21073 - Introdução às probabilidades e estatística bayesianas

Ano lectivo 2015/16

Docente: António Araújo

e-fólio B (8 a 15 de Janeiro)

,

Para a resolução do e-fólio, aconselha-se que:

- Verifique se o ficheiro que recebeu está correcto. O e-fólio consiste de 2 páginas com 3 problemas e termina com a palavra FIM.
- Como o e-fólio tem um tempo prolongado de resolução, espera-se que as respostas que enviar estejam legíveis, com boa apresentação e organização. Deve fazer à parte o trabalho auxiliar e enviar apenas uma versão final, "limpa". Deve digitalizar a sua resolução de forma legível, ou executá-la directamente em formato digital (aceita-se word, pdf, ou scans em jpeg, png ou tiff - se usar varios ficheiros envie apenas um arquivo com todos eles, em rar ou zip). Respostas ilegíveis não serão cotadas, por isso verifique bem o seu ficheiro antes de enviar.
- Justifique cuidadosamente todas as suas respostas. Apresente todos os cálculos que julgue necessários para a compreensão do seu raciocínio.
- Tenha em atenção o prazo de entrega do e-fólio e as indicações para submeter a resolução disponibilizadas na sala de aulas virtual.
- O e-fólio é um trabalho individual. Pode utilizar recursos externos (pesquisa online, literatura, etc) mas não pode pedir ajuda a terceiros nem discutir os problemas com os seus colegas.

Critérios de avaliação e cotação:

- Este e-fólio tem a cotação total de 4 valores, assim distribuídos: Cada questão vale 4/3 de valor.

Por favor preencha os seus dados:

- Nome:
- B.I:
- Nº de Estudante'
- Curso:

GRUPO I

Problema 1. *Considere o seguinte modelo em que $\lambda > 0$ é um parâmetro real desconhecido:*

$$p(x|\lambda) = \begin{cases} \dfrac{\lambda 2^{\lambda}}{x^{\lambda+1}}, & para\ x \geq 2 \\ 0, & para\ x < 2 \end{cases}$$

a) Mostre que a família $Gama(\alpha, \beta)$ é família conjugada de priors para este modelo, e obtenha a regra de actualização dos parâmetros α, β.

Nota: Recorde que

$$Gama(\lambda|\alpha, \beta) \propto \lambda^{\alpha-1} e^{-\beta\lambda}.$$

b) Suponha que λ tinha a distribuição prévia $p(\lambda) = Gamma(2.5, 0.8)$. Observou-se a amostra

$$X = \{46, 4, 16, 621, 96\}$$

Calcule a distribuição posterior de λ. Represente graficamente as distribuições prévias e posteriores (usando por exemplo a linguagem R). Calcule os seus valores médios e variâncias.

Solução

a)

Seja $\underline{x}$ o vector dos dados observados.

$$
\begin{aligned}
P(\lambda|\underline{x}) \propto p(\underline{x}|\lambda)p(\lambda|\alpha,\beta) &= \frac{\lambda^n 2^{n\lambda}}{\prod_i x_i^{\lambda+1}} \cdot \lambda^{\alpha-1}e^{-\beta\lambda} \\
&= \frac{\lambda^{n+\alpha-1}}{\prod_i x_i^{\lambda+1}} 2^{n\lambda} e^{-\beta\lambda} \\
&= \frac{\lambda^{n+\alpha-1}}{e^{\log(\prod_i x_i^{\lambda+1})}} e^{log(2^{n\lambda})} e^{-\beta\lambda} \\
&= \frac{\lambda^{n+\alpha-1}}{e^{(\lambda+1)\sum_i \log(x_i)}} e^{n\lambda\log(2)} e^{-\beta\lambda} \\
&= \lambda^{n+\alpha-1} exp\left(-(\lambda+1)\sum_i \log(x_i) + n\lambda\log(2) - \beta\lambda\right) \\
&= \lambda^{n+\alpha-1} exp\left(-\lambda\left(\sum_i \log(x_i) - n\log(2) + \beta\right)\right)\exp\left(-\sum_i \log(x_i)\right) \\
&\propto \lambda^{n+\alpha-1} exp\left(-\lambda\left(\sum_i \log(x_i) - n\log(2) + \beta\right)\right) \\
&\propto \lambda^{n+\alpha-1} exp\left(-\lambda\left(\sum_i \log(\frac{x_i}{2}) + \beta\right)\right)
\end{aligned}
$$

Sendo que esta expressão final é de novo o núcleo de uma Beta, demonstrámos o que se pretendia, e verificámos que a regra de actualização dos parâmetros é

$$(\alpha,\beta) \mapsto \left(\alpha+n, \beta+\sum_i \log\left(\frac{x_i}{2}\right)\right)$$

b) Basta aplicar a fórmula da alínea anterior:

$$(\alpha,\beta) \mapsto \left(\alpha+n, \beta+\sum_i \log\left(\frac{x_i}{2}\right)\right)$$

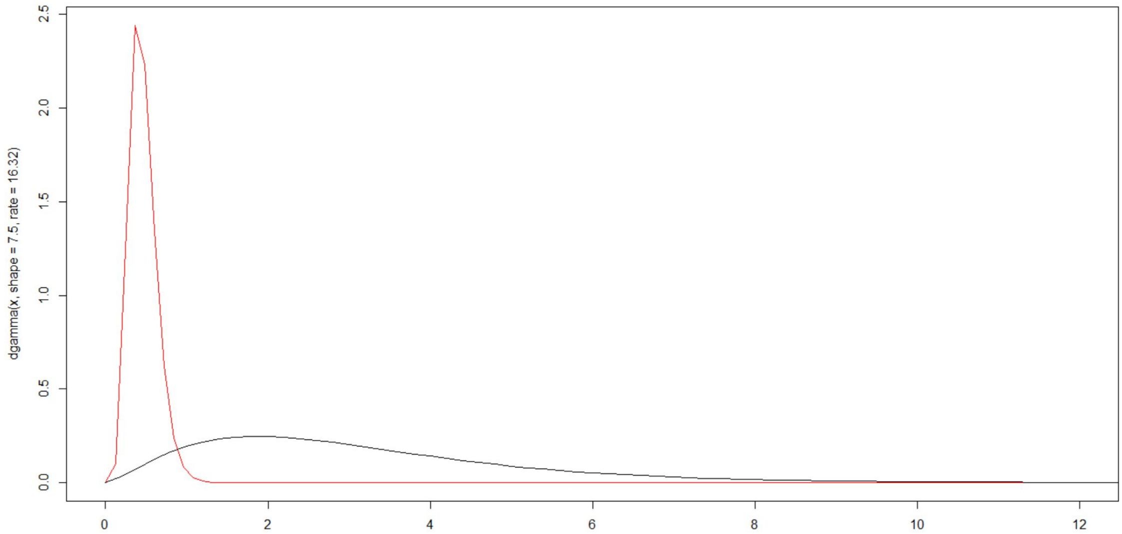
que fica

$$Gama(2.5, 0.8) \mapsto Gama\,(2.5+5, 0.8+15.52) = Gama\,(7.5, 16.32)$$

Para obter um gráfico do prior e do posterior basta aplicar os comandos de R seguintes, que foram descritos nas notas de apoio (com a simples mudança de distribuição da Beta para a Gama):

curve(dgamma(x,shape=7.5,rate=16.32),from=0, to=12 ,col="red")

curve(dgamma(x,shape=2.5,rate=0.8),from=0, to=20 ,col="black",add=TRUE) que resultam no gráfico seguinte (prior a preto, posterior a vermelho):



Para obter a média e desvio padrão de uma $Gama(x|\alpha, \beta)$ são conhecidas as fórmulas

$$E = \frac{\alpha}{\beta}, \quad V = \frac{\alpha}{\beta^2}$$

que aplicadas ao prior $Gama(2.5, 0.8)$ dão

$$E = 3.13, \quad V = 3.90$$

e aplicadas ao posterior $Gama(7.5, 16.32)$ dão

$$E = 0.46, \quad V = 0.03$$

É evidente no gráfico a clara diminuição da dispersão com a actualização dos parâmetros devida às observações.

Problema 2. *A probabilidade de uma certa peça ser fabricada com defeito segue um modelo binomial $X \sim B(\theta, n)$ em que o prior $p(\theta)$ é uma mistura convexa de Betas da forma*

$$p(\theta) = 0.1 Beta(6, 5) + 0.9 Beta(6, 2)$$

Diga qual é a distribuição posterior de θ se observar uma sequência de 30 peças e 20 estiverem avariadas.

Solução:
O posterior é

$$
\begin{aligned}
p(\theta|x=20,n=30) &\propto p(x=20|\theta,n=30)p(\theta) \\
&\propto \theta^{20}(1-\theta)^{10}\left(0.1Beta(6,5)+0.9Beta(6,2)\right) \\
&\propto \theta^{20}(1-\theta)^{10}\left(0.1\frac{\theta^5(1-\theta)^4}{B(6,5)}+0.9\frac{\theta^5(1-\theta)^1}{B(6,2)}\right) \\
&\propto 0.1\frac{\theta^{25}(1-\theta)^{14}}{B(6,5)}+0.9\frac{\theta^{25}(1-\theta)^{11}}{B(6,2)} \\
&\propto 0.1\frac{B(26,15)}{B(6,5)}\frac{\theta^{25}(1-\theta)^{14}}{B(26,15)}+0.9\frac{B(26,12)}{B(6,2)}\frac{\theta^{25}(1-\theta)^{11}}{B(26,12)} \\
&\propto 0.1\frac{B(26,15)}{B(6,5)}Beta(26,15)+0.9\frac{B(26,12)}{B(6,2)}Beta(26,12) \\
&= wBeta(26,15)+(1-w)Beta(26,12)
\end{aligned}
$$

onde $w = \dfrac{0.1\dfrac{B(26,15)}{B(6,5)}}{0.1\dfrac{B(26,15)}{B(6,5)}+0.9\dfrac{B(26,12)}{B(6,2)}} = 0.11 \approx 0.1$

Então

$$
p(\theta|x=4,n=10) = 0.1Beta(26,15)+0.9Beta(26,12)
$$

Nota 1: Lembrar que a função Beta é $B(a,b) = \dfrac{(a-1)!(b-1)!}{(a+b-1)!}$ quando a e b são inteiros positivos.

Nota 2: Na linguagem R a função Beta está disponível através do comando $beta(a,b)$. Não confundir com a distribuição Beta, disponível através do comando dbeta.

Problema 3. *Suponha que a posição de um ponto material sobre uma recta infinita é descrita por uma variável aleatória normal com variância conhecida,* $X = N(\mu, \sigma^2 = 9)$*. Suponha que tudo o que sabe sobre o prior* $p(\mu)$ *é que é uma distribuição com média* $\mu_0 = 0$ *e variância* $\sigma_0^2 = 16$*.*

a) Diga qual a forma do prior $p(\mu)$ *que escolheria, e porquê (tendo em conta que quer assumir o mínimo possível).*

b) Actualize o prior obtido em a) tendo em conta que observa a amostra $\overline{X} = \{4, 4.8, 3.7, 2.5, 3.5, 5.1\}$*.*

Solução:

a) O prior deverá ser normal, porque essa é a distribuição que minimiza a entropia quando apenas se conhece o valor da média e da variância. Nomeadamente, é a normal $N(\mu = 0, \sigma^2 = 16)$. Além disso, a família normal é conjugada de si própria.

b) Na fórmula

$$a' = \frac{\frac{1}{b^2}a + \frac{n}{\sigma^2}\overline{x}}{\frac{1}{b^2} + \frac{n}{\sigma^2}}, \quad b'^2 = \frac{1}{\frac{1}{b^2} + \frac{n}{\sigma^2}}$$

substituimos $\overline{x} = \sum x_i/n = 3.94$, $a = 0$, $b^2 = 16$, $\sigma^2 = 9$, $n = 6$, e obtemos que a distribuição posterior é $p(\theta|x, I) = N(a' = 3.6\,, b'^2 = 1.4)$. Como a distribuição é normal, basta ler numa tabela o intervalo de credibilidade. Uma região de credibilidade posterior para θ a 95% será portanto o intervalo $a' \pm 2b' = 3.6 \pm 2 \times 1.4 = [0.8, 6.4]$.

FIM