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

The negative binomial distribution models the number of failure x in a sequence of independent and identically distributed Bernoulli trials before a specified (non-random) number of failures (denoted r) occurs. Let X be negative binomial, $X|\theta \sim \text{Neg-Binom}(r, \theta)$, so that

$$p(x|\theta) = \binom{x+r-1}{r-1} \theta^r (1-\theta)^x$$

with $0 < \theta < 1$, and $r, x \in \{0, 1, 2, ...\}$. (Note that the negative binomial distribution is represented somewhat different from how it is done in the textbook). You can use in the following that

$$E[x|\theta] = r\frac{(1-\theta)}{\theta}.$$

As a prior distribution, assume $\theta \sim \texttt{Beta}(a, b)$,

$$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1}, \quad a, b > 0.$$

(a) Show that the maximum likelihood estimate for θ is $\hat{\theta}_{ML} = \frac{r}{x+r}$

(b) Derive the posterior density $p(\theta|x)$. Which distribution is this and what are its parameters?

(Continued on page 2.)

(c) Define conjugacy and explain why, or why not, the beta prior is conjugate with respect to the negative binomial likelihood.

(d) Show that

$$E[\theta|x] = \lambda\hat{\theta}_{ML} + (1-\lambda)E[\theta]$$

for an approriate choice of λ.

Give an interpretation of this result.

(e) Derive Jeffreys' prior and the resulting posterior distribution.

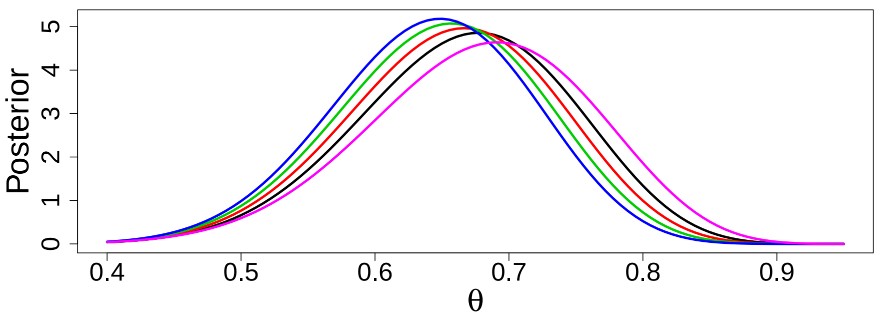
(f) Generalize now to the setting where $X_1, ..., X_n$ are conditionally (on θ) independent and all follow the Neg-Binom(r, θ) distribution. Derive the posterior distributions both for the Beta prior and Jeffreys' prior.

Show that the results in both cases correspond to the same results as if we would have one observation $x = \sum_{i=1}^n x_i$ from the Neg-Binom(nr, θ).

Argue why the influence on the posteriors are based on the same function of $(x_1, ..., x_n)$ in both cases.

(g) With $r = 5$, four values $(x_1, x_2, x_3, x_4) = (2, 3, 4, 0)$ were observed.

The plot below shows the posterior distribution corresponding to Beta priors with $a = b = j$ for $j = 2, 3, 4, 5$ and by the Jeffreys' prior. Decide on which curves that belong to the different priors and explain why.



Problem 2

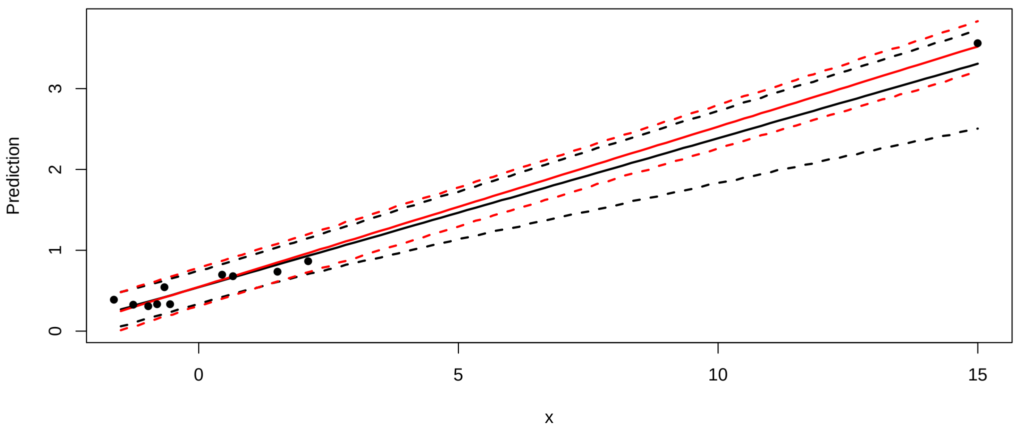
We will in this exercise consider a very simple regression model:

$$M_1 : y_i = \beta_0 + \beta_1 x_i + \sigma\varepsilon_i, \quad \varepsilon_i \sim N(0, 1)$$

Data were simulated from this model using $\beta_0 = 0.5, \beta_1 = 0.2$ and $\sigma = 0.1$. Our aim will be to give predictions and prediction intervals for new values of x.

(Continued on page 3.)

The plot below shows predictions (solid lines) and prediction intervals (dashed lines) based on bootstrapping (black lines) and Bayesian inference using non-informative priors (red lines). The points are the observed $n = 11$ datapoints. The boostrapping method used here is the non-parametric one where new bootstrap samples are obtained by sampling (with replacement) n points $\{(x_i^*, y_i^*), i = 1, ..., n\}$ from $\{(x_1, y_1), ..., (x_n, y_n)\}$.



(a) Explain the difference between the results obtained by Bootstraping compared to those obtained by the Bayesian approach. Relate this to general properties of Bayesian approaches.

We will now compare the linear model with one including a quadratic term in x:

$$M_2 : y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \sigma\varepsilon_i, \quad \varepsilon_i \sim N(0, 1)$$

The following table can be used for answering the questions below:

Model	Marginal log-likelihood	elpd
M_1	-23.357	-11.945
M_2	-24.026	-11.873

The marginal log-likelihoods are calculated by the leave-one-out (LOO) method while the expected log predictive densities (elpd) are calculated by bridge sampling.

(b) Based on the results in the table above, calculate the posterior probabilities for model M_1 and M_2. State the assumptions you make when calculating these probabilities.

Looking at the whole table, in which way would you recommend to do prediction?

(c) Explain the main ideas behind leave-one-out and bridge sampling.

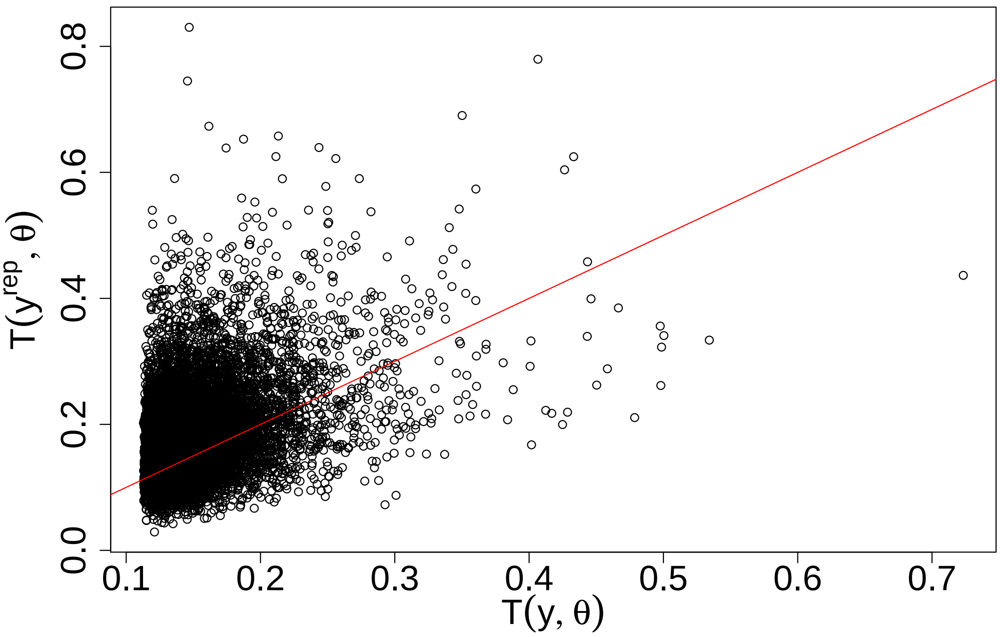
Would you trust the results obtained by these methods for this specific application?

(Continued on page 4.)

We will now turn to model checking applying the following discrepancy measure:

$$T(y, \beta) = \max_i |y_i - \beta_0 - \beta_1 x_i|.$$

The plot below is based on this measure within model M_1. The fraction of points that are above the line $y = x$ is 0.582.



(d) Explain the content of the plot above and how such plots can be used for model checking.

Based on the plot below, discuss whether these results indicate some problems with the model.