UNIVERSITY OF OSLO

Faculty of mathematics and natural sciences

Exam in: STK4021 — Applied Bayesian statistics - Home exam

Day of examination: June 5 - June 12 2020

Examination hours: 14.30 – 18.30.

This problem set consists of 5 pages.

Appendices: None

Permitted aids: Anything available

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

Problem 1

(a) We have that

$$\log p(y|\boldsymbol{\theta}) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y-\mu)^2$$

giving

$$\frac{\partial}{\partial\mu} = \frac{1}{\sigma^2}(y-\mu) \qquad \frac{\partial}{\partial\sigma^2} = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(y-\mu)^2$$

$$\frac{\partial^2}{\partial\mu^2} = -\frac{1}{\sigma^2} \qquad \frac{\partial^2}{\partial\mu\sigma^2} = -\frac{1}{\sigma^4}(y-\mu)$$

$$\frac{\partial^2}{\partial(\sigma^2)^2} = \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(y-\mu)^2$$

giving further

$$\boldsymbol{J} = E\begin{pmatrix} \frac{1}{\sigma^2} & \frac{1}{\sigma^4}(y-\mu) \\ \frac{1}{\sigma^4}(y-\mu) & -\frac{1}{2\sigma^4} + \frac{1}{\sigma^6}(y-\mu)^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

$$|\boldsymbol{J}| = \frac{1}{2\sigma^6}$$

and Jeffreys' prior model is $\pi^J(\boldsymbol{\theta}) \propto \sigma^{-3}$.

(b) We get the same calculations as in sec 3.2 (eq 3.2) except that we need

(Continued on page 2.)

to modify for the different prior. That is

$$
\begin{aligned}
p(\mu, \sigma^2|\boldsymbol{y}) \propto& \sigma^{-n-3} \exp\left(-\tfrac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\right) \\
p(\mu|\sigma^2, \boldsymbol{y}) =& N(\bar{y}, \sigma^2/n) \\
p(\sigma^2|y) \propto& \int_\sigma \sigma^{-(n+3)} \exp\left(-\tfrac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\right) d\mu \\
\propto& \sigma^{-(n+3)} \exp\left(-\tfrac{1}{2\sigma^2}(n-1)s^2\right) \sqrt{2\pi\sigma^2/n} \\
\propto& (\sigma^2)^{-(n+2)/2} \exp\left(-\tfrac{1}{2\sigma^2}(n-1)s^2\right) \\
\propto& \text{Inv-}\chi^2(n, s^2)
\end{aligned}
$$

(c) We have that under the given loss function that

$$
\begin{aligned}
\hat{\mu} =& \bar{y} \\
\hat{\sigma}^2 =& \frac{n}{n-2} s^2
\end{aligned}
$$

We see that we obtain the same result as in the book for $\hat{\mu}$ and an unbiased estimate for μ.

For $\hat{\sigma}^2$ we would have obtained $\frac{n-1}{n-3}s^2$ for the alternative prior. The expectations would then be $\frac{n}{n-2}\sigma^2$ and $\frac{n-1}{n-3}\sigma^3$, respectively, so a somewhat less biased estimator for the prior considered in this exercise.

(d) We have that

$$
\begin{aligned}
E[L(\theta, \hat{\theta})|\boldsymbol{y}] =& \int_\theta I(|\theta - \hat{\theta}| > \varepsilon) p(\theta|\boldsymbol{y}) d\theta \\
=& 1 - \Pr(|\theta - \hat{\theta}| \leq \varepsilon | \boldsymbol{y})
\end{aligned}
$$

This will, for small ε be minimized when $\hat{\theta}$ is close to the mode of $p(\theta|\boldsymbol{y})$.

This loss function do not consider the type of error made, only if an error is made, which is not a very reasonable setting for continous variables.

Problem 2

(a) We have

$$
f(y|\theta, \tau) = \theta\tau y^{\theta-1} \exp(-y^\theta \tau),
$$

(Continued on page 3.)

(b) We get

$$\begin{aligned}p(\tau|\boldsymbol{y}) \propto & p(\tau)\prod_{i=1}^{n} p(y_i|\tau,\sigma)\\ \propto & \tau^{a-1}e^{-b\tau}\prod_{i=1}^{n}\theta\tau y_i^{\theta-1}\exp(-y_i^{\theta}\tau)\\ \propto & \tau^{a+n-1}\exp(-(b+\sum_{i=1}^{n}y_i^{\theta})\tau)\\ \propto & \text{Gamma}(\tau, a+n, b+\sum_{i=1}^{n}y_i^{\theta})\end{aligned}$$

We then have

$$p(\sigma|\boldsymbol{y}) = \text{Gamma}(\sigma^{-\theta}, a+n, b+\sum_{i=1}^{n}y_i^{\theta})\theta\sigma^{-\theta-1}$$

by the transformation rule.

(c) We have

$$\begin{aligned}p(y) = & \int_{\tau} p(y|\tau)d\tau\\ = & \int_{\tau}\theta\tau y^{\theta-1}\exp(-y^{\theta}\tau)\frac{b^a}{\Gamma(a)}\tau^{a-1}\exp(-b\tau)d\tau\\ = & \theta\frac{b^a}{\Gamma(a)}y^{\theta-1}\int_{\tau}\tau^{a+1-1}\exp(-(b+y^{\theta})\tau)d\tau\\ = & \theta\frac{b^a}{\Gamma(a)}y^{\theta-1}\frac{\Gamma(a+1)}{(b+y^{\theta})^{a+1}}\\ = & \frac{\theta a b^a y^{\theta-1}}{(b+y^{\theta})^{a+1}}\end{aligned}$$

Problem 3

(a) We have

$$\begin{aligned}p(\theta|y_1,y_2) = & \frac{p(\theta)f(y_1|\theta)f(y_2|\theta)}{f(y_1,y_2)}\\ = & \frac{p(\theta)f(y_1|\theta)}{f(y_1)}\frac{f(y_1)p(y_2|\theta)}{f(y_1,y_2)}\\ = & p(\theta|y_1)\frac{f(y_2|\theta)}{f(y_2|y_1)}\end{aligned}$$

We see then that this corresponds to updating the information about θ when observing y_2 given a "prior" $p(\theta|y_1)$.

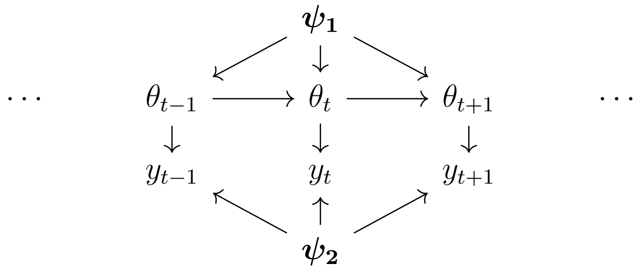
(Continued on page 4.)

(b) In general we have

$$\begin{aligned}
p(\theta|y_{1:t}) =& \frac{p(\theta)\prod_{s=1}^{t} f(y_t|\theta)}{f(y_{1:t})}\\
=& \frac{p(\theta)f(y_{1:t-1}|\theta)}{f(y_{1:t-1})}\frac{f(y_{1:t-1}|\theta)p(y_t|\theta)}{f(y_{1:t})}\\
=& p(\theta|y_{1:t-1})\frac{f(y_t|\theta)}{f(y_t|y_{1:t-1})}
\end{aligned}$$

where we see that we can sequentially update information as observations come in.

(c) See graph below:



(d) We now have

$$\begin{aligned}
p(\boldsymbol{\psi}, \theta_{1:t}|y_{1:t}) =& \frac{p(\boldsymbol{\psi})p(\theta_1|\boldsymbol{\psi}_1)p(y_1|\theta_1, \boldsymbol{\psi}_1)\prod_{s=1}^{t}[p(\theta_s|\theta_{s-1}, \boldsymbol{\psi}_1)f(y_s|\theta_s, \boldsymbol{\psi}_2)]}{f(y_{1:t})}\\
=& \frac{p(\boldsymbol{\psi})p(\theta_1|\boldsymbol{\psi}_1)p(y_1|\theta_1, \boldsymbol{\psi}_2)\prod_{s=1}^{t-1}[p(\theta_s|\theta_{s-1}, \boldsymbol{\psi}_1)f(y_s|\theta_s, \boldsymbol{\psi}_2)]}{f(y_{1:t-1})}\times\\
& f(y_{:t-1})\frac{p(\theta_t|\theta_{t-1}, \boldsymbol{\psi}_1)f(y_t|\theta_t, \boldsymbol{\psi}_2)}{f(y_{1:t})}\\
=& p(\boldsymbol{\psi}, \theta_{1:t-1}|y_{1:t-1})\frac{p(\theta_t|\theta_{t-1}, \boldsymbol{\psi}_1))p(y_t|\theta_t, \boldsymbol{\psi}_2}{f(y_t|y_{1:t-1})}
\end{aligned}$$

which can be used to update the posterior as new data arrive.

(e) There also seem to be a different regime around 01/04 which x_t does not capture.

When both θ_t and the categorical variable is included, there is an overparametrization which make the β_j's redundant. However, since they are given some estimated values, the definition of the θ's become somewhat different making also the estimates of σ, ρ different.

(f) We have

(Continued on page 5.)

BF	Model 1	Model 2	Model 3
Model 1	1.00e+00	0.00	0.00
Model 2	2.49e+39	1.00	0.41
Model 3	6.13e+39	2.46	1.00

Showing that the posterior probabilities becomes $0.00, 0.29$ and 0.71, respectively, so models 2 and 3 are comparable (and much better than model 1).

Similar results for waic.

(g) We actually now get a better model even though the linear prediction seems to neglect any variability in θ_t. This can be explained by that the Negative binomial distribution actually is a hierarchical model in it self with the Poisson distribution having a random rate. The variability in θ_t is then moved to the extra dispersion in the negative binomial distribution.