MINI - CURSO

Métodos computacionais para inferência com aplicações em R

LEG: Laboratório de Estatística e Geoinformação/UFPR

Equipe LEG :

Paulo Justiniano Ribeiro Jr

Wagner Hugo Bonat

Walmes Marques Zeviani

Elias Teixeira Krainski

Silvia Emiko Shimakura

http://www.leg.ufpr.br

{paulojus,wagner,walmes,elias,silvia}@leg.ufpr.br

57ª RBRAS

Piracicaba, SP, 05-09 de Maio de 2012

Tópicos

- Introdução
- Fundamentos sobre verossimilhança
  - Estimação/maximização (alguns métodos e algorítmos)
  - Intervalos (definições, propriedades assintóticas e aproximações)
  - Testes Hipótese (alternativas e racional)
  - Reparametrização (implicações)
- Verossimilhança (2 ou mais parâmetros)
  - Ortogonalidade
  - Reparametrizações
  - Perfilamento
- Modelos de regressão (GLM, Simplex, Subdisperso, Não Linear, Proc. Poisson não Hom.)
- Efeitos aleatórios (Espaciais, GLMM, Beta Longitudinal, TRI, Linear Dinâmico)
- Comentários adicionais/finais

Motivação

- Agradecimentos
- *Principal público alvo: graduação a início de PG*
- Motivações
- Experiências/exposição das gerações
- Facilidade de recursos computacionais e linguagens
- Uso de rotinas *versus* implementação/teste/ilustração/aprendizado
- Uso crítico e avaliação de limitações de rotinas

Exemplo introdutório

Revisitando 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 $\left(n, y = \frac{\sum_{i=1}^{n} x_i}{n}\right)$?

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

Exercícios I

Simular e fazer o gráfico da (log)verossimilhança para:

1. distribuição Exponencial
2. distribuição Poisson
3. distribuição Normal com variância unitária $N(\theta, 1)$
4. distribuição Normal com média zero $N(0, \theta)$
5. distribuição Gamma com parâmetro de forma (*shape*) igual a 2
6. processo $AR(1)$ de média zero e variância unitária
7. processo de Poisson homogêneo (1 dimensão) e não homogêneo com $\lambda(t) = \theta t$
8. Modelos: (usando $x = 1, 2, 3, \dots, 15$)
    1. $Y_i \sim N(\mu_i, 4)$ com $\mu_i = \theta x_i$
    2. $Y_i \sim B(p_i)$ com $\log\{p_i/(1 - p_i)\} = \theta x_i$
    3. $Y_i \sim P(\lambda_i)$ com $\log(\lambda_i) = \theta x_i$

Exercícios I (cont)

Para cada item anterior:

- escrever uma função de verossimilhança genérica (para qualquer amostra)
- verificar/comparar diferentes maneiras de escrever a função
- verificar se sua função está vetorizada para valores do parâmetro
- explorar as possibilidades de visualização do gráfico da função

Sob o mesmo paradigma

- $Y_1, \dots Y_n$ v.a. ; $N(\theta, 1)$
  - melhor estimativa pontual, IC ($\bar{y} \pm 1,96/\sqrt{n}$), interpretação do IC
- $Y_i = \alpha + \beta i + \epsilon_i$ com $\epsilon_i \sim N(0, \sigma^2)$
  - inferências com múltiplos parâmetros $(\alpha, \beta, \sigma^2)$
  - visualizações e parâmetros de interesse
- $Y \sim N(\mu, \sigma^2)$
  - $\theta = (\mu, \sigma^2)$
  - Interesse em inferências sobre funções dos parâmetros :

$$p = P[Y \geq u] = 1 - \Phi(\frac{u - \mu}{\sigma}) = g_1(\theta)$$

$$u = \mu + \sigma \Phi^{-1}(1 - p_0) = g_2(\theta)$$

- $Y_i \sim N(\beta_0 + \sum_{j=1}^{d} \beta_j x_{ij}, \sigma^2)$
  - x_{ij} valores das covariáveis, $\theta = (\beta_0, \beta_1, \dots \beta_d)$
  - interesse em efeito de um particular x_{ij}.
    Como avaliar se afetado pela demais covariáveis?
- outros modelos

Definições

- $Y_1, \ldots Y_n$ v.a. observáveis com distribuição conjunta dependendo de parâmetros desconhecidos $\theta = (\theta_1, \ldots \theta_d)$ com espaço paramétrico Θ
- Função de verossimilhança avaliada em (uma realização) $y = y_1, \ldots y_n$
- $L(\theta; y) = f_Y(y; \theta)$ (contínua) ou $L(\theta; y) = p(y; \theta)$ (discreta)
- Notação simplificada $L(\theta)$ e $l(\theta) = \log L(\theta)$
  $L(\theta_1) > L(\theta_2) \iff l(\theta_1) > l(\theta_2)$

Comentários:

- interpretação de $L(\theta)$ ou $l(\theta)$:
  Discreto: (log)probabilidade de observar x se θ é o parâmetro verdadeiro
  Contínuo: (log) $\lim_{\delta x \to 0} \frac{P\{x_i \leq X_i < x_i + \delta_i, i=1,\dots n; \theta\}}{\delta x_1 \dots \delta x_n}$

- verossimilhança relativa $L(\theta_1)/L(\theta_2)$ (ou $\log\{L(\theta_1)/L(\theta_2)\} = l(\theta_1) - l(\theta_2)$)

- observações i.i.d. $L(\theta) = \prod_{i=1}^{n} f(y_i; \theta) \;\; \theta \in \Theta$ (univariadas)

- observações independentes $L(\theta) = \prod_{i=1}^{n} f_{Y_i}(y_i; \theta) \;\; \theta \in \Theta$ (univariadas)

- observações bloco independentes $L(\theta) = \prod_{i=1}^{K} f_{Y_k}(y_k; \theta) \;\; \theta \in \Theta$ (multivariadas)

EMV (MLE)

$\hat{\theta}$ EMV de θ satisfaz:

$$L(\hat{\theta}) = \sup_{\theta \in \Theta}\{L(\theta)\} \text{ ou } l(\hat{\theta}) = \sup_{\theta \in \Theta}\{l(\theta)\}$$

Estritamente é um supremo e não um máximo

Definições complementares:

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

Sob certas condições de regularidade $\hat{\theta}$ EMV de θ satisfaz:

$$U(\theta) = 0 \text{ (equação de estimação)}$$

Obtendo estimativas (EMV)

- Solução analítica:
  estudando comportamento de $l(\theta)$ ou resolvendo $U(\theta) = 0$
- Métodos/aproximações numéricas
  - Solução da(s) equação(ões) de estimação (função escore)
    - com uso de derivadas (Newton-Raphson)
    - sem uso de derivadas
  - Maximização da função de (log)-verossimilhança
- Diversidade de algorítimos de maximização
- (re)parametrizações

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

Exercício I (cont)

Para cada item do Exercício obter (quando possível)

- $U(\theta)$ e seu gráfico
- $H(\theta)$
- EMV analítico
- EMV via Newton-Raphson
- EMV via solução de equação de estimação
- EMV via maximização da função de verossimilhança

EMV de funções do parâmetro

EMV de funções (1-1) do parâmetro

$$\phi_1 = g_1(\theta) = 1/\theta \longrightarrow \hat{\phi}_1 = 1/\hat{\theta} = \bar{y}$$

$$\phi_2 = g_2(\theta) = \log(\theta) \longrightarrow \hat{\phi}_2 = log(\hat{\theta}) = \log(\bar{y})$$

$$\phi_3 = g_3(\theta) = P[Y > u] = 1 - F(u;\theta) = \exp\{-\theta u\} \longrightarrow \hat{\phi}_3 = \exp\{-\hat{\theta} u\} = \exp\{-\bar{y} u\}$$

$$\phi_4 = g_4(\theta) = md = F^{-1}(0,5) = -\frac{\log(0,5)}{\theta} \longrightarrow \hat{\phi}_4 = -\frac{\log(0,5)}{\hat{\theta}} = -\log(0,5)\bar{y}$$

$$\phi_5 = g_5(\theta) = md = F^{-1}(0,5) = -\frac{\log(0,5)}{\theta} \longrightarrow \hat{\phi}_5 = -\frac{\log(0,5)}{\hat{\theta}} = -\log(0,5)\bar{y}$$

Função deviance

Uma representação alternativa (e conveniente) da função de verossimilhança

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

- chamada de função deviance
- interpretação como (log de) verosimilhança relativa
- (ver gráfico)
- características da representação gráfica
- localização do EMV (raiz)

Exercício I (cont)

- Traçar a função deviance para cada caso do Exercício I
- No exemplo da exponencial:
  - traçar a função de (log)verrossimilhança e deviance para ϕ_1 e ϕ_4
  - calcular $I(\hat{\phi}_1), I(\hat{\phi}_2), I(\hat{\phi}_3), I(\hat{\phi}_4)$ e comparar com $I(\theta)$
  - ajustar os valores dos eixos das abcissas "adequadamente"
  - discutir os resultados

Estimação intervalar

Definição "natural": valores com compatibilidade aceitável com amostra

Intervalo ou região de confiança para θ é um conjunto de valores que satisfaz uma das seguintes condições (equivalentes)

- $\{\theta \in \Theta | 0 \leq \frac{L(\theta)}{L(\hat{\theta})} \leq c_L\} \;\; 0 < c_L \leq 1$
- $\{\theta \in \Theta | 0 \leq l(\hat{\theta}) - l(\theta) \leq c_l\} \;\; c_l \geq 0$
- $\{\theta \in \Theta | 0 \leq D(\theta) \leq c_D\} \;\; c_l \geq 0$

$c_L \rightarrow c_l \rightarrow c_D$	$c_L \leftarrow c_l \leftarrow c_D$
$c_l = -\log(c_L)$	$c_l = c_d/2$
$c_D = 2c_l = -2\log(c_L)$	$c_L = \exp(-c_d/2)$

Exercício I (cont)

Considere as verossimilhanças relativas de 50, 26, 15 e 3,6%
(i.e. $c_L = 0,5; 0,15$ e $0,04$)

- encontrar os valores de c_I e c_D
- encontrar os limites dos intervalos para cada um desses valores e adicionar aos gráficos
  - das funções de log-verosimilhança e deviance para cada caso do Exercício I
  - das funções de log-verosimilhança e deviance para cada (re)parametrização da exponencial
- calcular $\sqrt{c_D}$ e $P[|Z| < \sqrt{c_D}]$ $(Z \sim N(0,1))$

Resposta parcial:

c_L	c_I	c_D	$P[\|Z\| < \sqrt{c_D}]$
50%	$0,693$	$1,386$	$0,761$
26%	1.661	$3,321$	$0,899$
15%	$1,897$	$3,794$	$0,942$
$3,6\%$	$3,324$	$6,648$	$0,990$

Dificuldades

Funções típicas e atípicas
(gráficos)

- não concava
- limites do espaço paramétrico dependente de parâmetro
- borda de espaços paramétricos
- intervalos disjuntos

Exemplo A

$$Y \sim N(\theta, 1) \quad \Theta = (-\infty, \infty)$$

$$l(\theta) = C - \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta)^2$$

$$U(\theta) = \sum_{i=1}^{n}(y_i - \theta)$$

$$H(\theta) = -n$$

$$U(\theta) = 0 \longrightarrow \hat{\theta} = \bar{y}$$

$$D(\theta) = \ldots = n(\bar{y} - \theta)$$

$$D(\theta) = c_D \longrightarrow (\hat{\theta}_I, \hat{\theta}_S) = (\bar{y} - \sqrt{c_D/n}\,,\ \bar{y} + \sqrt{c_D/n})$$

Aspectos relevantes

- exatamente quadrática (simétrica)
- côncava, espaço paramétrico Θ não depende de θ, $\hat{\theta}$ é interior a Θ
- completamente determinada por $\hat{\theta}, I(\hat{\theta})$ e $H(\hat{\theta})$
- comprimento dos IC's independe dos dados (o comprimento mesmo para diferentes experimentos)
- $\bar{y} \sim N(0, 1/n) \longrightarrow \sqrt{n}(\bar{y} - \theta) \sim N(0, 1) \longrightarrow n(\bar{y} - \theta) = D(\theta) \sim \chi^2_1$
- Pode-se calcular $P[D(\theta) \leq c_d]$ ou ...
- Pode-se encontrar c_D tal que $P[D(\theta) \leq c_d] = 1 - \alpha$
- IC probabilístico

Confiança	c_D	$\sqrt{c_D}$	c_L
90%	2,71	1,645	25,8%
95%	3,84	1,960	14,77%
99%	6,63	2,576	3,6%

Exemplo B

$$Y \sim U(0, \theta)$$

$$l(\theta) = \theta^{-n} \text{ (gráfico)}$$

$$U(\theta) = -n\theta^{-(n-1)}$$

$$H(\theta) = n(n-1)\theta^{-(n-1)}$$

$$D(\theta) = \ldots = 2n[\log(\theta) - \log(\hat{\theta})]$$

$$D(\theta) = c_D \longrightarrow (\hat{\theta}_I, \hat{\theta}_S) = (\hat{\theta}\ ,\ \hat{\theta} \exp\{c_D/2n\})$$

Aspectos relevantes

- não quadrática, assimétrica e não se torna quadrática com $n \rightarrow \infty$
- $U(\theta) = 0$ não produz $\hat{\theta}$
- EMV é óbvio pelo gráfico e/ou
  $$L(\theta) = \theta^{-n} \prod_{i=1}^{n} I_{[0,\theta]}(y_i) \longrightarrow L(\theta) = \theta^{-n} I_{[0,\theta]}(max(y_1, \ldots y_n))$$
- $\Theta = (max\{y_1, \ldots y_n\}, \infty)$
- informação desigual os redor do EMV:
  esquerda: $\hat{\theta}$ deve ser maior que $max\{y_1, \ldots y_n\}$
  direita: pouca informação
- verossimilhança só depende de $(n, max\{y_1, \ldots y_n\})$
- domínio da v.a. depende do parâmetro
- $D(\theta) \nsim \chi^2$

Exemplo C

$$Y \sim Exp(\theta)$$

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

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

$$H(\theta) = -\frac{n}{\theta^2}$$

$$\frac{d^i}{d\theta^i} = (-1)^{i+1}(i-1)!n\theta^{-i}$$

$$U(\theta) = 0 \longrightarrow \hat{\theta} = 1/\bar{y}$$

$$D(\theta) = \ldots = 2n[log(\hat{\theta}/\theta) + \bar{y}(\theta - \hat{\theta}]$$

Aspectos relevantes

- assimétrica, concava, polinomial de ordem infinita
- mas aproxima-se de simétrico com $n \to \infty$
- $H(\theta)$ depende do valor de θ
- IC resolvendo $D(\theta) = c_d$
    - "exata"por métodos numéricos
    - aproximação por Taylor

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

$$\tilde{D}(\theta) = c_D \longrightarrow (\hat{\theta}_I, \hat{\theta}_S) = \left(\hat{\theta}(1 - \sqrt{c_D/n})\,,\ \hat{\theta}(1 - \sqrt{c_D/n})\right)$$

- $D(\theta) \approx \chi^2_1$
- Intuição: por que $l(\theta)$ é assimétrica ao redor the $\hat{\theta}$ para n pequeno?

Exemplo C - Complementos/Exercícios

- Ilustrar que o comportamento de $I(\theta)$ com valores crescentes do tamanho de amostra, por exemplo $n = (10, 25, 100)$
- Deduzir as expressões de $D(\theta)$, $\tilde{D}(\theta)$ e $(\hat{\theta}_I, \hat{\theta}_S)$
- Escrever algorítmo que retorne: EMV pontual e intervalares (por métodos numéricos e aproximação quadrática)
- Programar e executar rotinas para um estudo de simulação para verificar a taxa de cobertura dos dois IC's (numérico e aproximação quadrática) para os diferentes tamanhos de amostra

Outros exemplos (iid)

- Dados arredondados e modelos contínuos para observações discretas
  Exemplo para observações inteiras:

$$L(\theta) = \prod_{i=1}^{n} \{F_Y(y_i + 1/2) - F_Y(y_i - 1/2)\} = \prod_{i=1}^{n} \int_{y_i-1/2}^{y_i+1/2} f_Y(y_i;\theta)dy_i \approx \prod_{i=1}^{n} f_Y(y_i) \times 1$$

  Aproximação só é boa se $f_Y(y)$ é relativamente constante no intervalo

- Dados censurados (direita, esquerda, intervalar)

$$L(\theta) = \prod_{i=1}^{n} f_Y(y_i,\theta)^{\delta_i} (P[L_I < Y_i < L_S])^{1-\delta_i}$$

δ_i é variável indicadora de censura
(L_I, L_S) são os limites dos intervalos das observações

Exemplo D

A contribuição de cada dados

- No modelo Gaussiano
- No modelo exponencial
- Observações de diferentes "tipos" na Gaussiana

Invariância

- Interesse em **inferências sobre $g(\theta)$**
- $g(\theta)$ função 1-1 (bijetora)
- $\phi_1, \phi_2, \phi_3, \phi_4, \phi_5$ (exemplo da exponencial)
- Intuição: $L(\phi) = L(g(\theta))$
  deformação do eixo das abcissas sem alterar ordenadas
- $\hat{\phi} = g(\theta), \ \hat{\phi}_I = g(\theta_I), \ \hat{\phi}_S = g(\theta_S)$
- Importância da invariância
  - apenas uma maximização (métodos numéricos) necessária para EMV e IC
  - $L(\cdot)$ pode ser quadrática em uma parametrização e altamente não quadrática em outra
  - sugere escolha de parametrização mais adequada

UFPR UNIVERSIDADE FEDERAL DO PARANÁ

Exercícios

No exemplo da distribuição exponencial

$$
\begin{aligned}
\phi_1 &= g_1(\theta) = 1/\theta \\
\phi_2 &= g_2(\theta) = \log(\theta) \\
\phi_3 &= g_3(\theta) = P[Y > u] = 1 - F(u; \theta) = \exp\{-\theta u\} \\
\phi_4 &= g_4(\theta) = md = F^{-1}(0,5) = -\log(0,5)/\theta \\
\phi_5 &= g_5(\theta) = q_{0,90} = F^{-1}(0,90) = -\log(0,90)/\theta
\end{aligned}
$$

- traçar a função de log-verosimilhança para $\phi_1, \phi_2, \phi_3, \phi_4, \phi_5$
- calcular $l(\hat{\phi}_1), l(\hat{\phi}_2), l(\hat{\phi}_3), l(\hat{\phi}_4), l(\hat{\phi}_5)$ e comparar com $l(\theta)$
- ajustar os valores dos eixos das abcissas "adequadamente"
- em cada caso encontrar os valores do parâmetro que correspondem a 20% do valor maximizado da verossimilhança
- traçar as funções deviance indicando os intervalos
- escrever funções R que sejam "genéricas"

Aproximações assintóticas

- Avaliar efeito da amostra no IC pode ser não trivial mesmo para exemplos simples
- Escolha (probabilística) de c_D
- Avaliar comportamento de $L(\theta)$ quando $n \to \infty$
- Aproximações por série de Taylor ao redor de $\hat{\theta}$

$$l(\theta) = l(\hat{\theta} + (\theta - \hat{\theta})l'(\hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})^2 l''(\theta^*) \text{ para } |\theta^* - \theta| \leq |\hat{\theta} - \theta|$$

$$l'(\theta) = l'(\hat{\theta} + (\theta - \hat{\theta})^2 l''(\theta^+) \text{ para } |\theta^+ - \theta| \leq |\hat{\theta} - \theta|$$

- Função Escore: $U(\theta) = l'(\theta)$
- Informação observada: $I_O(\theta) = -H(\theta) = l''(\theta)$
- Informação esperada: $I_E(\theta) = \mathrm{E}_Y[I_O(\theta)] = \mathrm{E}_Y \left\{ \frac{d^2 \log L(\theta; Y)}{d\theta^2} \right\}$
- (log) verossimilhança é função aleatória, v.a. determina $\hat{\theta}, U(\theta)$ e $I_O(\theta)$

Propriedades amostrais

Propriedades amostrais (substituindo y por Y)

- D1: $\mathrm{E}(T) = g(\theta)$: $T = T(Y)$ é estimador não viesado de $g(\theta)$
- L1: $\mathrm{E}[U(\theta)] = 0$ e $\mathrm{Var}[U(\theta)] = I_E(\theta)$
- L2: $\mathrm{E}(T) = \theta \longrightarrow \mathrm{Var}[T] \geq [I_E(\theta)]^{-1}$ (limite inferior de Cramér-Rao)
- L3: $\mathrm{E}(T) = g(\theta) \longrightarrow \mathrm{Var}[T] \geq [g'(\theta)]^2 [I_E(\theta)]^{-1}$

Sob 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

Resultados

Para problemas regulares com $n \to \infty$:

- T1: $\sqrt{I_E(\theta)}(\hat{\theta} - \theta) \sim N(0, 1)$ ou $\hat{\theta} \sim N(\theta, [I_E(\theta)]^{-1})$
  (distribuição assintótica do EMV)
  Elementos: aproximação quadrática, $I_O \to I_E$ e TLC
- C1: $\hat{\phi} = g(\hat{\theta}) \sim N(\phi, [g'(\theta)]^2[I_E(\theta)]^{-1})$
  Método *delta*:
  $$\mathrm{Var}(\hat{\phi}) = [g'(\theta)]^2[I_E(\theta)]^{-1} \longrightarrow \left[\mathrm{se}(\hat{\phi}) = |g'(\theta)|[I_E(\theta)]^{-1/2}\right]$$
- C2: Equivalência assintótica e conveniência ($\neq$'s convergências)
  $$\sqrt{I_O(\hat{\theta})}(\hat{\theta} - \theta) \approx \sqrt{I_O(\theta)}(\hat{\theta} - \theta) \approx \sqrt{I_E(\hat{\theta})}(\hat{\theta} - \theta) \approx \sqrt{I_E(\theta)}(\hat{\theta} - \theta) \sim N(0, 1)$$
- T2: $D(\theta) = 2[l(\hat{\theta}) - l(\theta)] \sim \chi^2_1$ (pela aproximação)

UFPR UNIVERSIDADE FEDERAL DO PARANÁ

leg

Discussão

- EMV $\hat{\theta}$ assintoticamente:
  não viciado, atinge limite inferior de Cramér-Rao com erro padrão $se(\hat{\theta}) = [I_E(\theta)]^{-1/2}$
  e intervalos de confiança (probabilísticos) $(1-\alpha)$ da forma $\hat{\theta} \pm z_{\alpha/2} se(\hat{\theta})$

- EMV $\hat{\phi} = \hat{\theta}$ assintoticamente:
  não viciado, atinge limite inferior de Cramér-Rao com erro pradrão
  $se(\hat{\phi}) = \sqrt{[g'(\theta)]^2 [I_E(\theta)]^{-1}}$
  e intervalos de confiança (probabilísticos) $(1-\alpha)$ da forma
  $g(\hat{\theta}) \pm z_{\alpha/2} se(\hat{\phi}) = g(\hat{\theta}) \pm z_{\alpha/2} |g'(\theta)| [I_E(\theta)]^{-1/2}$

- $I_E(\theta)$ pode ser substituído por $I_E(\hat{\theta})$, $I_O(\theta)$ ou $I_O(\hat{\theta})$ e $g'(\theta)$ por $g'(\hat{\theta})$

- $I_O(\hat{\theta})$: fácil obtenção, obtenção por métodos numéricos e boas propriedades (intuição!)

- IC $(\hat{\theta}_I, \hat{\theta}_S)$ mais razoável é obtido por: $\{\theta \in \Theta : D(\theta) \leq c_D\}$

- IC $(\hat{\phi}_I, \hat{\phi}_S) = (g(\hat{\theta}_I), g(\hat{\theta}_S))$
- Se transformação $g(\cdot)$ é não linear, invariância não é válida para aproximação quadrática

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

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

Exemplo A (cont)

$Y \sim N(\theta, 1)$

- $I_E(\theta) = I_E(\hat{\theta}) = I_O(\theta) = I_O(\hat{\theta}) = n^{-1}$
- log-verossimilhança é exatamente quadrática: $(\tilde{\theta}_I, \tilde{\theta}_S) = (\hat{\theta}_I, \hat{\theta}_S)$
- ambos IC's exatos para θ
- $\phi = P[Y \leq u] = \Phi(u - \theta)$
- $(\hat{\phi}_I, \hat{\phi}_S)$:
$$\{\Phi(u - \hat{\theta}_S), \Phi(u - \hat{\theta}_I)\}$$
- $(\tilde{\phi}_I, \tilde{\phi}_S)$:
$$(\hat{\phi} - z_{\alpha/2} se(\hat{\phi}), \hat{\phi} - z_{\alpha/2} se(\hat{\phi}))$$
$$\mathrm{Var}(\hat{\phi}) = [g'(\theta)]^2 [I_E(\theta)]^{-1} \approx [g'(\hat{\theta})]^2 [I_O(\hat{\theta})]^{-1} = f(u; \theta = \hat{\theta})/n$$
- Comparação gráfica

Exemplo C (cont)

$Y \sim Exp(\theta)$

- $I_E(\theta) = I_O(\theta) = n\theta^{-2}$ (constante, não depende dos valores da amostra)
- log-verossimilhança assimétrica $(\tilde{\theta}_I, \tilde{\theta}_S) \neq (\hat{\theta}_I, \hat{\theta}_S)$
- IC's podem ser inexatos para θ
- $(\hat{\theta}_I, \hat{\theta}_S)$ apenas numericamente
- $(\tilde{\theta}_I, \tilde{\theta}_S) = (\hat{\theta}(1 - z_{\alpha/2}/\sqrt{n}, \hat{\theta}(1 + z_{\alpha/2}/\sqrt{n})$ (já visto no exemplo)
- $se(\hat{\theta}) = [I_E(\hat{\theta})]^{-1/2} \approx \hat{\theta}/\sqrt{n}$
- $\phi = P[Y \leq u] = 1 - \exp\{-\theta u\}$

Exercício

- 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)

Exercício: família exponencial

- $f(x;\theta) = \exp\{h(\theta) + k(y) + c(\theta)t(y)\}$
- $I(\theta)$
- $U(\theta)$
- $H(\theta)$
- IC para $c(\theta)$
- IC para θ

Fazer gráficos para diferentes membros da família

Uniforme

Intervalo probabilístico para problemas não regulares:
requer examinar caso a caso
Exemplo: Distribuição Uniforme $Y \sim [0, \theta]$

$$\begin{aligned} P[\hat{\theta}] &= P[\hat{\theta} \exp c_D/(2n) < \theta] = P[\hat{\theta} d < \theta] \\ &= \ldots \\ &= P[max(Y_1, \ldots, Y_n) \leq \theta d] \\ &= \ldots \\ d^n &= 1 - \alpha \end{aligned}$$

Ponto de corte em $D(\theta)$ para intervalo probabilístico

$$d = (1-\alpha)^n = \exp\{c_D\}/(2n) \longrightarrow c_D = \log(2n) + n\log(1-\alpha)$$

Teste da razão de verossimilhança

$\theta = \theta_0$ *vs* hipótese irrestrita

Decorre naturalmente de avaliar:

$$\frac{L(\hat{\theta})}{L(\theta_0)}$$

ou equivalentemente

$$\log \frac{L(\hat{\theta})}{L(\theta_0)} = l(\hat{\theta}) - l(\theta_0) = D(\theta_0)/2$$

Interpretações:

- evidência relativa
- probabilística
- comportamento assintótico e aproximações como em IC

Comparações entre testes

Resumindo

- A informação dos dados está representada na função de verossimilhança
- "melhor" estimativa e intervalos são resumos da função
- inferência fácil (gráfico)
- avaliar $I(\theta)$ ou $D(\theta)$;
- encontrar $(\hat{\theta}_I, \hat{\theta}, \hat{\theta}_S)$ (métodos analíticos ou numéricos)
- atualização/adição de dados
- resultados analíticos (poucos casos)
- aproximações: como $I(\theta)$ se comporta quando $n \to \infty$

Paradigmas

Aitkim (2010)

- Verossimilhança "pura"(*pure likelihood*)
- Bayesiano (*Bayesian Theory*)
- Frequentista baseada em verossimilhança (*Likelihood-based repeated sampling theory*)
- por desenho guiada por modelo (*model-guided survey sampling theory*)

Exercícios II

$$Y_i \sim N(\mu, \sigma^2)$$
$$\text{Exemplo}: Y_i \sim N(10, 1)$$

- contornos razoavelmente elípticos
- eixos paralelos aos dos parâmetros (ortogonal)
- simétrico em μ e assimétrico em σ^2 (σ, $\log(\sigma)$)
- aspectos dos "cortes"
- ortogonalidade
- fixando EMV *vs* perfis de verossimilhança
- dados intervalares

Exercícios II

$$Y_i \sim N(a, b)$$
$$\text{Exemplo}: Y_i \sim U(0, 1)$$

- nada elípticos!
- aspecto triangular
- fixando $a = 0$ aspecto de Uniforme $[0, b]$

Exercícios II

$$Y_i \sim G(\alpha, \beta)$$
$$\text{Exemplo}: Y_i \sim G(5, 4)$$

- $\mathrm{E}[Y] = \alpha/\beta$
- inserir a linha $\beta = 4\alpha/5$ no gráfico
- elíptico
- aspectos dos "cortes"
- reparametrizações
- ortogonalização
- ICs e regiões

Exercícios II

$$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$$
$$\text{Exemplo}: Y_i \sim N(10 + 0.2x_i, 1) \quad x_i = i$$

- 4-D
- Gráficos 3D para cada parâmetro
  diferentes opções
- (β_0, β_1) não paralelos aos eixos
  Para manter $\mathrm{E}[Y_i] = \beta_0 + \beta_1 x_i$, o aumento de um implica na diminuição de outro!
- ortogonais a σ
- Modelo usando variável centrada: $\mathrm{E}[Y_i]\beta_0^* + \beta_1^* x_i$
  reparametrização!
- extensões/recomendações para múltiplas covariáveis

Exercícios II

$$(Y_1, Y_2, Y_3) \sim Multinomial(\theta_1, \theta_2, \theta_3)$$
$$\text{Exemplo :}(Y_1, Y_2, Y_3) \sim Multinomial(1/3, 1/3, 1/3) \; n = 25$$

- efeito da restrição $\theta_3 = 1 - \theta_1 - \theta_2$
- porém aproximadamente elípticos no "topo"
- cortes
- ICs

Exercícios II

$$Y_i \sim B(\theta_i) \quad \log(\theta_i/(1 - theta_i)) = \beta_0 + \beta_1 x_i$$
$$\text{Exemplo}: n = 25,\; x_i = i \;\beta_0 = 0 \;\; \beta_1 = 0,1$$

- contornos aproximadamente elípticos
- contornos aproximadamente ortogonais para variável centrada

Resultados assintóticos

Extensões do caso univariado

Parâmetros de interesse e nuisance

- Funções dos parâmetros
- Perfil de verossimilhança
  nos casos anteriores e outros formato de verosimilhança
- Vantagens interpretação e críticas
- Outras formas de "marginalização"
- Bayes/Integrada
  OBS: $f(\theta) = 1 \longrightarrow f(\theta^2) = 1/(2\theta)$