UNIVERSITY OF OSLO

Faculty of mathematics and natural sciences

Exam in:	STK4021/STK9021 — Applied Bayesian Analysis and Numerical Methods
Day of examination:	Wednesday 19th of December 2018
Examination hours:	09:00 – 13:00

This problem set consists of 4 pages.

Appendices:	None
Permitted aids:	One single sheet of paper with the candidate's own personal notes. Calculator.

Please make sure that your copy of the problem set is complete before you attempt to answer anything.

Solution

Problem 1 Poisson-Gamma

a

(i)

$$\theta \,|\, y_1 \sim \text{Gamma}(a + y_1, b + 1)$$

(ii)

$$\theta \,|\, \boldsymbol{y} \sim \text{Gamma}(a + n\bar{y}, b + n)$$

b

(i) $\text{E}(Y_i) = a/b \qquad \text{Var}(Y_i) = a(b+1)/b^2$
(ii) $\text{Cov}(Y_i, Y_j) = a/b^2 \qquad \text{Cor}(Y_i, Y_j) = 1/(b+1)$

c

$$f(y^* \,|\, \boldsymbol{y}) = \int_0^\infty f(y^* \,|\, \theta) p(\theta \,|\, \boldsymbol{y}) d\theta = \frac{1}{y^*!} \frac{\Gamma(a + n\bar{y} + y^*)}{\Gamma(a + n\bar{y})} \frac{(b+n)^{a+n\bar{y}}}{(b+n+1)^{a+n\bar{y}+y^*}}$$

To sample from predictive distribution:

- sample $\theta^{(j)}$ from posterior
- then sample observations from Poisson distribution with parameter $\theta^{(j)}$

(Continued on page 2.)

d

The Bayes estimator is the t minimising $g(t) = \mathrm{E}[L(\theta, t) \,|\, y]$ (the posterior loss). A bit of work gives $g(t) = \mathrm{E}[\theta \,|\, y] - 2t + 2t^2\mathrm{E}[1/\theta \,|\, y]$. Differentiating this wrt t and solving gives the right answer:

$$\widehat{\theta}_B = \left[\mathrm{E}\left(\frac{1}{\theta} \,|\, y\right)\right]^{-1}.$$

e

First integrate (and simplify) to find that $\mathrm{E}[1/\theta \,|\, y] = (b+n)/(a+n\bar{y}-1)$. Then we have

$$\widehat{\theta}_B = \frac{a+n\bar{y}-1}{b+n}$$

(some might note that this is the posterior mode). For the next subquestion it might be smart to express the estimator as,

$$\widehat{\theta}_B = \omega\frac{a-1}{b} + (1-\omega)\bar{y},$$

where $\omega = b/(b+n)$. (So a combination of prior mode and data mean).

f

I get

$$R(\widehat{\theta}_B, \theta) = \frac{(1-\omega)^2}{n} + \frac{\omega^2}{\theta}\left(\frac{a-1}{b} - \theta\right)^2,$$

and $R(\widehat{\theta}_F, \theta) = 1/n$ (a constant!). Bayes estimator is best when the true θ is reasonably close to the prior mode, i.e. when

$$\frac{1}{\theta}\left(\frac{a-1}{b} - \theta\right)^2 < \frac{1}{n} + \frac{2}{b}.$$

Problem 2 Seven exponential experiments

a

First we will consider each experiment separately.
(i) Here one should state the log-likelihood function for θ_i based on the data from experiment i, $\ell_{n_i}(\theta_i) = n_i \log\theta_i - \theta_i n_i \bar{y}_{i\cdot}$. Then differentiate wrt to θ_i and solve: $\widehat{\lambda}_{ML,i} = 1/\bar{y}_{i\cdot}$.
(ii)

$$\lambda_i \,|\, \boldsymbol{y_i} \sim \mathrm{Gamma}(\alpha + n_i, \beta + n_i\bar{y}_{i\cdot}).$$

(iii) Here it is sufficient to refer to the course: the Bayes estimate under quadratic loss is the posterior expectation. Then use formula for the expectation of a gamma:

$$\widehat{\lambda}_{B,i} = \widehat{\lambda}_i(\alpha, \beta) = \frac{\alpha + n_i}{\beta + n_i y_{i\cdot}} = \omega_i\frac{\alpha}{\beta} + (1-\omega_i)\frac{1}{\bar{y}_{i\cdot}}$$

with $\omega_i = \beta/(\beta + n_i\bar{y}_{i\cdot})$.
(iv) Integrate likelihood $L_{n_i}(\theta_i)$ times prior over θ_i space and find

$$f(\boldsymbol{y_i}) = \frac{\beta^\alpha}{(\beta + n_i\bar{y}_{i\cdot})^{\alpha+n_i}}\frac{\Gamma(\alpha+n_i)}{\Gamma(\alpha)}.$$

(Continued on page 3.)

b

$$\ell_k(\alpha,\beta) = \sum_{i=1}^{k} log f(\boldsymbol{y_i}) = k\alpha\log\beta - \sum_{i=1}^{k}(\alpha+n_i)\log(\beta+n_i\bar{y}_{i\cdot}) + \sum_{i=1}^{k}\log\Gamma(\alpha+n_i) - k\log\Gamma(\alpha).$$

A good answer should briefly state that this log-likelihood function can easily be programmed in R and then optimised wrt α and β (in order to obtain $\hat{\alpha}$ and $\hat{\beta}$).

c

$\widehat{\lambda}_{EB,i} = \widehat{\lambda}_i(\hat{\alpha},\hat{\beta})$ gives us the EB estimators (some might comment that the expression indicates that all the estimates are shrunk towards the estimated prior mean). Refer to the figure and comment that we see the typical "borrowing strength" effect of EB: the EB estimators are shrunk towards some common "mean". The student should indicate that experiments with small sample sizes seem to be shrunk more (more precisely maybe that experiments with a small $\sum_{j=1}^{n_i} y_{i,j}$ are shrunk more). Lastly, one can refer to "course knowledge" that EB estimators typically work well (in terms of risk functions!) in parts of the $\boldsymbol{\lambda}$ space (but not usually uniformly better than ML). Typically, EB will be good if the true λ_i are not too far from each other.

d

(i) Provide an expression for the full joint (unormalised) posterior:

$$\begin{aligned} p(\boldsymbol{\lambda},\alpha,\beta\,|\,\boldsymbol{y}) &\propto p(\alpha,\beta)p(\boldsymbol{\lambda}\,|\,\alpha,\beta)f(\boldsymbol{y}\,|\,\boldsymbol{\lambda}) \\ &\propto p(\alpha,\beta)\prod_{i=1}^{7}\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda_i^{\alpha-1}e^{-\beta\lambda_i}\prod_{i=1}^{7}\lambda_i^{n_i}e^{-\lambda_i n_i\bar{y}_{i\cdot}}. \end{aligned}$$

(ii) Given (α,β) the λ_i can be sampled directly (from gamma distributions). Provide an expression for $p(\alpha,\beta\,|\,\boldsymbol{y})$:

$$p(\alpha,\beta\,|\,\boldsymbol{y}) = \frac{p(\boldsymbol{\lambda},\alpha,\beta\,|\,\boldsymbol{y})}{p(\boldsymbol{\lambda},\,|\,\boldsymbol{y},\alpha,\beta)} \propto p(\alpha,\beta)\left[\frac{\beta^\alpha}{\Gamma(\alpha)}\right]^7\prod_{i=1}^{7}\frac{\Gamma(\alpha+n_i)}{(\beta+n_i\bar{y}_{i\cdot})^{\alpha+n_i}}$$

(iii) An answer might include

- a brief general overview of the general MCMC algorithm. Goal: make a markov chain whose stationary distribution is equal to the posterior distribution we are aiming at. After initialising the chain at (α^0,β^0), the algorithm basically contains three steps: (1) proposal of a value (α^p,β^p), (2) computation of acceptance probability, (3) decision to stay a previous value or accept the new value.
- the acceptance probability at the lth iteration in this specific case will be $min(1, p(\alpha^p,\beta^p\,|\,\boldsymbol{y})/p(\alpha^{l-1},\beta^{l-1}\,|\,\boldsymbol{y}))$ using the expression of $p(\alpha,\beta\,|\,\boldsymbol{y})$ from the previous question.
- I would use a simple binormal distribution for the proposal, with expectation equal to the previous values $(\alpha^{l-1},\beta^{l-1})$, zero correlation (for simplicity) and variances scaled to some reasonable values (not too large, not too small - one should monitor the overall acceptance rate in order to tune this).

(Continued on page 4.)

Problem 3 Submarine example

a

$$L_2(\theta) = \left(\frac{1}{10}\right)^2 I[y_{\max} - 5 \le \theta \le y_{\min} + 5]$$

b

Show that (here one needs to use Bayes formula, and compute the normalising constant)

$$p(\theta \mid y_1, y_2) = \begin{cases} \frac{1}{10-D}, & \text{for } \theta \in [y_{\max} - 5, y_{\min} + 5] \\ 0, & \text{otherwise,} \end{cases}$$

with $D = |y_1 - y_2|$, $y_{\max} = \max(y_1, y_2)$ and $y_{\min} = \min(y_1, y_2)$.

c

Posterior cdf:

$$P(\theta) = \begin{cases} 0, & \text{for } \theta < y_{\max} - 5 \\ \frac{\theta + 5 - y_{\max}}{10 - D}, & \text{for } \theta \in [y_{\max} - 5, y_{\min} + 5] \\ 1, & \text{for } \theta > y_{\min} + 5 \end{cases}$$

Posterior quantile function (inverse cdf):

$$P^{-1}(u) = y_{\max} - 5 + u(10 - D)$$

A 95% posterior interval for the following data: $y_1 = -4.1$ and $y_2 = 2.8$. $[-2.1225; 0.8225]$

d

Here one needs to consider $g(a)$, the posterior loss of each action. I find:

$$g(A) = 10P(0) \qquad g(B) = 20(1 - P(0)).$$

Both quantities depend on $P(0)$, the posterior cdf evaluated at 0. For the data provided we have $P(0) = 0.71$, giving

$$g(A) = 7.1 \qquad g(B) = 5.8.$$

Thus B is the Bayes action in this case.

THE END