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 5 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.

Problem 1 Poisson-Gamma

We will start with some questions concerning Poisson distributed data. In this exercise (and also the next one), you might find the following formula useful,

$$\int_0^\infty x^{a-1}e^{-bx}dx = \frac{\Gamma(a)}{b^a},$$

for positive numbers a and b. Also remember the following identity: $\Gamma(a+1) = a\Gamma(a)$.

Consider n independent observations from a Poisson distribution with parameter θ (and with density $f(y_i \mid \theta) = \theta^{y_i}e^{-\theta}/y_i!$). In this exercise we will use the following Gamma prior for θ,

$$p(\theta) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}.$$

The prior mean and variance can be expressed as $\mathrm{E}(\theta) = a/b$ and $\mathrm{Var}(\theta) = a/b^2$.

a

(i) Find the posterior distribution for θ for $n = 1$, i.e. conditional on one observation $\boldsymbol{y} = y_1$.
(ii) Then find the posterior distribution for θ for the general case of $n \geq 1$, i.e. conditional on data $\boldsymbol{y} = (y_1, \ldots, y_n)$.

b

(i) Provide expressions for the marginal mean and variance of y_i.
(ii) Find the marginal covariance and correlation between y_i and y_j, where $i \neq j$.

(Continued on page 2.)

c

Find an explicit formula for the posterior predictive distribution of a new observation y^*. Describe briefly how you would sample from this posterior predictive distribution in practice (using for example **R**).

d

In this subquestion, we forget the Poisson-Gamma set-up for a moment and consider a general model $f(y \mid \theta)$ and prior $p(\theta)$ leading to a posterior $p(\theta \mid y)$. Consider the loss function $L(\theta, \widehat{\theta}) = (\theta - \widehat{\theta})^2/\theta$ for a positive parameter θ. Show that the Bayes estimator (or Bayes action if you want) related to this loss function is

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

e

Now we are back to our n Poisson observations from Pois(θ), with a Gamma(a, b) prior as specified in the beginning of the exercise. Use the formula from **d** to find the Bayes estimator $\widehat{\theta}_B$ in this case.

f

Still considering the loss function $L(\theta, \widehat{\theta}) = (\theta - \widehat{\theta})^2/\theta$, find the risk function of the Bayes estimator $\widehat{\theta}_B$ from **e**. Also derive the risk function of the frequentist estimator $\widehat{\theta}_F = \bar{y}$. In what part of the θ space is the Bayes estimator better than the frequentist estimator?

Problem 2 Seven exponential experiments

Assume we have k different (and independent) experiments with potentially different sample sizes n_i in each experiment. Let $Y_{i,j}$ be the jth observation from the ith experiment, where $i = 1, \dots, k$ and $j = 1, \dots, n_i$. The observations are assumed to be independent and exponentially distributed,

$$Y_{i,j} \mid \lambda_i \sim \mathrm{Expo}(\lambda_i),$$

with experiment-specific parameters λ_i (the exponential density is, as usual, $f(y_{i,j} \mid \lambda_i) = \lambda_i e^{-\lambda_i y_{i,j}}$). Based on our knowledge of these experiments, we have chosen a common Gamma prior for the λ_i,

$$\lambda_i \sim \mathrm{Gamma}(\alpha, \beta).$$

In the table on the next page we provide the experiment-specific sample sizes and means ($\bar{y}_{i\cdot} = \sum_{j=1}^{n_i} y_{i,j}/n_i$) for $k = 7$ experiments. (This table is provided for concreteness, but you are not required to use the numbers in any calculations.)

(Continued on page 3.)

Experiment	1	2	3	4	5	6	7
n_i	4	9	12	3	5	7	8
$\bar{y}_{i\cdot}$	0.37	1.89	0.70	0.37	0.44	0.65	0.40

a

First we will consider each experiment separately.
(i) Show that the maximum likelihood estimator for λ_i (based on the data from study i only) is $\widehat{\lambda}_{ML,i} = 1/\bar{y}_{i\cdot}$.
(ii) Find the posterior distribution for λ_i based on the data from study i (using the Gamma prior above).
(iii) Show that the Bayes estimate under quadratic loss may be expressed as the following function of α and β:

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

with $\omega_i = \beta/(\beta + n_i \bar{y}_{i\cdot})$.
(iv) Find the marginal density of $\boldsymbol{y_i}$, the vector of observation from experiment i.

b

Consider now the full dataset from the k experiments. Use the result from question **a** (iv) and find an expression for the marginal log-likelihood function for the full dataset. Then explain how you would use this marginal log-likelihood function in order to estimate α and β from the data (using for example **R** or similar software).

c

Estimating prior parameters from the data corresponds to a so-called *empirical Bayes (EB)* strategy. Explain how you would use this strategy for estimating the ensemble of λ_i parameters. The results of the EB strategy are presented in Figure 1. Comment on the difference between EB and maximum likelihood estimates in this case. What seems to be the effect of the experiment-specific sample sizes (n_i)? Also comment briefly on how well you would expect this strategy to work, compared e.g. to the standard maximum likelihood method.

d

The empirical Bayes strategy may be considered an approximation to the fully Bayesian hierarchical model where we include a hyperprior for α and β. Let this hyperprior be $p(\alpha, \beta)$.
(i) Provide an expression for the full joint (unnormalised) posterior in this case, $p(\boldsymbol{\lambda}, \alpha, \beta \,|\, \boldsymbol{y})$; $\boldsymbol{\lambda}$ denotes the vector of seven λ_i and $\boldsymbol{y}$ denotes the full dataset from all the experiments.
(ii) This question and the next will concern how to sample from this joint posterior using computational methods. First, explain why, in this case, the structure of the model allows you to avoid making an MCMC for the full

(Continued on page 4.)

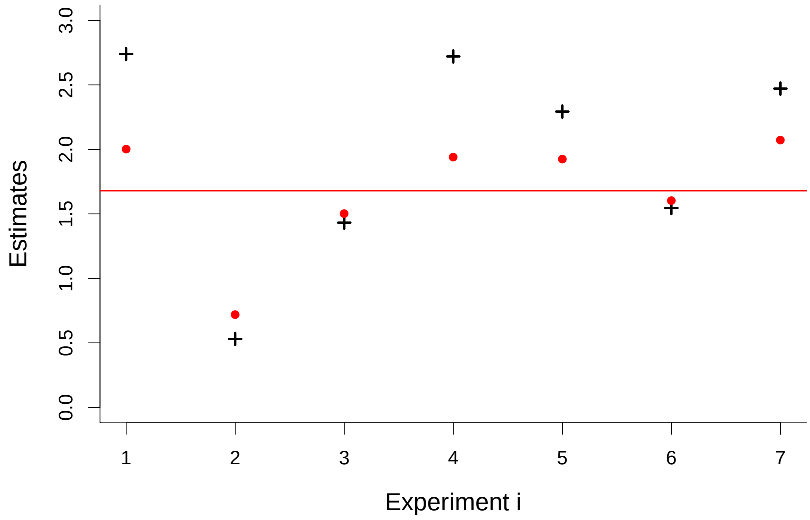

Figure 1: Estimated λ_i from the seven experiments. The black crosses are the estimates based on maximum likelihood (for each study separately). The red dots are estimates using an empirical Bayes strategy. The red line gives the estimated prior expectation.

vector of parameters $(\boldsymbol{\lambda}, \alpha, \beta)$. Then, provide an expression for $p(\alpha, \beta \,|\, \boldsymbol{y})$ (up to a constant of proportionality).
(iii) Briefly sketch how you would construct a Metropolis MCMC in order to sample from $p(\alpha, \beta \,|\, \boldsymbol{y})$. Include a brief description of the general algorithm, how you would compute acceptance probabilities and which proposal distribution you might use.

Problem 3 The lost submarine

In this exercise we consider two observations from a continuous uniform distribution of the following type,

$$Y_i \,|\, \theta \sim \text{Unif}(\theta - 5, \theta + 5)$$

with $i = 1, 2$. Thus, the maximal spread (distance) between the data is known, but the center θ is not (and lives on $(-\infty, +\infty)$). In parts of the literature, this set-up is presented as a story about an attempt to localise a submerged submarine, but this aspect is not crucial for the rest of this exercise. The density function is

$$f(y_i \,|\, \theta) = \begin{cases} \frac{1}{10}, & \text{for } y_i \in [\theta - 5, \theta + 5] \\ 0, & \text{otherwise.} \end{cases}$$

a

Give an expression for the likelihood function for θ based on the two observations.

(Continued on page 5.)

b

In the following, we will use the following improper prior for θ, $p(\theta) \propto 1$. Show that the posterior distribution for θ based on the two observations is equal to

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

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

c

Find an expression for the posterior cumulative distribution function and the posterior quantile function. Use this to construct a 95% posterior interval for the following data: $y_1 = -4.1$ and $y_2 = 2.8$.

d

A certain decision needs to be made, either A or B. The loss function is

$$L(\theta, A) = \begin{cases} 0 & \text{if } \theta \geq 0, \\ 10, & \text{if } \theta < 0, \end{cases}$$

$$L(\theta, B) = \begin{cases} 20 & \text{if } \theta \geq 0, \\ 0, & \text{if } \theta < 0. \end{cases}$$

Find the right decision (using the same two datapoints as in **d**).

THE END