Introdução à Inferência Bayesiana

Prof. Caio Azevedo

- Considere que desejamos estimar a renda média (em salários mínimos) do Brasil em 2012. Para isso, será retirada uma amostra segundo algum plano amostral.
- Podemos super algum modelo estatístico (distribuição de probabilidade) ou simplesmente considerar a média amostral como estimador (naturalmente o plano amostral deve ser levado em consideração).
- Suponha que o tamanho da população (N) é substancialmente maior do que o tamanho da amostra (n) e que o plano amostral foi AASs (sem reposição). Nesse caso, de um modo geral, podemos considerar que o plano foi AASc (com reposição).

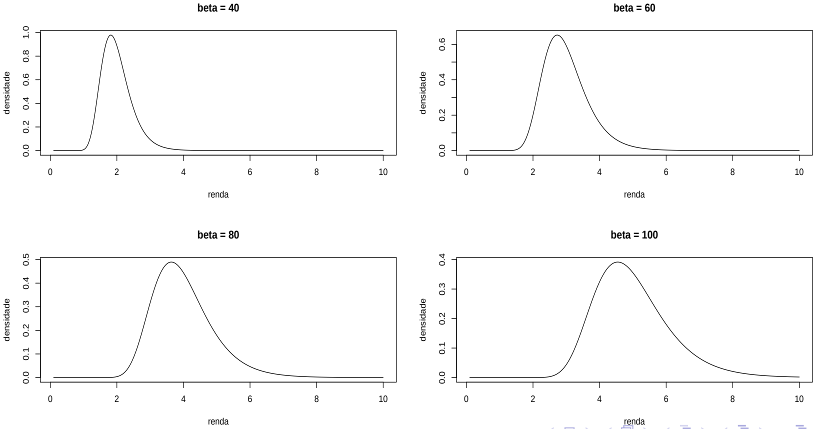
- Assumimos que as observações (amostra) são i.i.d $X \sim exp(\mu), \mathcal{E}(X) = \mu, p_X(x|\mu) = \frac{1}{\mu}e^{-x/\mu}\mathbb{1}_{(0,\infty)}(x)$.
- Parâmetro de interesse $\mathcal{E}(X) = \mu$
- Estimador de máxima verossimilhança de μ é $\widehat{\mu} = \overline{X}$ e $\mathcal{V}(\overline{X}) = \dfrac{\mu^2}{n}$.
- Resultado assintótico: $IC(\mu,\gamma) = \overline{X} \pm z_{\gamma}\frac{\overline{X}}{\sqrt{n}}, \frac{1-\gamma}{2} = P(Z > z_{\gamma})$.
  - Tamanho da amostra tem de ser suficientemente grande.
  - O IC permite valores negativos.
  - Como incorporar informações adicionais (além do plano amostral)?

- Suponha que em uma pesquisa anterior chegou-se à $E(X) = \widetilde{\mu} = 4,00$. Os especialistas afirmam que, muito provavelmente, a renda média do Brasil aumentou desde então e que sua distribuição deslocou-se para à direita (quartis aumentaram).
- Como incorporar essa informação acima e não utilizar resultados assintóticos?
- Vamos introduzir a informação acima em termos do cálculo de probabilidades, afirmando "à priori" que:

$$p_{\mu}(\mu|\alpha,\beta) = p_{\mu}(\mu) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\mu^{(-\alpha-1)}e^{-\beta/\mu}\mathbb{1}(\mu)_{(0,\infty)}$$

em que $\mathcal{E}(\mu) = \dfrac{\beta}{\alpha-1}, \alpha > 1$ e $\mathcal{V}(\mu) = \dfrac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \alpha > 2$.

Diferentes prioris, $\alpha = 21$



Ingredientes

- Conhecimento (Informação) à priori

  $p(\mu) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\mu^{-(\alpha+1)}e^{-\beta/\mu}\mathbb{1}(\mu)_{(0,\infty)}$.

- Informação obtida através do processo de amostragem

  (verossimilhança), $\mathbf{x} = (x_1, ...x_n)$.

$$L(\mu) = p(\mathbf{x}|\mu) = \prod_{l=1}^{n}\left\{\mu^{-1}e^{-x_i/\mu}\right\} = \mu^{-n}e^{-n\overline{x}/\mu}$$

- Conhecimento à posteriori (depois de observar os dados)

$$p(\mu|\mathbf{x}) = \frac{p(\mu,\mathbf{x})}{p(\mathbf{x})} = \frac{p(\mu)p(\mathbf{x}|\mu)}{\int_{\mathcal{R}^+} p(\mu)p(\mathbf{x}|\mu)d\mu}$$

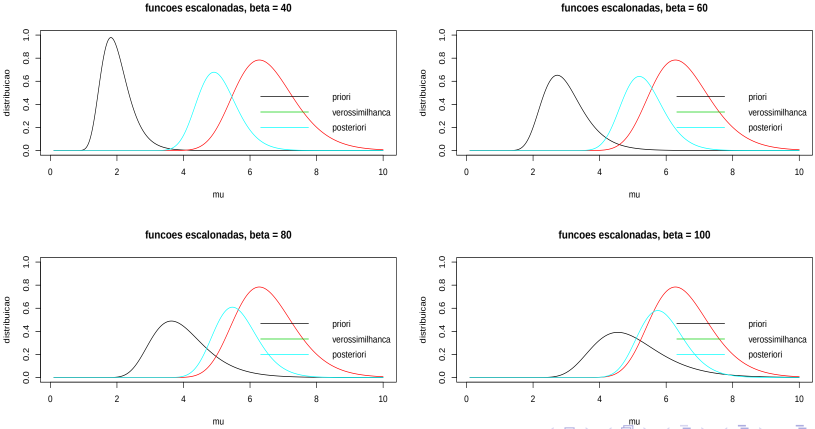
Ingredientes

- Pode-se provar que (exercício):

$$p(\mu|\mathbf{x}) = \frac{(n\overline{x} + \beta)^{n+\alpha}}{\Gamma(n+\alpha)} \mu^{-(n+\alpha+1)} e^{-\frac{n\overline{x}+\beta}{\mu}} \mathbb{1}_{(0,\infty)}(\mu)$$

- A disitribuição à priori normal inversa é conjugada para o modelo exponencial (com a parametrização adotada).
- Nesse caso, temos que: $\mathcal{E}(\mu|\mathbf{X}) = \dfrac{n\overline{x}+\beta}{n+\alpha-1}$ e $\mathcal{V}(\mu|\mathbf{X}) = \dfrac{(n\overline{x}+\beta)^2}{(n+\alpha-1)^2(n+\alpha-2)}$.

Priori, verossimilhança e posteriori $\alpha = 21$



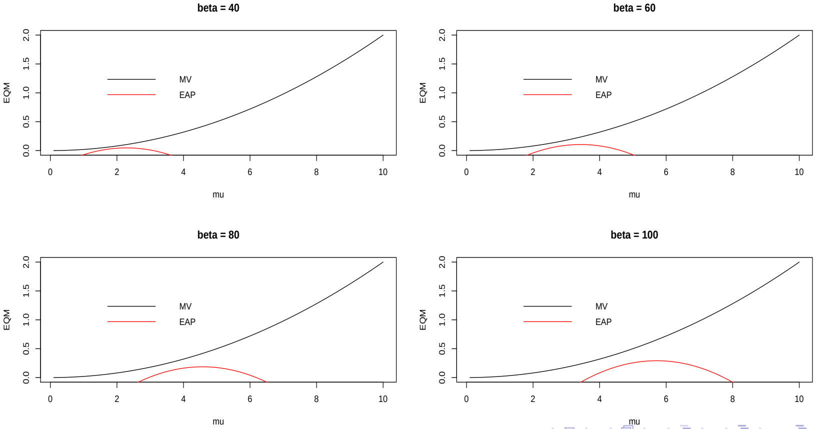
Comparação "frequenstista" entre os estimadores MV e EAP

- Temos que: $\mathcal{E}(\widehat{\mu}) = \mu$ e $EQM(\widehat{\mu}) = \mathcal{V}(\widehat{\mu}) = \dfrac{\mu^2}{n}$.
- Pode-se provar que: $\mathcal{E}_{X|\mu}(\mu|\mathbf{X}) = \dfrac{n\mu + \beta}{n + \alpha - 1}$,

  $\mathcal{V}_{X|\mu}(\mu) = \dfrac{\mu^2}{n(n + \alpha - 1)^2}$ e

  $EQM_{X|\mu}(\mu|\mathbf{X}) = \dfrac{n\mu^2 - [\beta - \mu(\alpha - 1)]^2}{(n + \alpha - 1)^2}$.

EQM $\alpha = 21$


beta = 40
beta = 60
beta = 80
beta = 100
EQM
mu
MV
EAP
0.0
0.5
1.0
1.5
2.0
0
2
4
6
8
10

Comparação

- Para uma amostra de tamanho $n = 50$ simulada de $X \sim exp(5, 5)$.

Estimador	Estimativa	Variancia Estimada
MV	6,28	0,89
EAP com $\beta = 40$	5,06	0,63
EAP com $\beta = 60$	5,34	0,63
EAP com $\beta = 80$	5,63	0,63
EAP com $\beta = 100$	5,91	0,63

Introdução

- Inferência Bayesiana: aprendizado, combinação entre o conhecimento previo e resultados de alguma pesquisa (no sentido amplo da palavra).
- Parâmetro: tratado como variável aleatória (pode ou não ser considerado "fixo").
- Incerteza: introduzida através da priori e atualizada combinando-a com a verossimilhança.

Definições e Notações

- Consideramos um modelo estatístico formado por $(\Omega, \mathcal{A}, \mathcal{P})$, (espaço amostral, sigma álgebra de eventos de Ω, família de medidas de probabilidade).
- $\mathcal{P} = \{\mathcal{P}_\theta : \theta \in \Theta\}$, $\Theta \subset \mathcal{R}^p$ (em geral infinito não-enumerável). Basicamente: $\mathcal{P}_\theta = F_{X|\theta}(.)$ (postulado).
- Realizar inferência Bayesiana (estimação pontual, intervalar e testar hipóteses estatísticas) em um dado espaço estatístico.
- $p^* = \{p_\theta : \theta \in \Theta\}$ (distribuição à priori). Os elmentos de p^* não precisam ser fdp (função de probabilidade ou função densidade de probabilidade).

Continuação

- Suporte da fdp:
  - $\Omega = \{x : f_{X|\theta} \geq 0\}$, $f_{X|\theta}(.)$ é a fdp associada à fda $F_{X|\theta}(.)$. Eventualmente os subíndices podem ser suprimidos.
- Assim, (Ω, Θ) passa a ser nosso conjunto de interesse.
- Amostra aleatória (aa): $X_1|\theta, ..., X_n|\theta$ são condicionalmente independentes (dado θ) e identicamente distrbuídas segundo $X|\theta \sim F_{X|\theta}$. Eventualmente, as suposições de independência e mesma distribuição podem ser relaxadas.
- Variável aleatória: $X_i|\theta$. Valor observado: x_i.

Continuação

- Se $X_1|\theta, ..., X_n|\theta$ é uma aa de $X|\theta$, $X_1, .., X_n$ serão permutáveis (mais adiante). $\underline{X} = \mathbf{X} = (X_1, ..., X_n)$.
- Estimador $\widehat{\theta}_{EAP} = \mathcal{E}(\theta|\mathbf{X})$.
- Estimativa $\widetilde{\theta}_{EAP} = \mathcal{E}(\theta|\mathbf{x})$.
- Estimador $\widehat{\theta}_{Mo} = \text{Moda}(\theta|\mathbf{X})$.
- Estimativa $\widetilde{\theta}_{Mo} = \text{Moda}(\theta|\mathbf{x})$.
- Estimador $\widehat{\theta}_{Md} = \text{Med}(\theta|\mathbf{X})$.
- Estimativa $\widetilde{\theta}_{Md} = \text{Med}(\theta|\mathbf{x})$.
- Medida de precisão do estimador (basicamente frequentista), $\mathcal{V}_{X|\theta}(\widehat{\theta})$.

Revisão de Cálculo de probabilidades

- Usaremos, em geral, p(.) para denotarmos fdp's.
- Sejam X, Y e Z va's em um mesmo espaço de probabilidade.

$$p(x|y) = \begin{cases} \dfrac{p(x,y)}{p(y)}, & \text{se } p(y) > 0, \\ 0, & \text{se } p(y) \end{cases}$$

$$p(x) = \begin{cases} \sum_y p(x,y), & \text{se y for discreto,} \\ \int_{\Omega_y} p(x,y)dy, & \text{se y for contínuo} \end{cases}$$

Revisão de Cálculo de probabilidades

- $E_X(X) = E_Y(E_{X|Y}(X|Y))$ e
  $\mathcal{V}_X(X) = \mathcal{E}_Y[\mathcal{V}_{X|Y}(X|Y)] + \mathcal{V}_Y[\mathcal{E}_{X|Y}[X|Y]]$.
- $p(x,y) = p(x|y)p(y) = p(y|x)p(x)$
- X e Y são independentes se e somente se $p(x,y) = p(x)p(y)$
  $(\rightarrow p(x|y) = p(x) e p(y|x) = p(y))$.
- X e Y são condicionalmente independentes dado Z se e somente se
  $p(x,y|z) = p(x|z)p(y|z)$
  $(\rightarrow p(x|y,z) = p(x|z) e p(y|x,z) = p(y|z))$.

Breve histórico

- Teorema de Bayes $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$, publicado em 1763, *An Essay towards solving a Problem in the Doctrine of Chances*.
- Fundamentos Bayesianos: Bruno de Finetti, Jim Berger entre outros.
- Existems algumas correntes de pensamento dentro da inferência Bayesiana:
  - Priori tem de ser completamente especificada usando conhecimento prévio.
  - Priori pode ser total ou parcialmente obtida através dos dados.

IF x IB

Aspecto	Inferência Frequentista	Inferência Bayesiana
Conhecimento:	verossimilhança	verossimilhança e priori
Parâmetros:	quantidades fixas	quantidades fixas mas de caráter aleatório ou totalmente aleatório
Teoria Assintótica	Fundamental	Pouco usada
Métodos computacionais	Extremamente dependente	Imprescindível na grande maioria das situações
Probabilidade:	limite de frequências relativas	medida de credibilidade

IF x IB

Aspecto	Inferência Frequentista	Inferência Bayesiana
Estimação pontual:	máxima verossimilhança, momentos, mínimos quadrados	quantis da posteriori e risco de Bayes
Estimação intervalar:	limites são aleatórios	parâmetro é aleatório
Teste de hipótese:	estatística, região crítica, poder, TUMP	Probabilidade das hipóteses serem verdadeiras