UNIVERSITY OF OSLO

Faculty of mathematics and natural sciences

Exam in: STK4021 — Applied Bayesian statistics - Home exam

Day of examination: November 27 -2020

Examination hours: 15.00 – 19.00.

This problem set consists of 4 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.

Notation:

- We will use $x_{1:t} = (x_1, ..., x_t)$.

Problem 1

(a) We have that

$$\begin{aligned}
\ell(\theta) \equiv \log L(\theta) =& \log \binom{x+r-1}{r-1} + r\log(\theta) + x\log(1-\theta) \\
\frac{\partial}{\partial\theta}\ell(\theta) =& \frac{r}{\theta} - \frac{x}{1-\theta} \\
\frac{\partial^2}{\partial\theta^2}\ell(\theta) =& -\frac{r}{\theta^2} - \frac{x}{(1-\theta)^2} < 0
\end{aligned}$$

and putting the derivative to zero, we obtain $\hat{\theta}_{ML} = \frac{r}{x+r}$ (which is a max point due to that the second derivative is negative).

(b) We have

$$\begin{aligned}
p(\theta|x) \propto& p(\theta)p(x|\theta) \\
\propto& \theta^{a-1}(1-\theta)^{b-1}\theta^r(1-\theta)^x \\
=& \theta^{a+r-1}(1-\theta)^{b+x-1} \\
\propto& \text{Beta}(r+a, x+b)
\end{aligned}$$

so a Beta distribution with parameters $a+r$ and $b+x$.

(c) Conjugacy means that the prior and the posterior belongs to the same class of distributions. That is the case here so the beta distribution is the conjugate family for this particular formulation of the negative binomial distribution.

(Continued on page 2.)

(d) We have that

$$\begin{aligned}E[\theta|x] =&\frac{a+r}{a+b+x+r}\\ =&\frac{x+r}{a+b+x}\frac{r}{x+r}+\frac{a+b}{a+b+x+r b}\frac{a}{a+b}\end{aligned}$$

showing the result with $\lambda=\frac{x+r}{a+b+x+r}$.

This shows that the posterior expectation is a weighted average of the prior expectation and the information obtained from data (here measured through the maximum likelihood estimate). We also see that the weight increase with x which corresponds to more information (seen through the observed information).

(e) We have that the expected Fisher information, $J(\theta)$ is given by

$$\begin{aligned}J(\theta) =&E[\frac{r}{\theta^2}+\frac{x-r}{(1-\theta)^2}]=\frac{r}{\theta^2}+\frac{\frac{r}{\theta}-r}{(1-\theta)^2}\\ =&r[\frac{1}{\theta^2}+\frac{1-\theta}{\theta(1-\theta)^2}]=r[\frac{1}{\theta^2}+\frac{1}{\theta(1-\theta)}]\\ =&\frac{r}{\theta}[\frac{1}{\theta}+\frac{1}{1-\theta}]=\frac{r}{\theta^2(1-\theta)}.\end{aligned}$$

Jeffreys' prior then becomes

$$p(\theta)\propto\frac{1}{\theta\sqrt{1-\theta}}$$

The posterior in this case becomes

$$\begin{aligned}p(\theta|x)\propto&\frac{1}{\theta\sqrt{1-\theta}}\theta^r(1-\theta)^{x-r}\\ \propto&\theta^{r-1}(1-\theta)^{x-r-0.5}\\ \propto&\text{Beta}(r,x-r+0.5)\end{aligned}$$

(f) For the Beta prior, we have

$$\begin{aligned}p(\theta|x_1,...,x_n)\propto&p(\theta)\prod_{i=1}^n p(x_i|\theta)\\ \propto&\theta^{a-1}(1-\theta)^{b-1}\theta^{nr}(1-\theta)^{\sum_{i=1}^n x_i}\\ =&\theta^{a+nr-1}(1-\theta)^{b+\sum_{i=1}^n x_i-1}\\ \propto&\text{Beta}(a+nr,b+\sum_{i=1}^n x_i)\end{aligned}$$

(Continued on page 3.)

For the Jeffreys' prior, we obtain

$$\begin{aligned} p(\theta|x_1, ..., x_n) \propto & p(\theta)\prod_{i=1}^{n} p(x_i|\theta) \\ \propto & \theta^{-1}(1-\theta)^{-0.5}\theta^{nr}(1-\theta)^{\sum_{i=1}^{n} x_i} \\ = & \theta^{nr-1}(1-\theta)^{\sum_{i=1}^{n} x_i - 0.5} \\ \propto & \text{Beta}(nr, \sum_{i=1}^{n} x_i + 0.5) \end{aligned}$$

They both depend on $\sum_{i=1}^{n} x_i$ due to that this is the sufficient statistic for θ in this case.

(g) We have that the maximum likelihood estimate becomes

$$\hat{\theta}_{ML} = \frac{nr}{\sum_{i=1}^{n} x_i + nr} = \frac{20}{27} = 0.74$$

while the prior expectations for the Beta priors are all 0.5. The higher the values of a, b, the more concentrated are the priors around 0.5 and will therefore push the posteriors towards 0.5. For the Jeffreys' prior, the mean is more close to the maximum likelihood estimate.

Based on this the purple curve correspons to Jeffreys' prior while the black, red, green and blue correspond to $a = 2, 3, 4, 5$.

Problem 2

(a) Bootstrapping is considering the frequentist properties when repeating experiments. In this case then some samples will not contain the extreme x observation in which case the estimates will be much less reliable.

The Bayesian approach, however, condition on the (lucky) setting where we have got this extreme observation which do give much information about the parameters. This then gives much less variability in the prediction intervals.

Note however that the Bayesian approach depend much more on the model assumption which might be questionable in the region where there are no x values.

(b) Assume $\Pr(M_1) = \Pr(M_2) = 0.5$. Assume further that these are the

(Continued on page 4.)

only two models under consideration. Then

$$
\begin{aligned}
Pr(M_1|y) =& \frac{\Pr(M_1)p(y|M_1)}{\Pr(M_1)p(y|M_1) + \Pr(M_2)p(y|M_2)} \\
=& \frac{p(y|M_1)}{p(y|M_1) + p(y|M_2)} \\
=& \frac{\exp(-23.357)}{\exp(-23.357) + \exp(-24.026)} = 0.661 \\
BF(M_1, M_2) =& \frac{p(y|M_1)}{p(y|M_2)} = e^{-23.357+24.026} = 1.952
\end{aligned}
$$

The Bayes factor and model probabilities give a slight preference for model 1. On the other hand, elpd gives a slight preference to model 2. The textbook prefers elpd due to the less sensitivity to priors. In this case both measures gives high uncertainty about which model to prefer, so model averaging would perhaps be preferable here.

(c) Cross-validation might be very unstable here due to the one extreme observation. Bridge sampling can also be somewhat problematic due to that the importance weights might be extreme when the last observation is left out. I would therefore be a bit sceptic to the results. However, visual inspection of the results do give similar support as the results shown.

(d) Predictive p-values are similar to classical p-values except that they take the uncertainties in the parameters into account. If the points are mainly on one side of the line, this indicate that there is some discrepancy in the model.

In this case, there is no clear evidence of discrepance, at least not with respect to what is checked here (which mainly is the most extreme residual).