Estatística bayesiana

Distribuições *a priori* não informativas

José Passos

ISEG-UTL

22 de Novembro de 2011

Tabela de conteúdos

Introdução

- Uma das características da inferência bayesiana é a possibilidade de expressar a informação sobre θ por intermédio de uma distribuição de probabilidade
- Vamos estudar a situação de ausência de informação *a priori*

Abordagem de Laplace

- Laplace foi o primeiro a utilizar distribuições *a priori* não informativas
- Estas distribuições baseiam-se no princípio da razão insuficiente de Bayes-Laplace que assumem a equiprobabilidade dos acontecimentos elementares na ausência de informação
- Um exemplo clássico é a população de Bernoulli de parâmetro θ. De acordo com o princípio da razão insuficiente a distribuição *a priori* não informativa é uniforme, $g(\theta) = I_{(0,1)}(\theta)$, com distribuição *a posteriori*,

$$h(\theta|x) \propto g(\theta)\ L(\theta|x) = x^{\theta}(1-x)^{1-\theta} I_{(0,1)}(\theta)$$

Problemas da abordagem de Laplace

- Se Θ não é compacto, $g(\theta) \propto k$ é imprópria, i.é., $\int g(\theta)\,d\theta = \infty$
- O princípio de acontecimentos equiprováveis não é coerente sob partições de Θ:
  - se $\Theta = \{\theta_1, \theta_2\}$ o princípio da razão insuficiente conduz a $g(\theta_1) = g(\theta_2) = 1/2$;
  - Se $\Theta = \{\theta_1, w_1, w_2\}$ tem-se de acordo com o mesmo princípio $g(\theta_1) = 1/3$, contrariando a 1ª condição.
- A distribuição *a priori* não é invariante sobre parametrizações de Θ: na parametrização $\eta = \tau(\theta)$, transformação biunivoca de θ, em geral, $g^*(\eta) \propto |d\tau^{-1}(\eta)/d\eta|$ não é constante

Problemas da abordagem de Laplace

- Exemplo: numa população de Bernoulli, $X \sim Be(\theta)$, tem-se $g(\theta) = I_{(0,1)}(\theta)$. Considerando a parametrização,

$$\eta = \tau(\theta) = \frac{\theta}{1-\theta}$$

a distribuição *a priori* na nova parametrização, $g^*(\eta)$, é dada por,

$$g^*(\eta) = g\left(\tau^{-1}(\eta)\right)\left|\frac{d\tau^{-1}(\eta)}{d\eta}\right| = \frac{1}{(1+\eta)^2}$$

que não é constante em η

1ª e 2ª regra de Jeffreys

- Jeffreys procura resolver alguns dos problemas associados à abordagem de Bayes- Laplace
- Propõe duas regras: uma para parâmetros de localização e outra para parâmetros de escala, com propriedades de invariância para certo tipo de transformações
- **1ª regra (parâmetros de localização):** se $\Theta = \{\theta : -\infty < \theta < \infty\}$ Jeffreys propõe $g(\theta) \propto k$
  - $g(\theta) \propto k$ é imprópria
  - $g(\theta) \propto k$ é invariante a parametrizações do tipo $\eta = a + b\theta$ pois ainda se tem $g^*(\eta) \propto k$

1ª e 2ª regra de Jeffreys

- **2ª regra (parâmetros de escala):** se $\Theta = \{\theta : 0 < \theta < \infty\}$ Jeffreys propõe $g(\theta) \propto 1/\theta$
  - $g(\theta) \propto 1/\theta$ é imprópria
  - $g(\theta) \propto 1/\theta$ é invariante a parametrizações do tipo $\eta = \theta^n$ pois ainda se tem $g^*(\eta) \propto 1/\eta$

Abordagem geral de Jeffreys

- Jeffreys fundamenta as duas regras num tratamento mais geral utilizando a medida de informação de Fisher sobre θ contida na amostra.
- Jeffreys explora as propriedades de invariância da medida de informação de Fisher propondo como distribuição *a priori* não informativa,

$$g(\theta) \propto I(\theta)^{1/2}$$

- A aplicação das duas regras ou da regra geral conduzem à mesma distribuição *a priori*

Abordagem geral de Jeffreys

- Se θ é um vector com k parâmetros Jeffreys propõe como distribuição *a priori* não informativa,

$$g(\theta) \propto [det\ I(\theta)]^{1/2}$$

- No caso k-paramétrico a aplicação das duas regras ou da regra geral não conduzem, em geral, à mesma distribuição *a priori*

Abordagem geral de Jeffreys

Exemplo: $X \sim N(\mu, \sigma)$ e considere uma amostra casual de dimensão n

- regra geral:

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

  com distribuição *a priori*,

$$g(\mu, \sigma) \propto |I(\mu, \sigma)|^{1/2} = \frac{n\sqrt{2}}{\sigma^2} \propto \frac{1}{\sigma^2}$$

Abordagem geral de Jeffreys

- 1ª e 2ª regra: supondo independência de μ e σ temos,

$$g(\mu, \sigma) = g(\mu) g(\sigma)$$

sendo μ um parâmetro de localização, $g(\mu) \propto k$, e sendo σ uma parâmetro de escala, $g(\sigma) \propto 1/\sigma$, tem-se a distribuição *a priori*

$$g(\mu, \sigma) = g(\mu) g(\sigma) \propto \frac{k}{\sigma} \propto \frac{1}{\sigma}$$

Outras abordagens

- Box-Tiao
- Entropia máxima (Zellner)
- Berger-Bernardo

Bibliografia

- Paulino; Turkman; Murteira (2003), *Estatística bayesiana*, Fundação Calouste Gulbenkian.