Distribuições à priori

Prof. Caio Azevedo

Motivação

- As distribuições à priori são de fundamental importância na inferência bayesiana.
- De fato, diferentes escolhas podem levar, algumas vezes, à resultados "significativamente" diferentes (para uma mesma verossimilhança).
- Contudo, existem alguns métodos para "construção", escolha e comparação de prioris.

Propriedades desejadas

- Fundamentalmente, uma priori deve apresentar as seguintes características:
  - Respeitar o espaço paramétrico.
  - Conduzir à uma posteriori própria (integrável).
  - Refletir, apropriadamente, o conhecimento de especialistas.
  - Conduzir à um processo de inferência com “boas propriedades”.
  - Não “dominar” a verossimilhança (à menos que exista uma contundente razão para isto).
- OBS: uma priori não, necessariamente, precisa ser uma fdp nem mesmo ser integrável.

Tipos de priori

- Quanto à propriedade (integrabilidade):
  - Própria: quando ela é integrável (ainda que não seja uma fdp).

    Exemplos:

$$\theta \sim gama(a,b); p(\theta) = I_{(0,1)}(\theta); p(\theta) = \frac{1}{\theta} I_{(a,b)}(\theta)$$

  - Imprópria: quando ela não é integrável. Exemplos:

$$p(\theta) \propto I_{(0,\infty)}(\theta); p(\theta) \propto I_{(a,\infty)}(\theta)$$

Tipos de priori (cont.)

- Quanto à depender ou não da amostra (verossimilhança) que se está analisando
  - "Subjetiva": não depende da amostra

$$\theta \sim N(a, b)$$

  - "Objetiva": depende da amostra. Exemplos: priori de Jeffreys, priori empírica, priori de referência (Berger e Bernardo).

Tipos de priori (cont.)

- Quanto ao nível de informação.
  - Não informativa: assumem ignorância total em relação ao parâmetro, ou seja, é proporcional à uma constante.

$$p(\theta) = I_{(0,1)}(\theta); p(\theta) = I_{(0,\infty)}(\theta)$$

  - Informativas: assumem algum grau de conhecimento acerca do parâmetro.
    - Pouco informativa ou vaga: $\theta \sim N(0, 10000)$.
    - Moderadamente informativa: $\theta \sim N(0, 100)$.
    - Muito informativa: $\theta \sim N(0, 1)$.

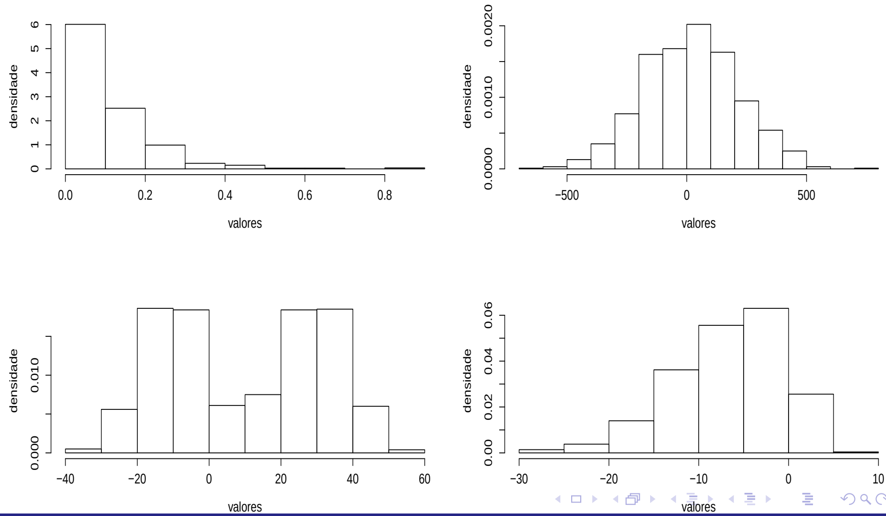
Observações importantes

- Prioris muito informativas podem conflitar com a verossimilhança.
- Prioris pouco informativas, ainda que levem à posterioris próprias, podem gerar uma certa instabilidade (quando se obtem a posteriori numericamente).
- Prioris impróprias, podem levar à posterioris impróprias. Neste caso, não se pode usar tais prioris. Ainda que a posteriori seja própria, a utilização de prioris impróprias pode comprometer o uso de certas metodologias bayesianas como o fator de Bayes.
- Prioris próprias, em geral, conduzem à posterioris próprias.
- Se a verossimilhança for integrável (uniformemente limitada) a posteriori será própria, ainda que a priori seja imprópria.

Método do histograma

- Consiste em construir um histograma usando informações sobre o problema como : os valores mais prováveis do parâmetros fornecidos por especialistas, ou fornecidos por um mesmo especialista (com as respectivas probabilidades) ou estimativas oriundas de estudos anteriores.
- Nesse caso, tal histograma fornece uma aproximação para distribuição à priori e ele pode ser usado como um indicador de uma possível distribuição.

Exemplos hipóteticos relativos ao método do histograma



Família conjugada

- Família conjugada: escolhe-se uma priori que leva à uma posteriori na mesma família. Exemplo: famílias conjugadas naturais.
- Em geral, as prioris dessas famílias dependem de hiperparâmetros.
- A escolha deles (hiperparâmetros) requer conhecimento prévio sobre o problema ou pode ser feita através da amostra.

Família conjugada (cont.)

- Para discutir a escolha dos hiperparâmetros, vamos considerar o Exemplo 4: verossimilhança de Poisson(λ) com priori conjugada ($gama(a, b^{-1})$).
- Se $\theta \sim gama(a, b^{-1})$, então $\mu = \mathcal{E}(\theta) = a/b$ e $\sigma^2 = \mathcal{V}(\theta) = a/b^2$. Podemos pedir ao pesquisador que fixe μ e σ^2 e, então, calculamos a e b a partir desses valores, ou seja:

$$a = \frac{\mu^2}{\sigma^2}; b = \frac{\mu}{\sigma^2}$$

Família conjugada (cont.)

- Podemos utilizar os dados (inferência bayesiana empírica) para obter os hiperparâmetros.
- Em nosso caso, note que

$$
\begin{aligned}
p(\mathbf{x}|\lambda)p(\lambda) &= \frac{e^{-n\lambda}\lambda^{n\bar{x}}}{\prod_{i=1}^{n} x_i!}\frac{b^a}{\Gamma(a)}e^{-b\lambda}\lambda^{a-1}\mathbb{1}_{(0,\infty)}(\lambda) \\
&= \frac{e^{-(n+b)\lambda}\lambda^{(n\bar{x}+a)-1}b^a}{\prod_{i=1}^{n} x_i!\Gamma(a)}\mathbb{1}_{(0,\infty)}(\lambda)
\end{aligned}
$$

Família conjugada (cont.)

- Assim, temos

$$p(\mathbf{x}|a,b) = \int_0^\infty p(\mathbf{x}|\lambda)p(\lambda)d\lambda = \frac{\Gamma(n\overline{x}+a)b^a}{(n+b)^{n\overline{x}+a}\Gamma(a)\prod_{i=1}^n x_i!}$$

- Assim, eliminou-se λ da função acima, originando uma espécie de verossimilhança para os hiperparâmetros. Portanto, podemos obter as estimativas de MV para (a,b) e usá-las na priori.

Família conjugada (cont.)

- Podemos, ainda, tratar os hiperparâmetros (a,b) como parâmetros de interesse, atribuir prioris para eles, e estimá-los através de suas posteriores.
- Essa estrutura é chamada de bayesiana hierárquica. Em nosso caso, podemos considerar:

$$\begin{aligned} X_i|\lambda &\overset{i.i.d}{\sim} Poisson(\lambda) \\ \lambda &\sim gama(a, b^{-1}) \\ a &\sim gama(c_1, d_1) \\ b &\sim gama(c_2, d_2) \end{aligned}$$

com (c_1, d_1, c_2, d_2) conhecidos.

Família conjugada (cont.)

- Neste caso, (c_1, d_1, c_2, d_2) passam a ser os hiperparâmetros e nosso objetivo é encontrar as posteriores marginais, ou seja:

$$p(\lambda|\mathbf{x}), p(a|\mathbf{x}), p(b|\mathbf{x})$$

- Na grande maioria dos caros, tais posteriores não apresentam forma analítica conhecida, e precisamos usar métodos numéricos para obter aproximações delas.
- Mesmo em casos em que se utiliza prioris que não correspondem à família conjugada, podemos utilizar as idéias apresentadas anteriormente para obter/escolher os hiperparâmetros.

Prioris não-informativas

- Se não dispomos de informações que possam nos levar à escolha de uma priori adequada, podemos utilizar prioris não informativas (proporcionais à uma constante).
- Uma outra forma, é obter as prioris de Jeffreys (PJ) ou de Jeffreys sob independência (PJI). Em geral, estes dois tipos de prioris são vagas ou pouco informativas.
- A PJ e a PJI são obtidas através da informação de Fisher ($I(\theta)$).
- Lembrando: quanto maior for a Informação de Fisher menor a variância do estimador de MV.

Prioris de Jeffreys

- Então, pensando na IF como uma distribuição de probabilidade, quanto maior a IF num determinado intervalo, maior a probabilidade (a priori) do parâmetro pertencer à este intervalo.
- Seja $\boldsymbol{\theta} = (\theta_1, .., \theta_k)$ um vetor de parâmetros e $I(.)$ a informação de Fisher obtida à partir de $p(\mathbf{x}|\boldsymbol{\theta})$. A PJ é definida por

$$p^J(\boldsymbol{\theta}) \propto |I(\boldsymbol{\theta})|^{1/2}$$

Prioris de Jeffreys (cont.)

- A priori de Jeffreys sob independência (PJI) consiste em se considerar os parâmetros como independentes (à priori) e obter a priori de Jeffreys para cada um deles. A PJI será o produtório destas prioris(de acordo com a suposição de independência desejada).
- Suponha que desejamos considerar os parâmetros independentes à priori. Neste caso, teremos:

$$p^{I}(\boldsymbol{\theta}) = \prod_{i=1}^{k} p^{J}(\theta_i)$$

Prioris de Jeffreys (cont.)

- Consideremos o Exemplo 4 (Poisson (λ)). Nesse caso, temos que $I(\lambda) = \frac{n}{\lambda}$, logo

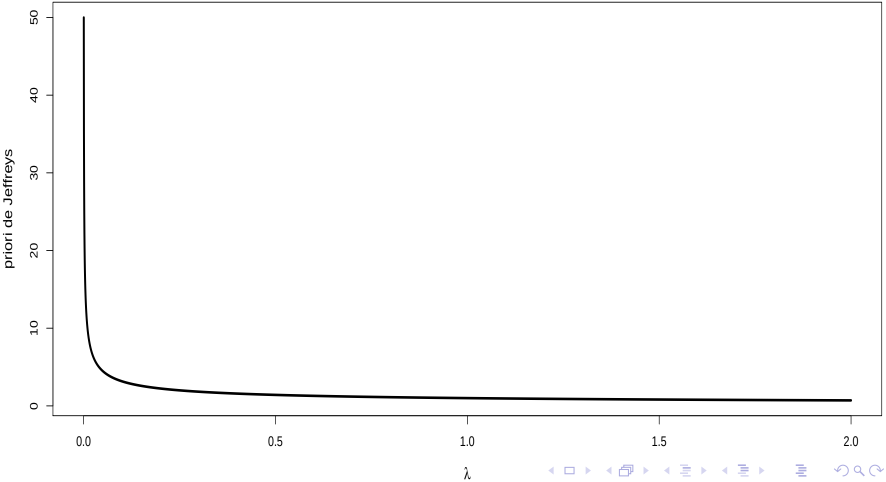
$$p^J(\lambda) \propto \lambda^{-1/2} I_{(0,\infty)}(\lambda)$$

a qual é imprópria. Por outro lado, a posteriori é dada por

$$p^J(\lambda|\mathbf{x}) \propto e^{-n\lambda} \lambda^{n\overline{x}-1/2-1} I_{(0,\infty)}(\lambda)$$

ou seja, $\lambda|\mathbf{x} \sim gama(n\overline{x} - 1/2, n^{-1})$, a qual é própria.

Priori de Jeffreys para o modelo de Poisson



Prioris de Jeffreys (cont.)

- Consideremos o Exemplo 8 (N(μ, σ^2)) com ambos os parâmetros desconhecidos. Nesse caso, temos que

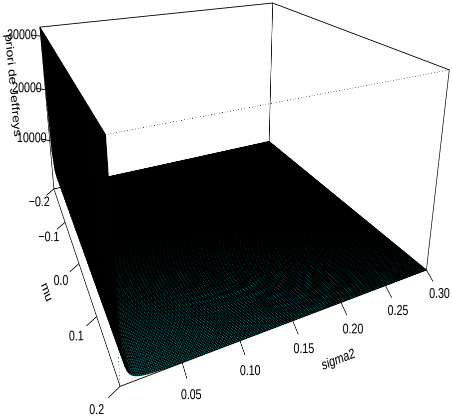
$$I(\boldsymbol{\theta}) = \begin{bmatrix} \dfrac{n}{\sigma^2} & 0 \\ 0 & \dfrac{n}{(\sigma^2)^2} \end{bmatrix}$$

- Assim,

$$p^J(\boldsymbol{\theta}) \propto \left(\sigma^2\right)^{-3/2} I_{(-\infty,\infty)}(\mu) I_{(0,\infty)}(\sigma^2)$$

a qual é imprópria.

Priori de Jeffreys para o modelo $N(\mu, \sigma^2)$



Prioris de Jeffreys (cont.)

- Contudo, note que

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{x}) &\propto exp\left\{-\frac{n\left(\mu-\overline{x}\right)^2+(n-1)s^2}{2\sigma^2}\right\}\left(\sigma^2\right)^{-(\frac{n}{2}+1)-\frac{1}{2}} \\ &\times I_{(-\infty,\infty)}(\mu)I_{(0,\infty)}(\sigma^2) \end{aligned}$$

a qual correspondes ao núcleo de uma distribuição $N-IG(\overline{x},n,n/2,(n-1)s^2/2)$, a qual é própria. Assim, $\mu|\mathbf{x}\sim t_{(n)}(\overline{x},\sqrt{(n-1)s^2}/n)$ e $\sigma^2|\mathbf{x}\sim IG(n/2,(n-1)s^2/2)$