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

Especialização DSBD, 8, 15 e 22/06/2018

Comentários

- Texto de referência:
  - 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
  - Experiências/exposição das gerações
  - 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

Visitando um exemplo simples:

- População: $X \sim B(\theta)$

Exemplo introdutório

Visitando um exemplo simples:

- População: $X \sim B(\theta)$
- Amostra: $x_1, \ldots, x_n$

Exemplo introdutório

Visitando um exemplo simples:

- População: $X \sim B(\theta)$
- Amostra: $x_1, \ldots, x_n$
- O que podemos falar sobre θ?
  - Qual a informação contida na amostra?
  - Consideram-se outras fontes de informação?

Laboratório de Estatística e Geoinformação

Exemplo introdutório

Visitando um exemplo simples:

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

Laboratório de Estatística e Geoinformação

Espaço do Modelo

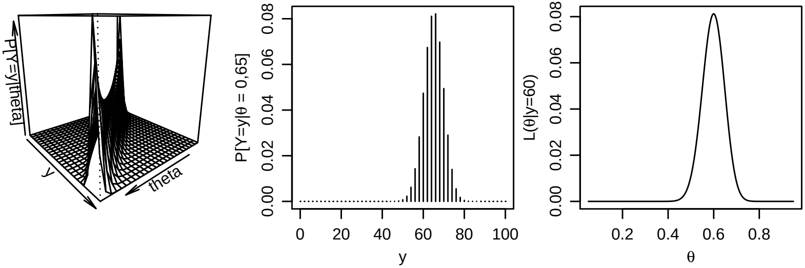
- O espaço definido pelo modelo (3-D)

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

Espaço do Modelo

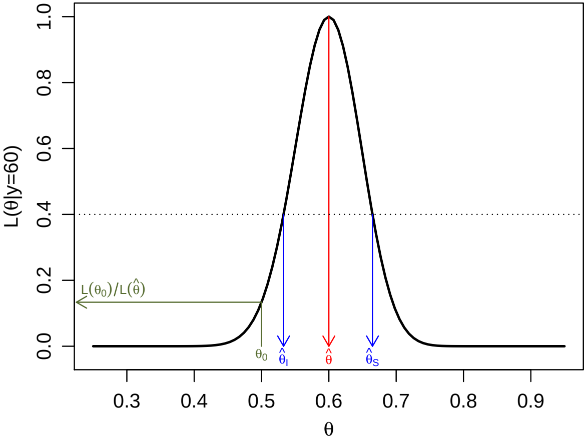
- O espaço definido pelo modelo (3-D)

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



Objetivos de Inferência

Objetivos de inferência e a função de verossimilhança



Elementos

- função de verossimilhança
  probabilidade da amostra obtida para diferentes valores de θ

Elementos

- função de verossimilhança
  probabilidade da amostra obtida para diferentes valores de θ
- melhor estimador

Laboratório de Estatística e Geoinformação

Elementos

- função de verossimilhança
  probabilidade da amostra obtida para diferentes valores de θ
- melhor estimador
- conjunto de valores razoavelmente compatíveis com a amostra

Elementos

- função de verossimilhança
  probabilidade da amostra obtida para diferentes valores de θ
- melhor estimador
- conjunto de valores razoavelmente compatíveis com a amostra
- decidir entre dois valores o mais compatível com a amostra

Elementos

- função de verossimilhança
  probabilidade da amostra obtida para diferentes valores de θ
- melhor estimador
- 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?

Elementos

- função de verossimilhança
  probabilidade da amostra obtida para diferentes valores de θ
- melhor estimador
- 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?
- suposições/pressupostos

Laboratório de Estatística e Geoinformação

Elementos

- função de verossimilhança
  probabilidade da amostra obtida para diferentes valores de θ
- melhor estimador
- 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?
- suposições/pressupostos
- relações e contrastes com outros métodos

Laboratório de Estatística e Geoinformação

Paradigmas para inferência

Diferentes Paradigmas no tratamento da incerteza

- Verossimilhança
- Frequentista (espaço amostral / Newman-Pearson)
- Bayesiano

Laboratório de Estatística e Geoinformação

Inferência Frequentista

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

Inferência 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.

Inferência 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`.

Laboratório de Estatística e Geoinformação

Inferência 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.

Inferência 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.

Inferência 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).

Inferência Bayesiana

- Extensão da definição do modelo:

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

- permite obter (via teorema de Bayes)

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

- 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).

Inferência Bayesiana

No contexto do exemplo de estimação de proporção:

- Priori: $[\theta] \sim \text{Beta}(a, b)$ (distribuição Beta)
- Verossimilhança: $[Y|\theta] \sim \text{Bin}(n, \theta)$ (em θ, é proporcional à distribuição Beta)
- Posteriori: $[\theta|y] \sim \text{Beta}(a + y, b + n)$ (distribuição Beta - conjugada)

(detalhes apresentados durante a aula)

Laboratório de Estatística e Geoinformação

Paradigmas para inferência

- Paradigmas . . .
  - Frequentista (espaço amostral / Newman-Pearson)
  - Verossimilhança
  - Bayesiano

Laboratório de Estatística e Geoinformação

Paradigmas para inferência

- Paradigmas ...
  - Frequentista (espaço amostral / Newman-Pearson)
  - Verossimilhança
  - Bayesiano
- ... respondem a diferentes perguntas:

Laboratório de Estatística e Geoinformação

Paradigmas para inferência

- Paradigmas . . .
  - Frequentista (espaço amostral / Newman-Pearson)
  - Verossimilhança
  - Bayesiano
- . . . respondem a diferentes perguntas:
  - O que devo fazer?

Laboratório de Estatística e Geoinformação

Paradigmas para inferência

- Paradigmas . . .
  - Frequentista (espaço amostral / Newman-Pearson)
  - Verossimilhança
  - Bayesiano
- . . . respondem a diferentes perguntas:
  - O que devo fazer?
  - O que os dados dizem?

Laboratório de Estatística e Geoinformação

Paradigmas para inferência

- Paradigmas . . .
  - Frequentista (espaço amostral / Newman-Pearson)
  - Verossimilhança
  - Bayesiano
- . . . respondem a diferentes perguntas:
  - O que devo fazer?
  - O que os dados dizem?
  - Em que devo acreditar?

Paradigmas para inferência

- Paradigmas ...
  - Frequentista (espaço amostral / Newman-Pearson)
  - Verossimilhança
  - Bayesiano
- ... respondem a diferentes perguntas:
  - O que devo fazer?
  - O que os dados dizem?
  - Em que devo acreditar?
- Uma provocação à reflexão (**Royall, 1997:**)
  "Fortunatelly we are not forced to choose either of these two evils, the sample-space dependence of the frequentists or the prior distribution of the Bayesians.
  Likelihood methods avoid both sources of subjectivity. *"*

Laboratório de Estatística e Geoinformação

Função de Verossimilhança

Definição informal:

Dada pela expressão da distribuição conjunta de todas as v.a.'s observáveis (Y) o modelo. ou *Distribuição conjunta de todas as v.a.'s especificadas no modelo, integrada sobre as não observáveis.*

Notação: $[\cdot]$ (função de) distribuição de $\cdot$

Casos de particular interesse:

- Distribuições e modelos de regressão

$$[Y|\theta]$$

Função de Verossimilhança

Definição informal:

Dada pela expressão da distribuição conjunta de todas as v.a.'s observáveis (Y) o modelo. ou *Distribuição conjunta de todas as v.a.'s especificadas no modelo, integrada sobre as não observáveis.*

Notação: $[\cdot]$ (função de) distribuição de $\cdot$

Casos de particular interesse:

- Distribuições e modelos de regressão

$$[Y|\theta]$$

- Modelos hierárquicos (mistos, efeitos aleatórios, longitudinais, espaciais, etc)

$$[Y|\theta] = \int [Y, b|\theta] d\, b = \int [Y|b, \theta][b|\theta] d\, b$$

Função de Verossimilhança

Definição informal:

Dada pela expressão da distribuição conjunta de todas as v.a.'s observáveis (Y) o modelo. ou *Distribuição conjunta de todas as v.a.'s especificadas no modelo, integrada sobre as não observáveis.*

Notação: $[\cdot]$ (função de) distribuição de $\cdot$

Casos de particular interesse:

- Distribuições e modelos de regressão

$$[Y|\theta]$$

- Modelos hierárquicos (mistos, efeitos aleatórios, longitudinais, espaciais, etc)

$$[Y|\theta] = \int [Y, b|\theta] d\, b = \int [Y|b, \theta][b|\theta] d\, b$$

- inclui Inf. Bayesiana com especificação de $[\theta]$

Expressão da Verossimilhança I

V.A. observável discreta (não há ambiguidade)

$$L(\theta) \equiv P_\theta[Y = y]$$

Exemplo: $Y \sim P(\lambda)$

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

$$\propto \exp\{-n\lambda\}\lambda^{\sum_{i=1}^{n} Y_i}$$

Laboratório de Estatística e Geoinformação

Expressão da Verossimilhança II

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

- Forma mais geral

$$L(\theta) = P_\theta[y_{1I} \leq y_1 \leq y_{1S}, \ldots, y_{nI} \leq y_1 \leq y_{nS}]$$

Laboratório de Estatística e Geoinformação

Expressão da Verossimilhança II

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

- Forma mais geral

$$L(\theta) = P_\theta[y_{1I} \leq y_1 \leq y_{1S}, \ldots, y_{nI} \leq y_1 \leq y_{nS}]$$

- Sob independência

$$\begin{aligned} L(\theta) &= P_\theta[y_{1I} \leq y_1 \leq y_{1S}] \cdot P_\theta[y_{2I} \leq y_2 \leq y_{2S}] \ldots P_\theta[y_{nI} \leq y_n \leq y_{nS}] \\ &= \prod_{i=1}^{n} P_\theta[y_{iI} \leq y_i \leq y_{iS}] \end{aligned}$$

Expressão da Verossimilhança II

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

- Forma mais geral

$$L(\theta) = P_\theta[y_{1I} \leq y_1 \leq y_{1S}, \ldots, y_{nI} \leq y_1 \leq y_{nS}]$$

- 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, $(y_i - \delta/2 \leq Y_i \leq y_i + \delta/2)$;

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

Expressão da Verossimilhança II (cont)

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

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

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

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

Laboratório de Estatística e Geoinformação

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:

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

Laboratório de Estatística e Geoinformação

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 $x_{(5)} = 3.5$

Laboratório de Estatística e Geoinformação

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:

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

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

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

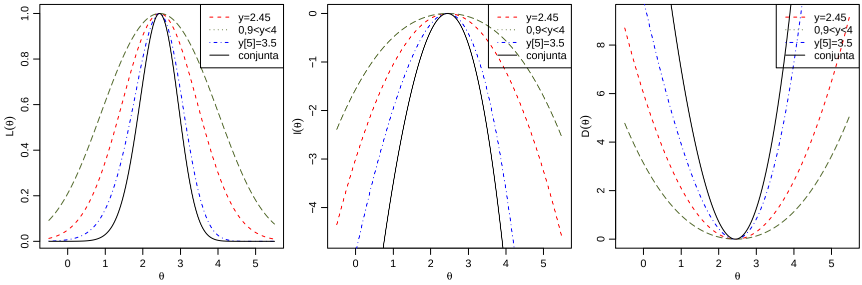
$$L_3 = L(\theta; x_{(5)} = 3.5) = n\{\Phi(x_{(n)} - \theta)\}^{n-1}\phi(x_{(n)} - \theta).$$

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)
```



Formas alternativas

- Verossimilhança:

$$L(\theta)$$

- Verossimilhança Relativa:

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

- log-Verossimilhança:

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

- *Deviance*:

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

EMV: $\hat{\theta} = \sup_{\Theta} L[\theta]$
Quando possível obtido por: $\hat{\theta} = \max_{\Theta} l[\theta]$

Laboratório de Estatística e Geoinformação

Exemplo: distribuição Poisson

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

Exemplo: Distribuição Poisson

1. Código 1

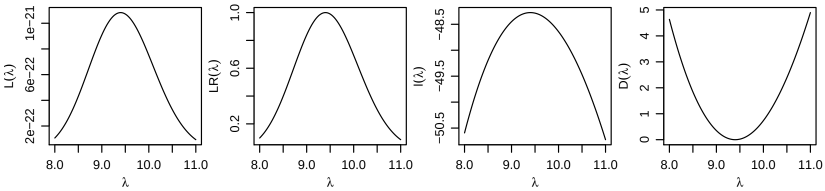
```
veroPois <- function(par, dados, tipo, maxlogL){
   tipo = match.arg(tipo, choices=c("L","LR","logL","dev"))
   ll <- sapply(par, function(p) sum(dpois(dados, lambda=p,
                                           log=TRUE)))
   return(switch(tipo, "L" = exp(ll),
                       "LR" = exp(ll-maxlogL),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

```
mll <- veroPois(emv, amostra=am, tipo="logL")
...
curve(veroPois(x, amostra=am, tipo="dev", maxlogL=mll), 8, 11,
      ylab=expression(D(lambda)), xlab=expression(lambda))
```



Funções de Interesse

- Função escore: $U(\theta) = l'(\theta)$
- Hessiano/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 $I_E(\hat{\theta})$

Laboratório de Estatística e Geoinformação

Funções de Interesse

- Função escore: $U(\theta) = l'(\theta)$
- Hessiano/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:

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

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

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

Laboratório de Estatística e Geoinformação

Exemplo: Distribuição Poisson (cont)

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

$$l(\lambda) = -n\lambda + (\sum_{i=1}^{n} Y_i)\log(\lambda) - \sum_{i=1}^{n} \log Y_i!$$

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

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

$$I_O(\lambda) = \frac{\sum_{i=1}^{n} Y_i}{\lambda^2} \; ; \; I_E(\lambda) = \frac{n}{\lambda} \; ; \; I_O(\hat{\lambda}) = I_E(\hat{\lambda}) = \frac{n^2}{\sum_{i=1}^{n} Y_i}$$

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

Laboratório de Estatística e Geoinformação

Exemplo: Distribuição Poisson (cont)

Função escore:

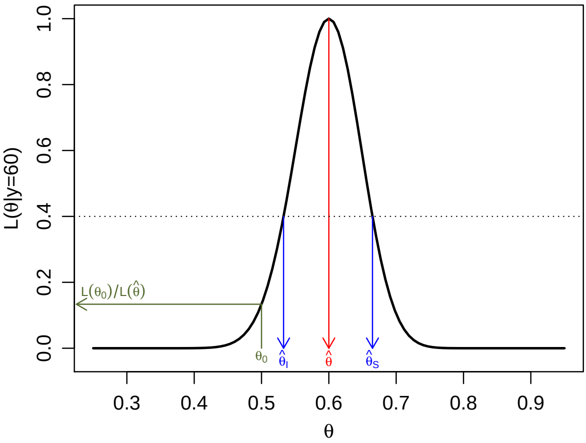
```
UPois <- function(lambda, amostra){
  return(with(amostra, n - soma/lambda))
}
```

Hessiano (informação estimada):

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

Estimação

Objetivos de inferência e a função de verossimilhança



Obtendo o EMV

- 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)

Laboratório de Estatística e Geoinformação

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 1^a ordem de $U(\theta)$:

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

```
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*: $-H(\lambda) \longrightarrow I_E(\lambda)$

Laboratório de Estatística e Geoinformação

Exemplo: Distribuição Poisson (cont)

- Solução de equação $U(\theta) = 0$:

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

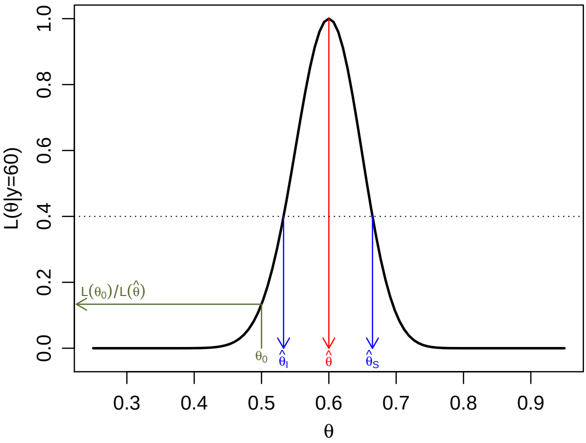
- Maximização da verossimilhança

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

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

Intervalo de Confiança

Objetivos de inferência e a função de verossimilhança



Intervalo de confiança

- Região definida por corte na função de verossimilhança
  - $LR(\theta) \geq r$
  - $l(\theta) \geq c$
  - $D(\theta) \leq c^*$

Laboratório de Estatística e Geoinformação

Intervalo de confiança

- Região definida por corte na função de verossimilhança
  - $LR(\theta) \geq r$
  - $l(\theta) \geq c$
  - $D(\theta) \leq c^*$
- Definição do ponto de corte
  a) comportamento assintótico $D(\theta) \sim \chi^2_p$
  b) interpretação probabilística direta em inf. Bayesiana (quantis ou HPD)
  c) interpretação de evidência relativa em $LR(\theta)$

Intervalo de confiança

- Região definida por corte na função de verossimilhança
  - $LR(\theta) \geq r$
  - $l(\theta) \geq c$
  - $D(\theta) \leq c^*$
- Definição do ponto de corte
  a) comportamento assintótico $D(\theta) \sim \chi^2_p$
  b) interpretação probabilística direta em inf. Bayesiana (quantis ou HPD)
  c) interpretação de evidência relativa em $LR(\theta)$
- Relações entre (a) e (c)

r	c	c^*	$P[\lvert Z \rvert < \sqrt{c^*}]$
50%	0,693	1,386	0,761
26%	1.661	3,321	0,899
15%	1,897	3,794	0,942
3,6%	3,324	6,648	0,990

- c.1) evidência por analogia sobre diferenças em $LR(\theta), l(\theta) ou D(\theta)$

analogia...

Limites do Intervalo

1. Solução de equação (analítica ou numérica)
   - $LR(\theta) = r$
   - $l(\theta) = c$
   - $D(\theta) = c^*$

Limites do Intervalo

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

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

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

$$D(\theta) = -(\theta - \hat{\theta})^2 l''(\hat{\theta}) \leq c^*$$

$$\tilde{\theta} = \hat{\theta} \pm \sqrt{\frac{c^*}{-l''(\hat{\theta})}}$$

Laboratório de Estatística e Geoinformação

Limites do Intervalo

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

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

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

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)

Intervalos de confiança

1. Corte na deviance: (solução apenas numericamente)

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

Exemplo: Exponencial (cont)

Intervalos de confiança

1. Corte na deviance: (solução apenas numericamente)

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

2. Aproximação quadrática:

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

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

Exemplo: Exponencial (cont)

Intervalos de confiança

1. Corte na deviance: (solução apenas numericamente)

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

2. Aproximação quadrática:

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

$$\left(\hat{\theta}_L \approx \hat{\theta}(1 - \sqrt{c^*/n}) \;\; , \;\; \hat{\theta}_U \approx \hat{\theta}(1 + \sqrt{c^*/n})\right)$$

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

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

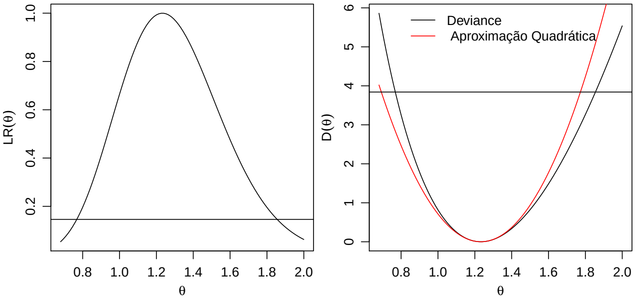
$$\hat{\theta}_L = \hat{\theta} - z_{\alpha}\hat{\theta}/\sqrt{n} \quad \text{e} \quad \hat{\theta}_U = \hat{\theta} + z_{\alpha}\hat{\theta}/\sqrt{n}$$

Laboratório de Estatística e Geoinformação

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))
}
```

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



Reparametrização

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

- $\hat{\phi} = g(\hat{\theta})$
- IC por corte: $(\hat{\phi}_I, \hat{\phi}_S) = (g(\hat{\theta}_I), g(\hat{\theta}_S))$
- Assintóticamente: $\hat{\phi} = g(\hat{\theta}) \sim N(\phi, [g'(\theta)]^2 [I_E(\theta)]^{-1})$
- Método *delta*:

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

- Se transformação $g(\cdot)$ é não linear, invariância não é válida para aproximação quadrática

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

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

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

Laboratório de Estatística e Geoinformação

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

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

Laboratório de Estatística e Geoinformação

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))
}
```

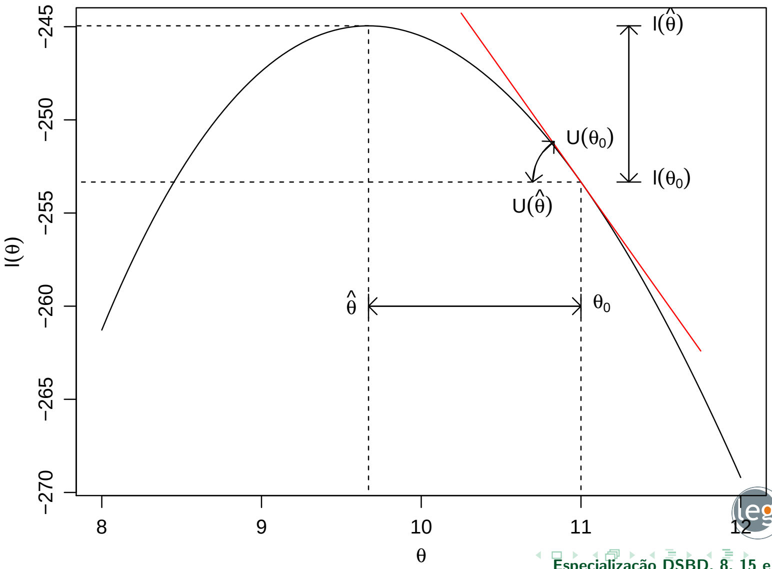
```
require(rootSolve)
uniroot.all(ICdevExp, interval=c(0,10), theta.hat=1/mean(y), y=y)
```

```
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)
```

Laboratório de Estatística e Geoinformação

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)
```

Laboratório de Estatística e Geoinformação

Exemplo: Distribuição Normal

log-Verossimilhança

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

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)

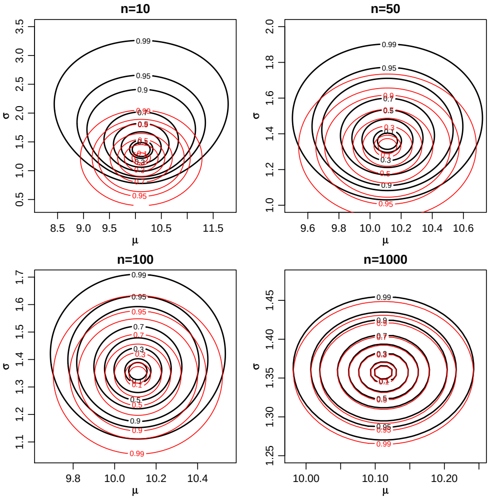
$$D(\mu, \sigma) \approx (\underline{\theta} - \underline{\hat{\theta}})^\top I_o(\underline{\hat{\theta}})(\underline{\theta} - \underline{\hat{\theta}}).$$

- assintótico

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

Laboratório de Estatística e Geoinformação

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 : $L(\theta) = \int [Y|\theta, \psi][\psi] d\,\psi$

Laboratório de Estatística e Geoinformação

Intervalos de confiança

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

Soluções usuais:

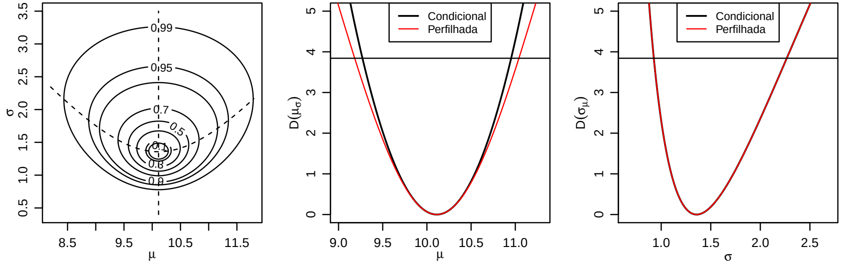
- Condicionando no EMV : $L(\theta) = L(\theta, \hat{\psi}) \equiv [Y|\theta, \hat{\psi}]$
- Verossimilhança Perfilhada : $L(\theta) \equiv L[\theta, \hat{\psi}_\theta]$
- Verossimilhanças marginais integradas Bayesianas : $L(\theta) = \int [Y|\theta, \psi][\psi] d\,\psi$

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:

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

$$L(\underline{\theta}) = f(y_i) \text{ para } y_i \text{ pontual},$$

$$L(\underline{\theta}) = 1 - F(85) \text{ para } y_i > 85,$$

$$L(\underline{\theta}) = 1 - F(80) \text{ para } y_i > 80,$$

$$L(\underline{\theta}) = F(80) - F(75) \text{ para } 75 < y_i < 80,$$

$$L(\underline{\theta}) = F(75) \text{ para } y_i < 85.$$

Laboratório de Estatística e Geoinformação

Dados intervalares (cont)

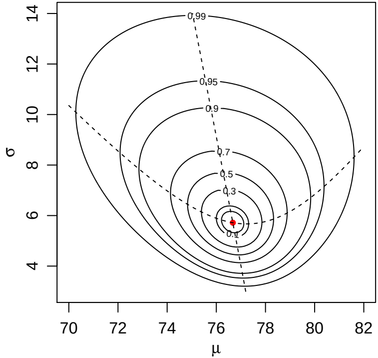
```
nllnormI <- function(par, xp, XI) {
    ll1 <- sum(dnorm(xp, mean = par[1], sd = par[2], log = T))
    L2 <- pnorm(XI, mean = par[1], sd = par[2])
    ll2 <- sum(log(L2[, 2] - L2[, 1]))
    return(-(ll1 + ll2))
}
```

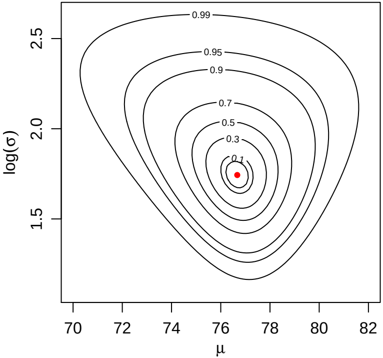
```
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
[1,]   85   80   75   75   75   75 -Inf -Inf -Inf  -Inf  -Inf  -Inf
[2,]  Inf  Inf   80   80   80   80   75   75   75    75    75    75

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))
}
```

Laboratório de Estatística e Geoinformação

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`