partial l(\mu,\sigma)}{\partial \mu} = \frac{\sum_{i=1}^{n} y_i}{\sigma^2} - \frac{n\mu}{\sigma^2}$$

$$U(\sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(y_i - \mu)^2.$$

EMV

$$\hat{\mu} = \frac{\sum_{i=1}^{n} y_i}{n} \quad \text{e} \quad \hat{\sigma^2} = \frac{\sum_{i=1}^{n}(y_i - \mu)^2}{n}.$$

Informação

$$I_O(\hat{\mu},\hat{\sigma}) = \begin{bmatrix} \frac{n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{2n}{\hat{\sigma}^2} \end{bmatrix}.$$

Intervalos de confiança

Conjuntos

- corte

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

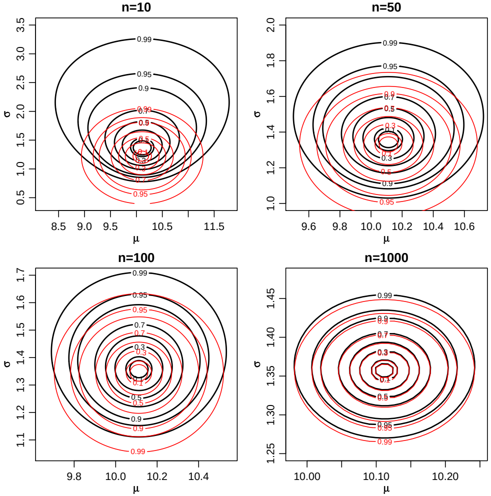
- elipse (aproximação quadrática)

$$D(\mu,\sigma) \approx (\underline{\theta} - \underline{\hat{\theta}})^\top I_o(\underline{\hat{\theta}})(\underline{\theta} - \underline{\hat{\theta}}).$$

- assintótico

$$\begin{bmatrix}\hat{\mu}\\ \hat{\sigma}\end{bmatrix} \sim NM_2\left(\begin{bmatrix}\mu\\ \sigma\end{bmatrix}, \begin{bmatrix}\hat{\sigma}^2/n & 0\\ 0 & \hat{\sigma}^2/2n\end{bmatrix}\right)$$

Exemplo: Distribuição Normal (cont)



Intervalos de confiança

Parâmetros de interesse e de inconveniência (*nuisance*): (θ, ψ)

Soluções usuais:

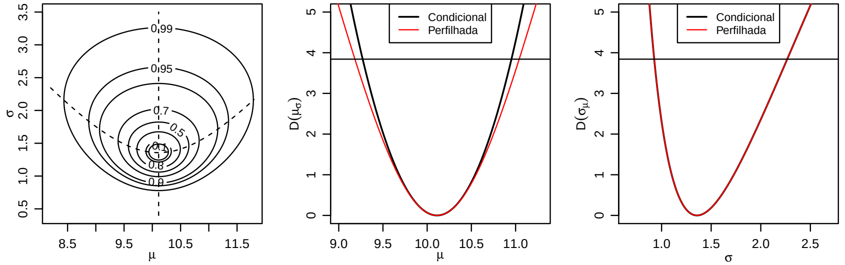
- Condicionando no EMV : $L(\theta) = L(\theta, \hat{\psi}) \equiv [Y|\theta, \hat{\psi}]$
- Verossimilhança Perfilhada : $L(\theta) \equiv L[\theta, \hat{\psi}_\theta]$
- Verossimilhanças marginais integradas Bayesianas : $L(\theta) = \int [Y|\theta, \psi][\psi] d\,\psi$

  Exemplo Normal: $1/\sigma^2 \sim G(a, b)$

$$f(y|\mu) = \frac{\Gamma(n/2+1)}{\pi^{n/2}\Gamma(a)(\sum_i (y_i - \mu)^2 + 2b)^{n/2+a}}$$

Integrações analíticas e por simulação

Exemplo: Distribuição Normal (cont)



Exemplo: Distribuição Normal (cont)

```
pl.mu <- function(sigma, mu, dados){
    pll <- sum(dnorm(dados, mean=mu, sd=sigma, log=TRUE))
    return(pll)}
##
pl.sigma <- function(mu, sigma, dados){
    pll <- sum(dnorm(dados, mean=mu, sd=sigma, log=TRUE))
    return(pll)}
```

```
grid.mu <- seq(9, 11.3, length=200)
grid.sigma <- seq(0.65, 2.7, length=200)
## Condicionais:
mu.cond <- sapply(grid.mu, pl.sigma, sigma=sqrt(var(y10)*9/10), dados=y10)
sigma.cond <- sapply(grid.sigma, pl.mu, mu=mean(y10), dados=y10)

mu.perf <- matrix(0, nrow=length(mu), ncol=2)
for(i in 1:length(mu)){
mu.perf[i,] <- unlist(optimize(pl.mu,c(0,200),
                      mu=mu[i],dados=y10,maximum=TRUE))}
sigma.perf <- matrix(0, nrow=length(sigma), ncol=2)
for(i in 1:length(sigma)){
sigma.perf[i,] <- unlist(optimize(pl.sigma,c(0,1000),
                sigma=sigma[i],dados=y10,maximum=TRUE))}
```

Exemplo: Distribuição Normal (Dados intervalares)

Dados intervalares e parâmetros $\theta = (\mu, \sigma)$

observações "pontuais":

`72,6  81,3  72,4  86,4  79,2  76,7  81,3` ;

observações intervalares:

uma observação com valor acima de 85,
uma observação com valor acima de 80,
quatro observações com valores entre 75 e 80,
seis observações com valores abaixo de 75.

Contribuições para verossimilhança

$$L(\theta) = f(y_i) \text{ para } y_i \text{ pontual,}$$
$$L(\theta) = 1 - F(85) \text{ para } y_i > 85,$$
$$L(\theta) = 1 - F(80) \text{ para } y_i > 80,$$
$$L(\theta) = F(80) - F(75) \text{ para } 75 < y_i < 80,$$
$$L(\theta) = F(75) \text{ para } y_i < 75.$$

Exemplo: Distribuição Normal (Dados intervalares)

Expressão da verossimilhança no exemplo:

$$L(\theta) = \left(\prod_{1=1}^{7} f(y_i)\right) \cdot (1 - F(85)) \cdot (1 - F(80)) \cdot (F(80) - F(75))^4 \cdot (F(75))^6$$

De forma mais geral para n_p dados pontuais e n_I dados intervalares com valores entre a_i e b_i:

$$\begin{aligned} L(\theta) &= \prod_{1=1}^{n_p} f(y_i) \cdot \prod_{1=1}^{n_I} (F(b_i) - F(a_i)) \\ &= \prod_{1=1}^{n_p} \phi(\frac{y_i - \mu}{\sigma}) \cdot \prod_{1=1}^{n_I} (\Phi(\frac{a_i - \mu}{\sigma}) - \Phi(\frac{b_i - \mu}{\sigma})) \end{aligned}$$

No exemplo:

a	85	80	75	75	75	75	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
b	∞	∞	80	80	80	80	75	75	75	75	75	75

Dados intervalares (cont)

$$l(\theta) = \sum_{1=1}^{n_p} \log\left(\phi(\frac{y_i-\mu}{\sigma})\right) + \sum_{1=1}^{n_I} \log\left((\Phi(\frac{a_i-\mu}{\sigma}) - \Phi(\frac{b_i-\mu}{\sigma}))\right)$$

```
nllnormI <- function(par, xp, XI) {
    ll1 <- sum(dnorm(xp, mean = par[1], sd = par[2], log = T))
    L2 <- pnorm(XI, mean = par[1], sd = par[2])
    ll2 <- sum(log(L2[, 2] - L2[, 1]))
    return(-(ll1 + ll2))
}
```

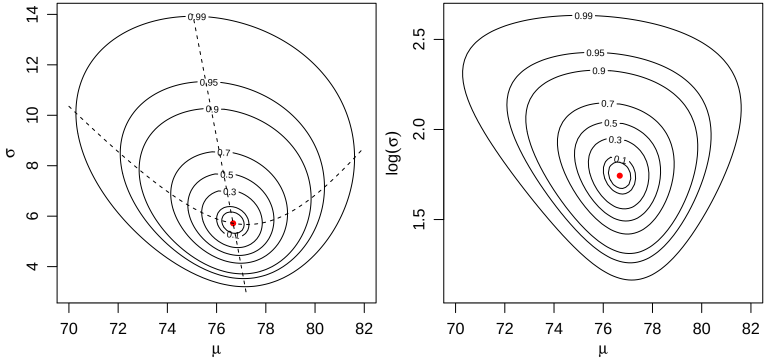
```
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
[1,]   85   80   75   75   75   75 -Inf -Inf -Inf  -Inf  -Inf  -Inf
[2,]  Inf  Inf   80   80   80   80   75   75   75    75    75    75
```

```
ini <- c(mean(y), sd(y))
ests <- optim(, nllnormI, x=y, XI=yI)$par
```

Dados intervalares (cont)

Função deviance genérica.

```
devFun <- function(theta, est, llFUN, ...){
  return(2 * (llFUN(theta, ...) - llFUN(est, ...)))
}
devSurf <- Vectorize(function(x,y, ...) devFun(c(x,y), ...))
```



Dados intervalares (cont)

Código mais geral e cuidadoso

```
nllnormI <- function(par, xp, XI, logsigma=FALSE){
    if(logsigma) par[2] <- exp(par[2])
        ll1 <- ifelse(missing(xp), 0,
                 sum(dnorm(xp, mean=par[1], sd=par[2], log=T)))
    if(missing(XI)) ll2 <- 0
    else{
      if(ncol(XI) != 2 || any(XI[,2] <= XI[,1]))
       stop("XI deve ser matrix com 2 colunas com XI[,2] > XI[,2]"
      L2 <- pnorm(XI, mean=par[1], sd=par[2])
            ll2 <- sum(log(L2[,2] - L2[,1]))
  }
  return(-(ll1 + ll2))
}
```

Outros exemplos (texto)

1. AR1
2. Outro exemplo de reparametrização
3. Gamma
4. Binomial Negativa
5. Processo de Poisson não-homogêneo
6. Modelo espacial Geoestatístico
7. Códigos Genéricos:
    - `mle` (**stat4**) e `mle2` (**bbmle**)
    - `profile` e `confint`