UNIVERSITETET I OSLO

Matematisk Institutt

Exam in:	**STK 4020/9020:**
	Bayesian Statistics
With:	**Nils Lid Hjort**
Time for exam:	**4 December 2008 at 11:55 – 15 December s.y. at 14:00**

This is the exam set for STK 4020, autumn semester 2008. It is made available on the course website as of *Thursday 4 December 12:00*, and candidates must submit their written reports by *Monday 15 December 14:00* (or earlier), to the reception office at the Department of Mathematics, *in duplicate*. The supplementary oral examinations take place December 18 and 19 (practical details for these will be provided later).

Reports may be written in nynorsk, bokmål, riksmål, English or Latin, and should preferably be text-processed (TeX, LaTeX, `Word`), but may also be hand-processed. Give your name on the first page. Write concisely (in der Beschränkung zeigt sich erst der Meister; brevity is the soul of wit; краткость – сестра таланта). Relevant figures need to be included in the report. Copies of machine programmes used (in `R`, or `matlab`, or similar) are also to be included, perhaps as an Appendix to the report. The full exam set *is* (admittedly) laborious, and candidates are graciously allowed not to despair if they do not manage to answer all questions well.

Importantly, each student needs to submit *two special extra pages* with her or his report. *The first* (page A) is the 'erklæring' (self-declaration form), properly signed; it is available at the webpage as 'Exam Project, page A, declaration form'. *The second* (page B) is the student's one-page summary of the exam project report, which should also contain a brief self-assessment of its quality.

This exam set contains three exercises and comprises eight pages.

Exercise 1

Do you read me? In her PhD thesis *The Effects on an Elaborated Directed Reading Activity on the Metacomprehension Skills of Third Graders* from 1987, Maribeth Schmitt conducted an experiment with young children (of age around eight) to see whether certain reading activities in the classroom would improve aspects of their reading skills. 21 children went through these extra activities for a period of eight weeks whereas a control group of 23 children were exposed to the same basic curriculum, but without the extra activities. After the eight-week period, the pupils took a certain 'degree of reading power' test, with the following results:

- the experiment group:
  `24 43 58 71 43 49 61 44 67 49 53 56 59 52 62 54 57 33 46 43 57`
- the control group:
  `42 43 55 26 62 37 33 41 19 54 20 85 46 10 17 60 53 42 37 42 55 28 48`

(a) We shall view the data $y_{1,1},\ldots,y_{1,n_1}$ from the control group and $y_{2,1},\ldots,y_{2,n_2}$ from the experimental group as independent and normally distributed with the same standard deviation σ and with mean parameters respectively μ_1 and μ_2. Assume for this and the following point that $\sigma = 12.00$ is a known quantity (the more realistic case of σ also unknown is being worked with below), and that the prior distribution that takes μ_1 and μ_2 independent, with

$$\mu_1 \sim \mathrm{N}(\mu_{1,0},\sigma^2/v_1) \quad \text{and} \quad \mu_2 \sim \mathrm{N}(\mu_{2,0},\sigma^2/v_2),$$

is an adequate representation of prior knowledge, with prior means $\mu_{1,0}$ and $\mu_{2,0}$ set to 45.00 and relative precision numbers v_1 and v_2 set to 2.00. Show first that

$$\begin{pmatrix}\mu_1\\ \bar y_1\end{pmatrix} \sim \mathrm{N}_2(\begin{pmatrix}\mu_{1,0}\\ \mu_{1,0}\end{pmatrix}, \sigma^2\begin{pmatrix}1/v_1 & 1/v_1\\ 1/v_1 & 1/v+1/n_1\end{pmatrix}),$$

by working through the required means, variances and covariance, and then show that

$$\mu_1 \,|\, \mathrm{data} \sim \mathrm{N}\Big(\frac{v_1\mu_{0,1}+n_1\bar y_1}{v_1+n_1}, \frac{\sigma^2}{v_1+n_1}\Big),$$

where $\bar y_1 = (1/n_1)\sum_{i=1}^{n_1} y_{1,i}$ (and with an analogous results for $\mu_2\,|\,\mathrm{data}$). You may e.g. use Hjort's 'Course Notes and Exercises', Exercise 4.

(b) Find the posterior distribution of $\delta = \mu_2 - \mu_1$, and give in particular a posterior 90% credibility interval for this parameter.

(c) More realistically σ is not a known parameter, however, and an extended Bayesian analysis is called for. With $\lambda = 1/\sigma^2$ the inverse variance, agree to say that the three parameters follow a gamma-normal distribution, and write

$$(\lambda,\mu_1,\mu_2) \sim \mathrm{GN}(a,b,\mu_{1,0},\mu_{2,0},v_1,v_2),$$

to indicate that

(i) $\lambda \sim \mathrm{Gamma}(a,b)$,

(ii) given λ (and hence $\sigma = 1/\sqrt{\lambda}$), μ_1 and μ_2 are independent and respectively $\mathrm{N}(\mu_{1,0},\sigma^2/v_1)$ and $\mathrm{N}(\mu_{2,0},\sigma^2/v_2)$.

Write down a clear expression for the prior density $p(\lambda,\mu_1,\mu_2)$. Also show that the likelihood function for the data may be expressed as

$$L(\lambda,\mu_1,\mu_2) \propto \lambda^{(n_1+n_2)/2}\exp\big[-\tfrac12\lambda\{Q_1+Q_2+n_1(\mu_1-\bar y_1)^2+n_2(\mu_2-\bar y_2)^2\}\big],$$

in terms of the group means $\bar y_1$ and $\bar y_2$ along with

$$Q_1 = \sum_{i=1}^{n_1}(y_{1,i}-\bar y_1)^2 \quad \text{and} \quad Q_2 = \sum_{i=1}^{n_2}(y_{2,i}-\bar y_2)^2.$$

(d) Show that

$$(\lambda, \mu_1, \mu_2) \,|\, \text{data} \sim \text{GN}(a^*, b^*, \mu_1^*, \mu_2^*, v_1 + n_1, v_2 + n_2),$$

where

$$\begin{aligned} a^* &= a + \tfrac{1}{2}(n_1 + n_2), \\ b^* &= b + \tfrac{1}{2}\{Q_1 + Q_2 + d_1(\bar y_1 - \mu_{1,0})^2 + d_2(\bar y_2 - \mu_{2,0})^2\}, \\ \mu_1^* &= \frac{v_1\mu_{1,0} + n_1\bar y_1}{v_1 + n_1}, \\ \mu_2^* &= \frac{v_2\mu_{2,0} + n_2\bar y_2}{v_2 + n_2}, \end{aligned}$$

and where finally $d_1 = v_1 n_1/(v_1 + n_1)$ and $d_2 = v_2 n_2/(v_2 + n_2)$. Comment on the particular case where prior precision numbers v_1 and v_2 go to zero.

(e) We are now in a position to carry out full Bayesian analysis for the 'degree of reading power' data. To select a prior of the gamma-normal type above, choose first gamma distribution parameters (a, b) for $1/\sigma^2$ by demanding that the 0.10 and 0.90 prior quantiles of σ are 10.0 and 20.0, respectively; then use $v_1 = v_2 = 0$ above, in addition to prior mean numbers 45.00 and 45.00 for μ_1 and μ_2. Draw 10^4 triples $(\sigma_j, \mu_{1,j}, \mu_{2,j})$ from the posterior distribution, and display a histogram for the distribution of $\delta = \mu_2 - \mu_1$. Also, find the 0.05, 0.50, 0.95 posterior quantiles of δ.

(f) Give an exact representation of the posterior distribution of δ above, in terms of a certain t-distribution, i.e. one that may be used without recourse to simulations. You may exploit the fact that if $U \sim \text{Gamma}(a, b)$, then $2bU \sim \chi^2_{2a}$. Check the exact 0.05, 0.50, 0.95 quantiles with those found above using simulations.

(g) The school authorities needed to make a decision after the study described above, among

$$\begin{cases} \text{action } a = 1, \text{ to not allow further experiments,} \\ \text{action } a = 2, \text{ to ask for a follow-up experiment,} \\ \text{action } a = 3, \text{ to immediately implement new pedagogical guidelines.} \end{cases}$$

Assume the loss function associated with the situation may be adequately presented as

$$\text{loss}((\sigma, \mu_1, \mu_2), a) = \begin{cases} 0 & \text{if } \delta < 2.0 \text{ and } a = 1, \\ 0 & \text{if } 2.0 \le \delta \le 7.0 \text{ and } a = 2, \\ 0 & \text{if } \delta > 7.0 \text{ and } a = 3, \\ 1 & \text{otherwise.} \end{cases}$$

Which decision are you recommending for the school authorities?

– jeg har ikke lært noe siden jeg lærte å lese

Ingvild Burkey (2008): Den mest tenkelige av alle verdener

Exercise 2

FOR HOW LONG HAS A CERTAIN PHENOMENON LASTED? Such a question is harder to answer when one does not know when it started and when it ended, only that its existence could be verified on certain occasions. There are many situations where questions like the above are pertinent, and there are several statistical variations on the theme. The one we focus on here is as follows: observations $y_1, \ldots, y_n$ are available and seen as having arisen independently from a uniform distribution on $[a, b]$. The task of primary interest is to assess the size of $\gamma = b - a$. For the present illustration, these are the data points:

```
30.02 26.70 24.93 22.96 32.58
32.60 31.00 23.39 24.45 24.31
27.01 29.28 27.89 28.41 29.24
```

In particular, denoting the data $y_1, \ldots, y_n$,

$$u_n = \min_{i \le n} y_i = 22.96, \quad v_n = \max_{i \le n} y_i = 32.60, \quad \Delta_n = v_n - u_n = 9.64.$$

(a) Show that the likelihood function may be written as follows, expressed as a function of (a, γ) rather than of (a, b):

$$L_n(a, \gamma) = \Big(\frac{1}{\gamma}\Big)^n I\{a \le u_n \text{ and } v_n \le a + \gamma\} = \begin{cases} (1/\gamma)^n & \text{if } a \le u_n \text{ and } v_n \le a + \gamma, \\ 0 & \text{otherwise.} \end{cases}$$

Find the maximum likelihood estimates for a and γ.

(b) With a flat prior on a, independently of a prior $p(\gamma)$ for γ, show that the posterior distribution of γ is

$$p(\gamma \,|\, \text{data}) \propto p(\gamma)(\gamma - \Delta_n)(1/\gamma)^n \quad \text{for } \gamma \ge \Delta_n,$$

where $\Delta_n = v_n - u_n$ is the precise range of the n data points.

(c) Show that a Bayesian analysis, starting with a prior that takes a and γ flat and independent on respectively the full line and the positive half-line, leads to the posterior density

$$p(\gamma \,|\, \text{data}) = (1/k_n)(\gamma - \Delta_n)(1/\gamma)^n \quad \text{for } \gamma \ge \Delta_n,$$

with k_n the required integral. For the data given above, plot the posterior density in a diagram. Also, compute the posterior mode $\widehat{\gamma}_{\text{mode}}$. Is this a Bayes estimate, under a suitable loss function?

(d) Discuss the use of flat priors here.

(e) Use a suitable technique to generate say 10,000 random γ from the posterior distribution, and display a histogram of these in the same diagram as the exact posterior density. Compute the 0.10, 0.50, 0.90 quantiles of the posterior.

(f) Dr. Phil. A.T. List is a stamp collector of some fame, and among his specialties are philatelic matters related to famous scientists. Hundred years after the death of Niels Henik Abel (1802–1829), the Norwegian postal office issued a certain stamp and a 'first day cover' envelope commemorating him. As the attached facsimile indicates (see page 8), these carry 'R numbers' (as in 'rekommandert post'), and R numbers from five such Abel 1929 envelopes, from various philatelic sales lists and auctions over the past few years, and known to List, and further communicated to me, are

280, 304, 308, 310, 328.

Neither List nor other philatelists do know where this sub-sequence of Abelian R numbers started or when it ended, however. Take these five known R numbers to be an ordered random sample of five numbers from $\{a+1,\ldots,a+\gamma\}$ (where all draws of five numbers have the same probability), with a and γ unknown. With a reasonable prior for the parameters, find and display the posterior distribution of γ, and, in particular, give a point estimate for the number of Abel 1929 envelopes that were once produced.

How should one go about building one's expertise?
– By studying the masters and not their pupils.
Niels Henrik Abel (1802–1829)

Exercise 3

HOW LONG IS A LIFE? A unique set of life-lengths in Roman Egypt was collected by W. Spiegelberg in 1901 *(Ägyptische und griechische Eigennamen aus Mumienetiketten der römischen Kaiserzeit)* and analysed by Karl Pearson (1902) in the very first volume of *Biometrika*. The data set contains the age at death for 141 Egyptian mummies in the Roman period, 82 men and 59 women, dating from the last century B.C. The life-lengths vary from 1 to 96 years, and Pearson argued that these can be considered a random sample from one of the better-living classes in that society, at a time when a fairly stable and civil government was in existence (as we recall, the violent 'tax revolt' with ensuing long-lasting complications took place under Antoninus Pius later, in 139 AD). – To access these data, go to the site `www.econ.kuleuven.be/public/ndbaf45/modelselection/index.html` of the Claeskens and Hjort (2008) book and check 'Mortality in ancient Egypt'.

Pearson did not attempt to fit any parametric models for these data, but discussed differences between the Egyptian age distribution and that of England 2000 years later. The purpose of the present exercise is to analyse aspects of the data from a Bayesian perspective, using the parametric Weibull model.

(a) The two-parameter Weibull model for life-times has a cumulative distribution function of the form

$$F(y,a,b) = 1 - \exp\Big\{-\Big(\frac{y}{a}\Big)^b\Big\} \quad \text{for } y > 0,$$

where a and b are positive parameters (typically unknown).

(i) Find a formula for the median of the distribution.

(ii) Show that the probability of surviving age t, given that one has survived up to t_0, is $\exp[-\{(t/a)^b - (t_0/a)^b\}]$, for $t > t_0$.

(iiii) Show that the density can be expressed as

$$f(y, a, b) = \exp\Big\{-\Big(\frac{y}{a}\Big)^b\Big\} \frac{by^{b-1}}{a^b} \quad \text{for } y > 0.$$

(b) You're the Bayesian statistician now, and learn that you soon will have the chance to analyse a set of Egyptian life-lengths from the era of Gaius Cornelius Gallus, Gaius Aelius Gallus and Octavianus Augustus. In preparation for this, you are to construct a reasonable prior $p(a, b)$ for the two parameters of the Weibull model. You do not need to be overly ambitious here, but your prior should be 'reasonable' – you should demonstrate that it is by drawing say 1000 random life-times from the associated prior predictive distribution, and displaying these in a histogram.

(c) The so-called Jeffreys prior is meant to be a non-informative 'default prior' for use for parametric models in absence of clear prior information, and is defined as

$$p_0(a, b) \propto |J(a, b)|^{1/2},$$

in terms of the Fisher information matrix of the model. This latter matrix is defined as

$$J(a, b) = \mathrm{Var}_{a,b} \begin{pmatrix} U(Y) \\ V(Y) \end{pmatrix},$$

where Y is from the indicated $f(y, a, b)$ distribution, and $U(y) = \partial \log f(y, a, b)/\partial a$ and $V(y) = \partial \log f(y, a, b)/\partial b$. – For the present Weibull model, show that the Jeffreys prior is proportional to $1/a$.

(d) We take now all the $n = 141$ life-lengths to come from the same population (i.e. we are not yet distinguishing between men's and women's survival distributions). By programming the log-likelihood function

$$\ell_n(a, b) = \sum_{i=1}^n \log f(y_i, a, b)$$

and using a suitable numerical optimiser, find both the maximum likelihood estimates $(\widehat{a}, \widehat{b})$ and the associated second derivatives matrix

$$\widehat{J} = -\frac{\partial^2 \ell_n(\widehat{a}, \widehat{b})}{\partial\theta\partial\theta^{\mathrm{t}}} = -\begin{pmatrix} \partial^2 \ell_n(\widehat{a}, \widehat{b})/\partial a^2 & \partial^2 \ell_n(\widehat{a}, \widehat{b})/\partial a\, \partial b \\ \partial^2 \ell_n(\widehat{a}, \widehat{b})/\partial a\, \partial b & \partial^2 \ell_n(\widehat{a}, \widehat{b})/\partial b^2 \end{pmatrix}.$$

Here θ is the parameter vector $(a, b)^{\mathrm{t}}$. [The algorithm `nlm(minusloglik, starthere, hessian=T)` may e.g. be used, in `R`. I find $\widehat{a} = 33.563$ and $\widehat{b} = 1.404$.] Plot the estimated density curve along with a histogram of the data.

(e) Regardless of your efforts related to working with points (b) and (c), we now take simply a flat uniform prior $p(a, b) \propto 1$ on the region $[10, 50] \times [0.10, 3.00]$ in the positive quadrant. Find first a bi-normal approximation to the posterior $p(a, b \mid \text{data})$. Give approximate 95% credibility intervals for a and for b. Draw say 10^4 random pairs (a_j, b_j) from this approximate posterior distribution, and use these to give point estimates and approximate 95% credibility intervals for two quantities of interest:
   (i) the median of the life-length distribution;
   (ii) the probability κ that a person survives past 40 years, given that the person has survived at least up to 20 years.

(f) Then try to simulate random pairs from the *exact* posterior distribution $p(a, b \mid \text{data})$, using e.g. a Markov Chain Monte Carlo strategy. Use these simulations to give point estimates and 95% credibility intervals for a, b, the median, and κ. Compare with answers reached via the bi-normal approximation. Comment on the form and nature of the prior used here.

(g) The analysis has up to now taken all 141 life-times as having come from the same population; indeed Pearson remarks: 'In dealing with [these data] I have not ventured to separate the male and female mortality, the numbers being far too insignificant'. We shall nevertheless attempt to assess the difference between the male and female distributions, and specifically aim at making inference for the difference of medians

$$\delta = \text{med}_{\text{men}} - \text{med}_{\text{women}}.$$

Using a Weibull distribution for the men data, and another Weibull distribution for the women data, along with flat priors as in point (e), give a Bayesian analysis for the δ parameter.

(h) Which modelling approach appears best, of model M_1, using one Weibull model for the full data set, and model M_2, using one Weibull for the men and another Weibull for the women? There are various model selection strategies for answering such a question, but the approach to be taken now is as follows. Suppose these two models are a priori equally likely, and then 'let the data decide': find the posterior probabilities for models M_1 and M_2. When working with these, use again and for each case a uniform prior on $[10, 50] \times [0.10, 3.00]$ for the two Weibull parameters in question.

> – *Either man is constitutionally fitter to survive to-day [than two thousand years ago], or he is mentally fitter, i.e. better able to organise his civic surroundings. Both conclusions point perfectly definitely to an evolutionary progress.*
>
> *Karl Pearson*

A philatelic rarity: a first-day envelope with stamps commemorating the hundred years passing since Abel's death; cf. Exercise 2.