Prioris, Prioris Conjugadas, Prioris não informativas e Distribuições Preditivas

Comentários Gerais sobre Prioris

- Prioris e posterioris não são absolutos. Podem mudar de acordo com a informação disponível ao analista.

- Suponha uma observação (x, y) com verossimilhança $p(x, y \mid \theta) = p(x \mid \theta)\, p(y \mid x, \theta)$.

  - $p(\theta \mid x) \propto p(\theta)\, p(x \mid \theta)$
  - $p(\theta \mid x, y) \propto p(\theta)\, p(x, y \mid \theta) \propto \{p(\theta)\, p(x \mid \theta)\}\, p(y \mid x, \theta)$

  Recomenda-se que prioris sejam fundamentadas em observações de experimentos anteriores.

  A melhor priori a usar pode mudar dependendo do problema sob estudo.

- Visto que $p(\theta \mid x) \propto p(\theta)\, p(x \mid \theta)$, devemos concentrar o esforço para especificar *elucidar / elicitate* a priori na região onde (esperamos que) a verossimilhança vai ser significativamente maior do que zero.

Questiono essa última afirmação. (Pessoa procurando chaves)

Famílias Conjugadas

- Exponential family: $p(x \mid \theta) \propto c(\theta)\ \exp\{\sum_{j=1}^{p} T_j(x)\eta_j(\theta)\}$;
- Uma priori da forma $p(\theta) \propto b(\theta)\ \exp\{\sum_{j=1}^{p} \alpha_j \eta_j(\theta)\}$ vai ser conjugada;

Usualmente chama de família conjugada natural.

- Beta-Binomial, Multinomial-Dirichlet, Poisson-Gama, Exponencial-Gama (mais geral: Gama com α fixo-Gama); Normal-Normal (variância conhecida), Normal-Normal-Gama (precisão desconhecida) etc.
- Priori conjugada fora de famílias exponenciais: Uniforme(0,θ)-Pareto

Inferência para a média de uma Normal com variância conhecida

$$p(x_1, \ldots, x_n \,|\, \mu) \propto \exp\{-\sum_{i=1}^{n}(x_i - \mu)^2/2\sigma^2\} \qquad p(\bar{x} \,|\, \mu) \propto \exp\{\frac{-n}{2\sigma^2}(\bar{x} - \mu)^2\}$$

Considere $X_1, \ldots, X_n \,|\, \mu \overset{iid}{\sim} \mathrm{N}(\mu,\sigma^2)$ com σ^2 conhecida e suponha que a priori $\mu \sim \mathrm{N}(\mu_0,\tau^2)$. Para calcular a distribuição a posteriori o seguinte Lema será muito útil.

Um Lema útil

Lema 1. *Seja Y uma variável aleatória que toma valores em toda a reta real tal que a sua densidade é $f(y) = k\,e^{-Q(y)}$, onde k é uma constante que não depende de y e $Q(y) = ay^2 + by + c$ $(a > 0)$ é uma forma quadrática em y. Logo, Y segue uma distribuição Normal com média e variância*

$$\mathbb{E}(Y) = -\frac{b}{2a} = -\frac{Q'(0)}{Q''(0)} \quad e \quad \mathrm{Var}(Y) = \frac{1}{2a} = \frac{1}{Q''(0)}\,.$$

Demonstração: Primeiro complete o quadrado.

$$ay^2 + by + c = a\left(y + \frac{b}{2a}\right)^2 + c - \frac{b^2}{4a}\,,$$ assim $$f(y) = k' \exp\left\{-a\left(y + \frac{b}{2a}\right)^2\right\}\,,$$

onde $k' = k\exp\{c - b^2/(4a)\}$ continua sendo uma constante que não depende de y. Dessa forma, a densidade de y é proporcional a uma densidade Normal com média $\mu = -\frac{b}{2a}$ e variância $\sigma^2 = \frac{1}{2a}$ e, como é uma densidade, a constante de proporcionalidade tem que ser igual a um. $\square$

Usando o lema para calcular a posteriori

Retornando ao problema do início desta Seção, para achar a distribuição a posteriori note que (i) $p(\mu) \propto \exp\{-\frac{1}{2\,\tau^2}\,(\mu - \mu_0)^2\}$ e (ii) como $\bar{x} = \sum_{i=1}^{n} x_i$ é um estatístico suficiente e $\bar{x}\,|\,\mu \sim \mathrm{N}(\mu,\sigma^2/n)$, $p(x_1,\ldots,x_n\,|\,\mu) \propto p(\bar{x}\,|\,\mu) \propto \exp\{-\frac{n}{2\,\sigma^2}\,(\bar{x}-\mu)^2\}$ (aqui e em todo o exemplo "$\propto$" significa "proporcional a menos de uma constante que não depende de μ"). Logo,

$$
\begin{aligned}
p(\mu\,|\,x_1,\ldots x_n) = \frac{p(\mu)\,p(x_1,\ldots,x_n\,|\,\mu)}{\int_0^1 p(\mu)\,p(x_1,\ldots,x_n\,|\,\mu)\,d\mu} &\propto p(\mu)\,p(x_1,\ldots,x_n\,|\,\mu) \propto p(\mu)\,p(\bar{x}\,|\,\mu) \\
&\propto \exp\left\{-\frac{1}{2\,\tau^2}\,(\mu-\mu_0)^2\right\}\,\exp\left\{-\frac{n}{2\,\sigma^2}\,(\bar{x}-\mu)^2\right\} \\
&\propto \exp\left\{-\left[\frac{1}{2\,\tau^2}\,(\mu-\mu_0)^2 + \frac{n}{2\,\sigma^2}\,(\bar{x}-\mu)^2\right]\right\}\,. \quad (1)
\end{aligned}
$$

Usando o lema (continuição)

Definindo $Q(\mu) = \frac{1}{2\,\tau^2}\,(\mu - \mu_0)^2 + \frac{n}{2\,\sigma^2}\,(\bar{x} - \mu)^2$ e usando o Lema 1, segue que a distribuição a posteriori é Normal com média e variâncias dadas por

$$\mu^* = \mathbb{E}(\mu \,|\, x_1, \ldots, x_n) = -\frac{Q'(0)}{Q''(0)} = \frac{\frac{1}{\tau^2}\,\mu_0 + \frac{n}{\sigma^2}\,\bar{x}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} \tag{2}$$

e

$$\tau^{*2} = \mathrm{Var}(\mu \,|\, x_1, \ldots, x_n) = \frac{1}{Q''(0)} = \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}\,. \tag{3}$$

Interpretação pelas precisões

$$\mu^* = \mathbb{E}(\mu \,|\, x_1, \ldots, x_n) = -\frac{Q'(0)}{Q''(0)} = \frac{\frac{1}{\tau^2}\,\mu_0 + \frac{n}{\sigma^2}\,\bar{x}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} \qquad (2)$$

$$\tau^{*2} = \mathrm{Var}(\mu \,|\, x_1, \ldots, x_n) = \frac{1}{Q''(0)} = \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}\,. \qquad (3)$$

Algumas vezes na literatura Bayesiana a variância da distribución a priori é parametrizada pelo recíproco $\phi = \tau^{-2}$, denominada de *precisão*. De forma semelhante, a *precisão* da distribução amostral de $\bar{x}$ é $[\mathrm{Var}(\bar{x} \,|\, \mu)]^{-1} = n/\sigma^2$. Com essa nova terminologia, a equação (3) diz simplesmente que "*a precisão a posteriori* $[\mathrm{Var}(\mu \,|\, x_1, \ldots, x_n)]^{-1}$ *é igual a precisão a priori* τ^{-2} *mais a precisão amostral* n/σ^2". De forma semelhante, a equação (2) diz que "*a média a posteriori é uma média ponderada da média a priori* μ_0 *e da média amostral* $\bar{x}$ *com pesos proporcionais as respectivas precisões*". Por exemplo, quanto maior for o tamanho amostral n, se τ^2 e σ^2, mais perto de $\bar{x}$ vai estar a média a posteriori.

As balanças incertas

Exemplo 1. Um objeto foi pesado $n = 10$ vezes numa balança retornando média $\bar{x} = 3.53$g. Segundo o fabricante da balança, para esse tipo de objetos a balança é muito precisa e o desvio padrão das pesadas deve ser da ordem de $\sigma = 0.2$g.

Uma outra balança, menos precisa, também estava disponível e nela o objeto foi pesado $m = 50$ vezes retornando média e desvio padrão iguais a $\bar{y} = 3.25$ e $s_y = 0.5$g.

Construa um modelo para analisar esses dados e com base nele obtenha um intervalo que contem μ com probabilidade a posteriori igual a 95%.

As balanças (continuação)

Podemos usar a informação da balança menos precisa para justificar uma distribuição a priori $\mu \sim$ Normal(3.25,0.5^2) (formalmente, a distribuição a posteriori resultante de usar uma priori não informativa com os dados da balança menos precisa seria t de Student com 49 graus de liberdade—veja as Seções 1.2.1 e 1.2.2 abaixo, mas como n é grande, a distribuição Normal é uma boa aproximação). Logo, podemos tomar $\mu_0 = \bar{y} = 3.25$ e $\tau^2 = s_y^2 = 0.5^2$. Usando o valor de $\bar{x} = 3.53$ usamos as equações (2) e (3) para calcular

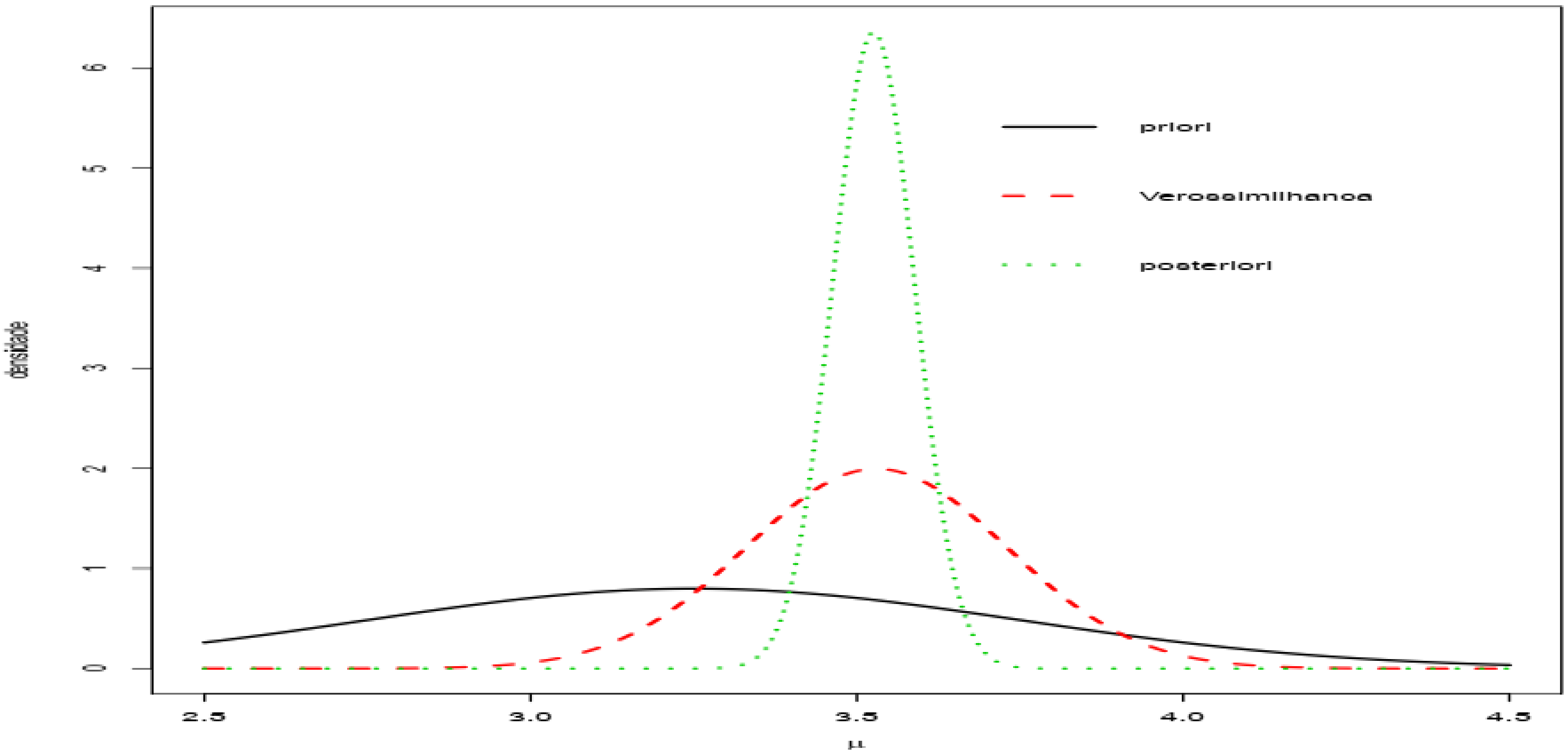
$$\mu^* = \frac{\frac{1}{0.5^2}(3.25) + \frac{10}{0.2^2}3.53}{\frac{1}{0.5^2} + \frac{10}{0.2^2}} \doteq 3.5256 \qquad \tau^{*2} = \frac{1}{\frac{1}{0.5^2} + \frac{10}{0.2^2}} \doteq (0.0627)^2 .$$

Assim a posteriori, $\mu \,|\, x_1, \ldots x_n \sim$ Normal(3.5256,0.0627^2). O intervalo desejado é então

$$\mu^* \pm z_{\alpha/2}\,\tau^* \doteq 3.5256 \pm (1.96)\,(0.0627) \doteq (3.4026; 3.6486) .$$

As balanças (conclusão)

$n/\sigma^2 = 250 \qquad 1/\tau^2 = 4$



Prioris "não informativas"

Na situação descrita nesta Seção tem duas formas de pensar numa distribuição a priori *pouco informativa*. Uma, é fazer a variância τ^2 da distribuição a priori muito grande. Em particular, quando $\tau^2 \to \infty$, tomando limites nas equações (2) e (3), vemos que a distribuição a posteriori tende a uma Normal com média $\bar{x}$ e variância σ^2/n, que depende somente da verossimilhança (veja que o valor de μ_0 também some uma vez que fazemos $\tau^2 \to \infty$). Outra forma de proceder é pensar numa "distribuição uniforme" para μ, algo assim como especificar em (1) $p(\mu)$ igual a uma constante $c > 0$ (como a constante c no numerador cancela com a do denominador, usualmente escreve-se $p(\mu) \propto 1$). Porém, a distribuição uniforme não está definida num intervalo infinito como a reta real ou, em outras palavras, não é possível normalizar a densidade para que integre um pois $\int_{-\infty}^{\infty} c\, d\mu = \infty$ (lembre da discussão sobre $w(\theta)$ e $w^*(\theta) = w(\theta)/\int w(\theta)\, d\theta$ na Seção **??**), de forma que essa possibilidade está, ao menos formalmente, fora do paradigma Bayesiano.

Prioris "não informativas" (continuação)

Ainda, é possível pôr $p(\mu) \propto 1$ na equação (1) e, formalmente, fazer as mesmas contas que faríamos com uma distribuição a priori "*verdadeira*". Se o fizermos, chegaríamos a que

$$p(\mu \,|\, x_1, \ldots, x_n) \propto p(x_1, \ldots, x_n \,|\, \theta) \propto p(\bar{x} \,|\, \theta) \propto \exp\{-\frac{n}{2\sigma^2}(\mu - \bar{x})^2\}\,,$$

o que implica que $\mu \,|\, x_1, \ldots, x_n \sim \text{Normal}(\bar{x}, \sigma^2/n)$, que é o mesmo resultado ao qual chegamos fazendo $\tau^2 \to \infty$.

Prioris "não informativas" (continuação)

Na literatura, quando se usa uma "distribuição a priori" que, como no caso anterior, tem integral infinita, fala-se de uma "*distribuição a priori* **imprópria**". As aspas foram usadas acima pois, pelo menos na opinião do autor destas notas, distribuições a priori impróprias não são distribuições a priori. De qualquer forma, distribuições impróprias ou com variância infinita podem ser apropriadas em situações nas quais deseja-se usar uma distribuição a priori relativamente "achatada" (*flat*) na região onde a verossimilhança é significativamente diferente de zero. No caso particular desta seção, a verossimilhança $p(x_1, \ldots, x_n \mid \mu) \propto \exp\{-n(\mu - \bar{x})^2\}$ comporta-se semelhante a uma densidade Normal e decresce muito rapidamente a zero quando μ se afasta de $\bar{x}$. Na maioria das situações práticas, o valor da densidade a priori $p(\mu)$ para valores de μ que distam de $\bar{x}$ (digamos) mais do que 3 ou 4 vezes o desvio padrão $\sigma/\sqrt{n}$ é irrelevante. Nessas situações uma distribuição a priori imprópria $p(\mu) \propto 1$ pode ser uma aproximação conveniente e mais fácil de especificar que uma distribuição própria.

Prioris "não informativas" (conclusão)

Seja fazendo $\tau^2 \to \infty$ ou usando $p(\mu) \propto 1$, a seguinte observação merece ser feita. Uma vez que chegamos no resultado $\mu \,|\, x_1, \ldots, x_n \sim \text{Normal}(\bar{x}, \sigma^2/n)$, se denotarmos por $z_{\alpha/2}$ o quantil $(1-\alpha/2)$ da distribuição Normal Padrão, podemos escrever que

$$\mathbb{P}\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \;\Big|\; x_1, \ldots, x_n\right) = 1 - \alpha\,. \tag{4}$$

Em comparação, na estatística clássica, quando calcula-se o intervalo de confiança usual nesta situação, tem-se que

$$\mathbb{P}_\mu\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha\,. \tag{5}$$

Note a diferença na interpretação dessas duas afirmações. Em (4), a variável aleatória é μ, enquanto $\bar{x}$ está fixa no valor que foi efetivamente observado. Em (5), a variável aleatória é $\bar{X}$, enquanto μ está fixa num valor desconhecido.

Distribuições Preditivas

$X_1, \ldots, X_n, X_{n+1}, \ldots, X_{n+m} \mid \mu \stackrel{iid}{\sim} \text{Normal}(\mu,\sigma^2)$ com a distribuição a priori $\mu \sim \text{Normal}(\mu_0,\tau^2)$. Com base na observação de $\bar{x}_{1:n} = n^{-1} \sum_{i=1}^{n} x_i$ desejamos prever $\bar{X}_{(n+1):(n+m)} = m^{-1} \sum_{i=n+1}^{m+n} X_i$ ou, em outras palavras, procuramos a distribuição condicional de $\bar{X}_{(n+1):(n+m)}$ dado $\bar{x}_{1:n}$. (se for desejado somente a distribuição preditiva de uma única observação futura X_{n+1}, tome $m = 1$ nos resultados a seguir).

Distribuições Preditivas (continuação)

Essa distribuição pode ser achada diretamente da distribuição conjunta de $(\theta, \bar{X}_{1:n}, \bar{X}_{(n+1):(n+m)})$ integrando primeiro com respeito a θ e depois usando a definição de densidade condicional. Neste caso, porém, é mais fácil proceder da seguinte forma. Primeiro, veja que a distribuição conjunta do vetor $(X_1, \ldots, X_{n+m}, \theta)$ é Normal $(n + m + 1)$-variada. Logo, todas as distribuições marginais e condicionais também serão Normais multivariadas. Em particular, $X_{n+1}, \ldots, X_{n+m} \mid \bar{x}_{1:n}$ segue uma distribuição Normal m-variada e, sendo $\bar{X}_{(n+1):(n+m)}$ uma combinação linear de $X_{n+1}, \ldots, X_{n+m}$, concluimos que a distribuição preditiva de $\bar{X}_{(n+1):(n+m)}$ dado $\bar{x}_{1:n}$ deve ser Normal univariada. Assim, é suficiente achar os dois primeiros momentos da distribuição preditiva, o que pode ser feito facilmente com as propriedades (**??**)–(**??**). Isto é, lembrando que os dois primeiros momentos da distribuição a posteriori de $\mu \mid \bar{x}_{1:n}$ são dados pelas equações (2)–(3),

Distribuições Preditivas (continuação)

$$\begin{aligned}\mathbb{E}(\bar{X}_{(n+1):(n+m)}\,|\,\bar{x}_{1:n}) &= \mathbb{E}_{\mu\,|\,\bar{x}_{1:n}}\,\mathbb{E}(\bar{X}_{(n+1):(n+m)}\,|\,\bar{x}_{1:n},\mu)\\ &= \mathbb{E}_{\mu\,|\,\bar{x}_{1:n}}\,\mathbb{E}(\bar{X}_{(n+1):(n+m)}\,|\,\mu) = \mathbb{E}_{\mu\,|\,\bar{x}_{1:n}}(\mu) = \frac{\frac{1}{\tau^2}\,\mu_0 + \frac{n}{\sigma^2}\,\bar{x}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}} = \mu^*\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(\bar{X}_{(n+1):(n+m)}\,|\,\bar{x}_{1:n}) &= \mathbb{E}_{\mu\,|\,\bar{x}_{1:n}}\,\mathrm{Var}(\bar{X}_{(n+1):(n+m)}\,|\,\bar{x}_{1:n},\mu) + \mathrm{Var}_{\mu\,|\,\bar{x}_{1:n}}\,\mathbb{E}(\bar{X}_{(n+1):(n+m)}\,|\,\bar{x}_{1:n},\mu)\\ &= \mathbb{E}_{\mu\,|\,\bar{x}_{1:n}}\,\mathrm{Var}(\bar{X}_{(n+1):(n+m)}\,|\,\mu) + \mathrm{Var}_{\mu\,|\,\bar{x}_{1:n}}\,\mathbb{E}(\bar{X}_{(n+1):(n+m)}\,|\,\mu)\\ &= \mathbb{E}_{\mu\,|\,\bar{x}_{1:n}}\left(\frac{\sigma^2}{m}\right) + \mathrm{Var}_{\mu\,|\,\bar{x}_{1:n}}(\mu) = \frac{\sigma^2}{m} + \frac{1}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}} = \frac{\sigma^2}{m} + \tau^{*2}\,.\end{aligned}$$

Veja que está última variância reflete duas fontes de incerteza com respeito a $\bar{X}_{(n+1):(n+m)}$: o primeiro termo σ^2/m tem a ver com a variabilidade inerente de $\bar{X}_{(n+1):(n+m)}$, enquanto o segundo termo τ^{*2} reflete a incerteza com respeito à média μ.