UNIVERSITY OF OSLO

Faculty of mathematics and natural sciences

Exam in:	STK9021 — Applied Bayesian statistics - Home exam
Day of examination:	June 5 - June 12 2020
Examination hours:	14.30 – 18.30.

This problem set consists of 6 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

Consider a normal model

$$y \sim N(\mu, \sigma^2).$$

Define $\boldsymbol{\theta} = (\mu, \sigma^2)$. Recall that the expected Fisher information matrix is given by

$$\boldsymbol{J}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left[-\frac{\partial^2}{\partial \boldsymbol{\theta}\boldsymbol{\theta}^T}\log p(y|\boldsymbol{\theta})\right]$$

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

(b) Assume you now have observed $y_1, y_2, ..., y_n$. Derive the posterior distribution for $\boldsymbol{\theta}$.

(c) Consider a loss function $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$ where θ is either μ or σ^2. Find the Bayes estimator for both μ and σ^2.

Compare the results with the results obtained with a prior $p(\boldsymbol{\theta}) \propto (\sigma^2)^{-1}$ as considered in the textbook (both by looking at the estimators and on their expectations).

(d) Consider now instead the loss function

$$L(\theta, \hat{\theta}) = \begin{cases} 0 & \text{if } |\hat{\theta} - \theta| \leq \varepsilon \\ 1 & \text{if } |\hat{\theta} - \theta| > \varepsilon \end{cases}$$

(Continued on page 2.)

Show that the Bayes estimate in this case becomes the mode of the posterior distribution $p(\theta|\boldsymbol{y})$ when we let $\varepsilon \to 0$. Derive the Bayes estimates for μ and σ^2 in this case.

Discuss both the result you obtained here compared to (*c*) and also the rationality of this loss function.

Problem 2

The *Weibull* distribution is defined by

$$p(y|\theta,\sigma) = \frac{\theta}{\sigma^\theta} y^{\theta-1} \exp(-(y/\sigma)^\theta), \quad \theta, \sigma > 0$$

and is often used for lifetime data or extreme value data. Here θ is a *shape* parameter while σ is a *scale* parameter. For simplicity, we will assume θ is known in this exercise.

(a) Define $\tau = \sigma^{-\theta}$. Write the distribution for y expressed with (θ, τ) instead.

(b) Show that the Gamma distribution is a conjugate family for τ in this parametrisation of the Weibull distribution.

Derive the posterior distribution for τ based on samples $y_1, ..., y_n$ which are assumed independently distributed given (θ, σ). Also derive from this the posterior distribution for σ.

(c) Still with θ assumed known, derive the marginal distribution for y.

Problem 3

Consider a setting where you have a prior $p(\theta)$ on a parameter θ and then two datasets/observations y_1 and y_2, both with densities $f(y_t|\theta)$ for $t = 1, 2$ (and independent given θ).

(a) Show that the posterior distribution of θ based on y_1 can be considered as a prior distribution for θ when updating information with respect to y_2, that is

$$p(\theta|y_1, y_2) = p(\theta|y_1)\frac{f(y_2|\theta)}{f(y_2|y_1)}. \tag{1}$$

(b) Assume now we have a whole sequence of data $y_1, y_2, ...,$ each conditional independent and with densities $f(y_t|\theta)$ for $t = 1, 2, ...$ Show that you can get a similar updating scheme from $t - 1$ to t as (1) and discuss how this can be used to update information about θ sequentially.

(Continued on page 3.)

Consider now a more general setting where there is one θ_t for each observation y_t such that

$$p(\boldsymbol{\psi}, \theta_{1:T}, y_{1:T}) = p(\boldsymbol{\psi})p(\theta_1|\boldsymbol{\psi}_1)p(y_1|\theta_1, \boldsymbol{\psi}_2)\prod_{t=2}^{T} p(\theta_t|\theta_{t-1}, \psi_1)p(y_t|\theta_t, \boldsymbol{\psi}_2)$$

where $\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2)$ are some global hyperparameters influencing the $\{\theta_t\}$ process and the observation distributions, respectively.

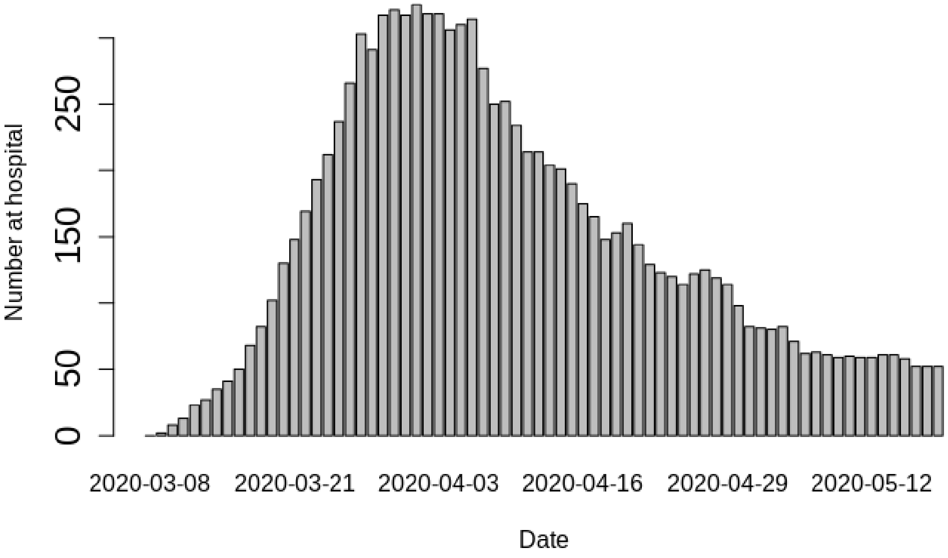
(c) Draw a graph similar to Figure 5.1 in the textbook that describes this model.

(d) Assume now that we want to make sequential inference about $\theta_{1:t}$ (and $\boldsymbol{\psi}$) based on data $y_{1:t}$. Assume you have available $p(\boldsymbol{\psi}, \theta_{1:t-1}|y_{1:t-1})$ available at time $t-1$, show how you can update this to $p(\theta_{1:t}|y_{1:t})$ at time t.

In the following we will consider a specific model where

$$\begin{aligned} p(\theta_t|\theta_{t-1}) =& N(\rho\theta_{t-1}, \sigma^2) \\ p(y_t|\theta_t) =& \text{Poisson}(\exp(\boldsymbol{\beta}^T\boldsymbol{x}_t + \theta_t)) \end{aligned}$$

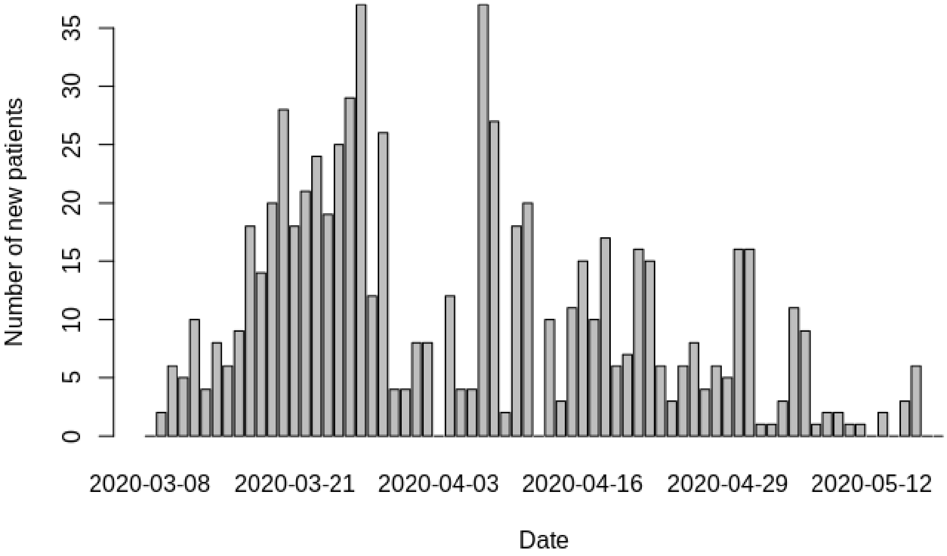
where $\boldsymbol{x}_t$ are some covariates observed at time t. Here $\boldsymbol{\psi}_1 = (\rho, \sigma^2)$ and $\boldsymbol{\psi}_2 = \boldsymbol{\beta}$.
We will in particular look at the number of Covid-19 patients present at hospitals for each day (data which are publically available from `Helsedirektoratet`). The plot below shows the total number of all patients within Norway.



Now, in order to analyse such data, it is better to look at the number of *new* cases each day. These data are not publically available, but a proxy can be

(Continued on page 4.)

made by looking at the differences in the total numbers between following days in addition to distributions of the length of hospitalization. The plot below shows a simulation of such data which will be used in the rest of the exercise:



Norway introduced several restrictions and reopenings. We will first define a categorical covariate

$$x_t = \begin{cases} 0 & \text{if } t \text{ is before March 15} \\ 1 & \text{if } t \text{ is between March 15 and May 1} \\ 2 & \text{if } t \text{ is after May 1} \end{cases}$$

We will let $\boldsymbol{x}_t$ denote the dummy variables related to x_t so that

$$\boldsymbol{\beta}^T\boldsymbol{x}_t = \begin{cases} \beta_0 & \text{if } x_t = 0 \\ \beta_1 & \text{if } x_t = 1 \\ \beta_2 & \text{if } x_t = 2 \end{cases}$$

We will consider three different models, all special cases of the general model described before.
Model 1:

$$p(y_t) = \text{Poisson}(\exp(\boldsymbol{\beta}^T\boldsymbol{x}_t))$$

Model 2:

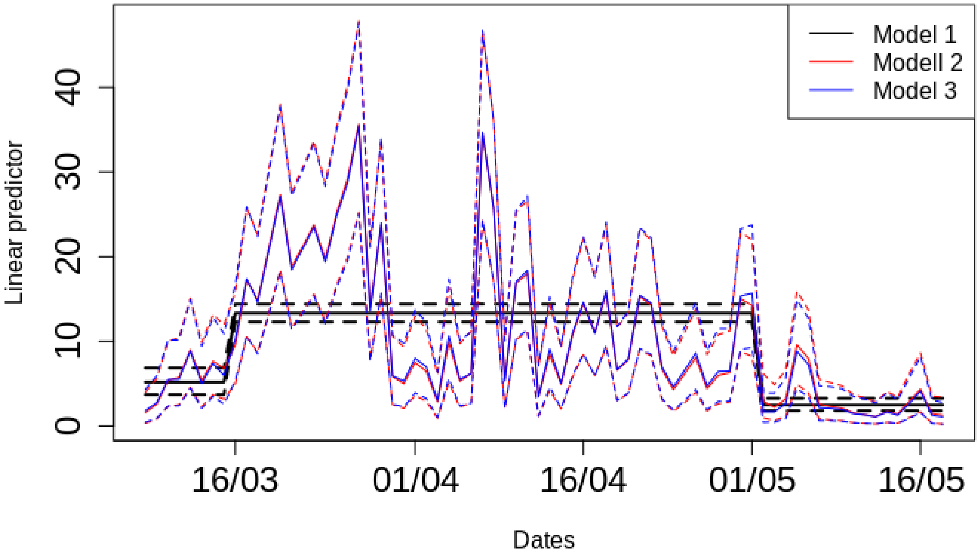
$$\begin{aligned} p(\theta_t|\theta_{t-1}) =& N(\beta_0 + \rho\theta_{t-1}, \sigma^2) \\ p(y_t|\theta_t) =& \text{Poisson}(\exp(\theta_t)) \end{aligned}$$

Model 3:

$$\begin{aligned} p(\theta_t|\theta_{t-1}) =& N(\rho\theta_{t-1}, \sigma^2) \\ p(y_t|\theta_t) =& \text{Poisson}(\exp(\boldsymbol{\beta}^T\boldsymbol{x}_t + \theta_t)) \end{aligned}$$

(Continued on page 5.)

The plot below shows the estimates of the linear predictor on the exponential scale ($\exp(\boldsymbol{\beta}^T\boldsymbol{x}_t+\theta_t)$) for the three models with the solid lines corresponding to the posterior means while the dashed lines correspond to the 95% credibility intervals (the lines for model 2 and 3 are overlapping).



The table below shows posterior means of the parameters involved, the marginal likelihoods (mlik) and waic measures.

	β_0	β_1	β_2	σ	ρ	mlik	p_{waic}	waic
Model 1	1.63	2.59	0.90			-334.72	12.26	659.51
Model 2	1.69			1.06	0.70	-245.93	37.70	393.29
Model 3	1.42	2.36	0.47	0.75	0.42	-244.90	35.96	387.92

(e) Discuss the plot of the linear predictors. In particular, answer the following questions:

- Is the covariate x_t reasonably defined here?
- Why do you think the linear predictors for models 2 and 3 are so similar in this case?
- Given that the linear predictors for models 2 and 3 are so similar, why do you think the estimates of σ and ρ are so different?

(f) Based on the marginal likelihoods, calculate Bayes factors between the different models.

Which model would you prefer?

What would your choice of model be if you used the waic criterion instead?

Discuss possible discrepancies here.

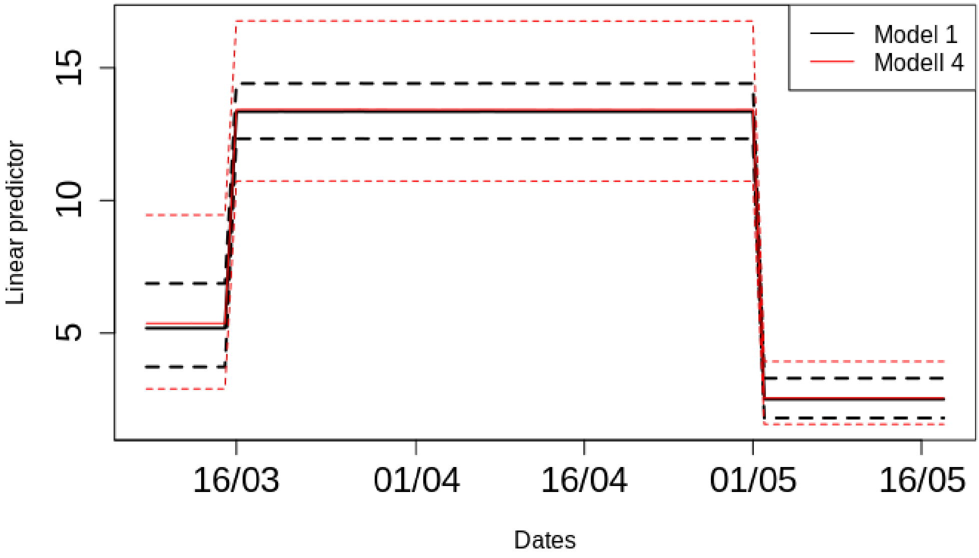
(Continued on page 6.)

Consider now yet another model,
Model 4:

$$\begin{aligned} p(\theta_t|\theta_{t-1}) =& N(\rho\theta_{t-1}, \sigma^2) \\ p(y_t|\theta_t) =& \text{Neg.binomial}(\exp(\boldsymbol{\beta}^T\boldsymbol{x}_t + \theta_t), \alpha) \end{aligned}$$

where the negative binomial distribution is parametrized by the *mean* (first parameter) and the *size* parameter.
The plot below shows the posterior mean of the linear predictor $\boldsymbol{\beta}^T\boldsymbol{x}_t + \theta_t$ in this case, combined with the results from model 1. Discuss the results obtained in this case

	β_0	β_1	β_2	σ	ρ	α	mlik	p_{waic}	waic
Model 4	1.63	0.96	-0.74	0.01	-0.00	1.99	-239.67	4.16	456.23


Model 1
Modell 4
Linear predictor
16/03
01/04
16/04
01/05
16/05
Dates