UNIVERSITETET I OSLO

Matematisk Institutt

EXAM IN:	**STK 4021/9021 – Bayesian Statistics**
	Part I of two parts: The project
WITH:	**Nils Lid Hjort**
TIME FOR EXAM:	**6.–19.xii.2013**

This is the exam project set for STK 4021/9021, autumn semester 2013. It is made available on the course website as of *Friday 6 December 12:00*, and candidates must submit their written reports by *Thursday 19 December 14:00* (or earlier), to the reception office at the Department of Mathematics, in duplicate. The supplementary four-hour written examination takes place *Monday December 9* (practical details concerning this are provided elsewhere). 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. Candidates are required to work on their own (i.e. without cooperation with any others). They 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. Note that *the STK 9021 students need to answer also Exercise 4*, whereas the STK 4021 students can confine their attention to Exercises 1–3.

Exercise 1

THE FIRST CONDITION OF PROGRESS is the removal of censorship, claims G.B. Shaw. In various statistics situations censored data cannot be avoided, however, as illustrated below. Assume first that certain lifetimes $y_1, \ldots, y_n$ of a type of technical components are independent and exponentially distributed with parameter θ, i.e. stemming from the density $\theta \exp(-\theta y)$ for y positive.

(a) Write down the log-likelihood function for these data, find the maximum likelihood estimator, and describe its precise and/or approximate distribution.

(b) Assume a Bayesian analysis of such data starts out with θ having a Gamma prior with parameters (a, b), i.e. with density

$$\frac{b^a}{\Gamma(a)}\theta^{a-1}\exp(-b\theta)$$

for θ positive. Derive the posterior distribution of θ. Find explicit formulae for $\widehat{\theta}_1$ and $\widehat{\theta}_2$, the Bayes estimators under loss functions

$$L_1(\theta, \widehat{\theta}) = (\widehat{\theta} - \theta)^2 \quad \text{and} \quad L_2(\theta, \widehat{\theta}) = \Big(\frac{1}{\widehat{\theta}} - \frac{1}{\theta}\Big)^2,$$

respectively. Comment on your results.

(c) Assume now that at the day when a statistical report must be written up not all lifetimes are actually observed, for the simple and healthy reason that some of the technical components under scrutiny are still working. Let $\delta_i = 1$ if y_i is observed, but $\delta_i = 0$ if what is observed concerning the lifelength of object i is only of the form $y_i > z_i$. The data may then be represented in the form (z_i, δ_i) for $i = 1, \ldots, n$, where $y_i = z_i$ if $\delta_i = 1$ (non-censored data) but $y_i > z_i$ if $\delta_i = 0$ (censored data). Explain that the likelihood takes the form

$$\prod_{i:\delta_i=1} \theta\exp(-\theta z_i) \prod_{i:\delta_i=0} \exp(-\theta z_i).$$

Find the posterior distribution for θ, after again having started from a Gamma prior (a, b), and give the Bayes estimator under squared error loss.

(d) Consider the following little dataset of (z_i, δ_i), having arisen as explained above (with time-scale being in 1000 hours, though this does not concern us here):

z	delta	z	delta
0.155	1	0.006	0
0.378	1	0.253	0
0.427	1	0.403	0
0.530	1	0.459	0
0.814	1	2.368	0

Starting with a Gamma prior π_0 with parameters $(0.1, 0.1)$, give the posterior distributions π_A using only the five first observations; π_B using only the last five observations; and finally the full posterior distribution π_{AB} using all ten observations. Display these three posterior densities in a diagram. Also compute and display the 0.05, 0.50, 0.95 quantiles for θ, for the prior, for the π_A, for the π_B, and finally for the π_{AB} distributions. Comment on your findings.

(e) Assume that there is one more observation, pertaining to technical object no. eleven, for which it is known that its lifetime is in the interval $[1.222, 1.666]$, but one did not manage to measure the time more accurately. Compute and display the posterior distribution π_{ABC} based on all eleven pieces of information, and supplement the little table above with 0.05, 0.50, 0.95 posterior quantiles for θ.

Exercise 2

FAILURE IS THE CONDIMENT that gives success its flavor (argues Truman Capote). A certain type of electronic equipment once in a while experiences a failure, under certain experimental conditions, which are varied from occasion to occasion in the course of nine experiments. These counts (taken from a more elaborate failure analysis data set for a certain Australian technical plant, some years ago) are as follows:

5, 1, 4, 14, 17, 17, 13, 7, 12.

We view these as independent Poisson counts, and our concern here is to estimate the associated intensity parameters, along with credibility intervals.

(a) Assume $y_1, \ldots, y_n$ are independent Poisson variables with parameters $\theta_1, \ldots, \theta_n$, and that these are to be estimated with loss function

$$L(\theta, \widehat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} (\widehat{\theta}_i - \theta_i)^2.$$

Show that the maximum likelihood estimators are the raw data themselves, i.e. $\widehat{\theta}_{i,\mathrm{ml}} = y_i$. Find the risk function (expected loss) $R(y_{\mathrm{ml}}, \theta)$ for this standard estimation strategy.

(b) Take now the parameters to be independent and stemming from a Gamma distribution prior, with parameters (a, b). Find the posterior distribution for the θ_i, and show that the Bayes estimators take the form

$$\widehat{\theta}_{i,\mathrm{B}} = \frac{a + y_i}{b + 1} \quad \text{for } i = 1, \ldots, n.$$

(c) Work out an explicit expression for the risk function $R(\widehat{\theta}_{\mathrm{B}}, \theta)$ for $\widehat{\theta}_{\mathrm{B}}$, in terms of the average and the variance

$$\bar{\theta} = \frac{1}{n} \sum_{i=1}^{n} \theta_i \quad \text{and} \quad V = \frac{1}{n} \sum_{i=1}^{n} (\theta_i - \bar{\theta})^2$$

of the parameters. Try also to characterise the part of the parameter space where $\widehat{\theta}_{\mathrm{B}}$ performs better than the raw estimator $\widehat{\theta}_{\mathrm{ml}}$.

(d) For the marginal distribution of y_i, with the Gamma prior above, show that

$$\mathrm{E}\, y_i = \frac{a}{b}, \quad \mathrm{Var}\, y_i = \frac{a}{b}\Big(1 + \frac{1}{b}\Big).$$

Argue from an empirical Bayes perspective that

$$\theta_i^* = \frac{b\bar{y} + y_i}{b + 1} \quad \text{for } i = 1, \ldots, n$$

is a natural class of estimators.

(e) When working with the estimator above,

$$\theta_i^* = \frac{b\bar{y} + y_i}{b+1} = w\bar{y} + (1-w)y_i \quad \text{for } i = 1, \ldots, n,$$

one may choose to view b here, or equivalently w, as merely an algorithmic parameter, identifying the θ^* estimator (i.e. θ^* here makes good sense even without the potential connection to the parameter of a Gamma prior). Work out an expression for the risk function $R(\theta^*, \theta)$ for θ^*, for a given $w = b/(b+1)$, again in terms of $\bar{\theta}$ and V. Use this to suggest a data-dependent value for the parameter w (or, equivalently, for b).

(f) Going back to the Gamma prior related use and interpretation of (a, b) in the formula for $\widehat{\theta}_{i,\mathrm{B}}$, use suitable empirical Bayes arguments to give a recipe for data-based values $(\widehat{a}, \widehat{b})$. For the technical component failure data above, compute the resulting empirical Bayes estimates $\widehat{\theta}_{i,\mathrm{eB}}$. Give in fact a 9×5 table, where the five columns contain

$$(y_i, \widehat{\theta}_{i,\mathrm{eB}}, q_i(0.05), q_i(0.50), q_i(0.95)),$$

where the q_i numbers are the relevant quantiles of the estimated posterior distribution for θ_i. Also, construct a suitable plot, where I suggest having experiments $1, \ldots, 9$ along the y-axis, then displaying estimates and intervals horizontally. You may attempt to more or less reconstruct my *Figure A* (page 7), where the experiments are also ordered by the size of the y_i.

(g) An alternative to the empirical Bayesian analysis above is a full Bayes construction, involving a background prior for (a, b). Your task now is to carry out such an analysis for the failure count data, where the setup is (i) taking a flat prior for (a, b); (ii) for given (a, b), having $\theta_1, \ldots, \theta_n$ independent from the Gamma (a, b); (iii) as above, for given $\theta_1, \ldots, \theta_n$, taking $y_1, \ldots, y_n$ independent and Poisson with these parameters. Use simulation to generate realisations $(a, b, \theta_1, \ldots, \theta_n)$ from the appropriate posterior distribution. Give another 9×5 table, supplementing the one from the previous point, containing

$$(y_i, \widehat{\theta}_{i,\mathrm{fB}}, r_i(0.05), r_i(0.50), r_i(0.95)),$$

where the $\widehat{\theta}_{i,\mathrm{fB}}$ are the associated Bayes estimates, and the r_i numbers the quantiles of the relevant posterior distributions. Make a plot of these results, similar to the one from the previous point, and comment on similarities and differences.

(h) The Bayes estimates $\widehat{\theta}_{i,\mathrm{B}}$ of point (b) associated with the Gamma (a, b) prior (which in later points led to certain empirical Bayes estimates) were derived under the loss function $L(\theta, \widehat{\theta})$ given in point (a). This loss function penalises errors equally, regardless of size; estimating a true parameter of 101.5 with the value 103.3, for example, is precisely as good as estimating a true parameter 1.5 with the value 3.3. For various Poisson count contexts, this might not be a satisfactory way of measuring loss. Derive a formula for the Bayes estimates $\widetilde{\theta}_{i,\mathrm{B}}$ for the θ_i with the alternative loss function

$$\widetilde{L}(\theta, \widehat{\theta}) = \sum_{i=1}^{n} (\widehat{\theta}_i - \theta_i)^2 / \bar{\theta}.$$

Exercise 3

I HAVE NOTHING TO OFFER but blood, toil, tears and sweat. The data pairs (x_0, y_0) given here relate to $n = 17$ patients in a study about factors influencing survival chances for a certain serious type of leukaemia. Here y_0 is the survival time after diagnosis, measured in weeks, and x_0 is the $\log_{10}$ white blood cell count. There were certain additional characteristics associated with this particular group of patients, but these do not concern us here.

x0	y0	x0	y0	x0	y0
3.36	65	4.00	121	4.54	22
2.88	156	4.23	4	5.00	1
3.63	100	3.73	39	5.00	1
3.41	134	3.85	143	4.72	5
3.78	16	3.97	56	5.00	65
4.02	108	4.51	26		

(a) Read the data into your computer, and transform weeks to years via $y = y_0/52$. Our basic model for these survival data will be that they are independent with

$$y_i \sim \mathrm{Gamma}(\exp(\beta_0 - \beta_1 x_i), \exp(\delta)) \quad \text{for } i = 1, \ldots, n,$$

where $x_i = x_{0,i} - \bar{x}_0$ and $\bar{x}_0 = n^{-1} \sum_{i=1}^n x_{0,i}$ the average value of the x_0. Write down the log-likelihood function for the data, as a function of the three model parameters $(\beta_0, \beta_1, \delta)$. Find the maximum likelihood estimates, e.g. by programming the log-likelihood function and using `nlm` in `R`. My programme yields estimates $(0.367, 1.117, 0.408)$. Use these estimates to reproduce a version of my *Figure B* (page 8), which displays the 17 data pairs $(x_{0,i}, y_i)$ along with the estimated curves

$$F^{-1}(0.05, x_0), \quad F^{-1}(0.50, x_0), \quad F^{-1}(0.95, x_0),$$

quantiles in the distribution of y given x_0, across the range of x_0 values. Comment briefly on this plot and what it conveys.

(b) We are to carry out Bayesian analysis based on this data set and the Gamma model above, and need a prior for $(\beta_0, \beta_1, \delta)$. You are now to try out a few simple priors, as part of the initial 'get to know your model' exercising before we attack the data. Make them simple, and to check whether they make sense (or perhaps not), generate survival times, for imaginary patients with x_0 equal to respectively $x_{0,\min} = \min_{i \le n} x_{0,i}$ and $x_{0,\max} = \max_{i \le n} x_{0,i}$. (It will not be necessary or required here to carry your investigations very far; the point is to learn which areas of the parameter space are reasonable and which are not.)

(c) Regardless of your investigations of the previous point, let now the prior for $(\beta_0, \beta_1, \delta)$ be essentially flat over the full parameter range (if insisting on a proper prior, one may take it as flat on $[-100, 100]^3$). Set up a Markov Chain Monte Carlo scheme to simulate realisations from the posterior distribution of $(\beta_0, \beta_1, \delta)$, and report on 0.05, 0.50, 0.95 quantiles for each parameter.

(d) For a patient with white blood cell count corresponding to $x_0 = 3.33$, use the MCMC simulations to display the distribution of (low, high), where

$$\text{low} = F^{-1}(0.05, x_0) \quad \text{and} \quad \text{high} = F^{-1}(0.95, x_0)$$

are quantiles of that person's survival distribution, and summarise this information in a suitable way. Compare this with the result of applying a 'lazy Bayesian' approximation, starting with the approximate normal distribution for $(\beta_0, \beta_1, \delta)$ obtained from the maximum likelihood procedure. Comment briefly on what you find.

(e) For this patient, with $x_0 = 3.33$, use posterior simulation to generate say 10^4 survival times after diagnosis. Display this distribution in a suitable plot, perhaps using `plot(density(lives))` or a fine histogram. Do the same for another patient, with $x_0 = 4.44$, and comment on your findings.

(f) The leukaemia data above have been used by several statisticians to illustrate various aspects of modelling techniques, methodological development, etc. In particular, other models than the one worked with above (which I have invented for the present purposes) have been used. D.R. Cox, for example, once used these data to illustrate the two-parameter model that takes the y_i to be independent and exponentially distributed, with parameters $\theta_i = \exp(\gamma_0 - \gamma_1 x_i)$. Assume that the model used above and Cox's model are a priori equally likely, and use the so-called Bayesian information criterion (BIC) to approximate the posterior probabilities for the two models. Which of the two models is more likely to be correct, the one you have worked with here, or Sir David's?

– I note here that `R` has preprogrammed various algorithms related to the gamma distribution, like `dgamma, pgamma, qgamma, rgamma`, which of course may be used when programming a log-likelihood function, finding quantiles, simulating realisations, etc., without having to start from scratch, so to speak. It may also be practical to be able to simulate realisations from the multinormal distribution, which may be done in `R` by first writing `library(MASS)` and then using `mvrnorm`.

Exercise 4 – for the PhD students taking STK 9021 only

The number of PhD candidates in the kingdom of Norway has more than doubled over the past ten years (from 4124 in 2002 via 7883 in 2008 to 9532 in 2012, actually). This is mindboggingly spellbindingly fantastic.

By the general rules of the Faculty of Mathematics and Natural Sciences those taking the PhD STK 9021 version of this course are required to be examined and evaluated in a somewhat more extensive manner from those taking the STK 4021 version. We solve this here by demanding that the STK 9021 candidates work also with the present Exercise 4 (those among the STK 4021 students eager to work with this exercise too are however welcomed to do so). This exercise is as follows.

I have uploaded Bradley Efron's 1986 article *Why Isn't Everyone a Bayesian?* to the course website, taken from the *American Statistician* journal, along with discussion contributions by Herman Chernoff, Dennis Lindley, Carl Morris, S. James Press, Adrian Smith, and Efron's rejoinder. Read through the Efron 1986 paper and ensuing discussion, and write up a short essay (perhaps two or three pages) where you (a) briefly sum up just a few points from this discussion and (b) choose one or two of these themes for further elaboration from your side. You are very much invited to present your own views as relevant for your own work (ongoing or prospective). I emphasise that you are not necessarily required to care about all of the details or aspects of the Efron 1986 discussion; you are instead supposed to find something concerning Bayes there worth discussing further from your own views or tastes, and of reasonable relevance for the STK 9021 course.

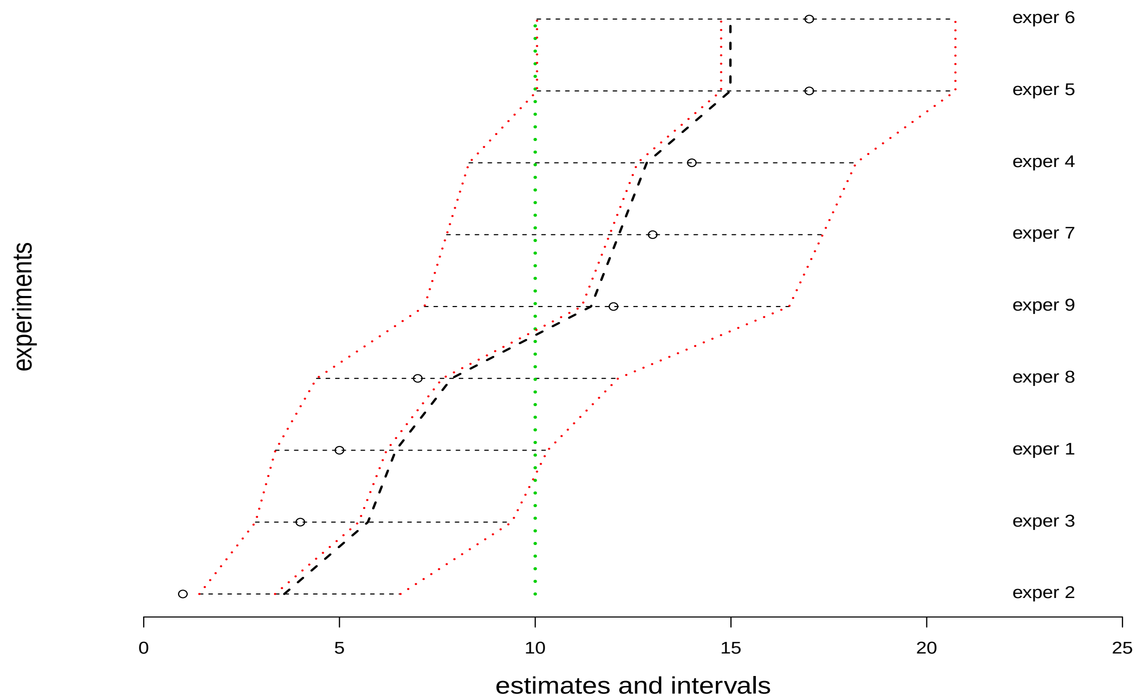

Figure A: The nine experiments reported on in Exercise 2 are ordered according to the size of the failure count y_i (with this value indicated by a small circle). The figure displays certain empirical Bayes estimates $\widehat{\theta}_{i,\mathrm{eB}}$ (the black, dotted line in the middle) along with 90% empirical Bayes credibility intervals and the posterior median estimates (red, dotted lines). The green vertical line corresponds to the overall average estimate $\bar{y}$. There are several viable empirical Bayes constructions, and the figure above reflects one of these schemes, which might not be identical to the one chosen by the exam candidate.

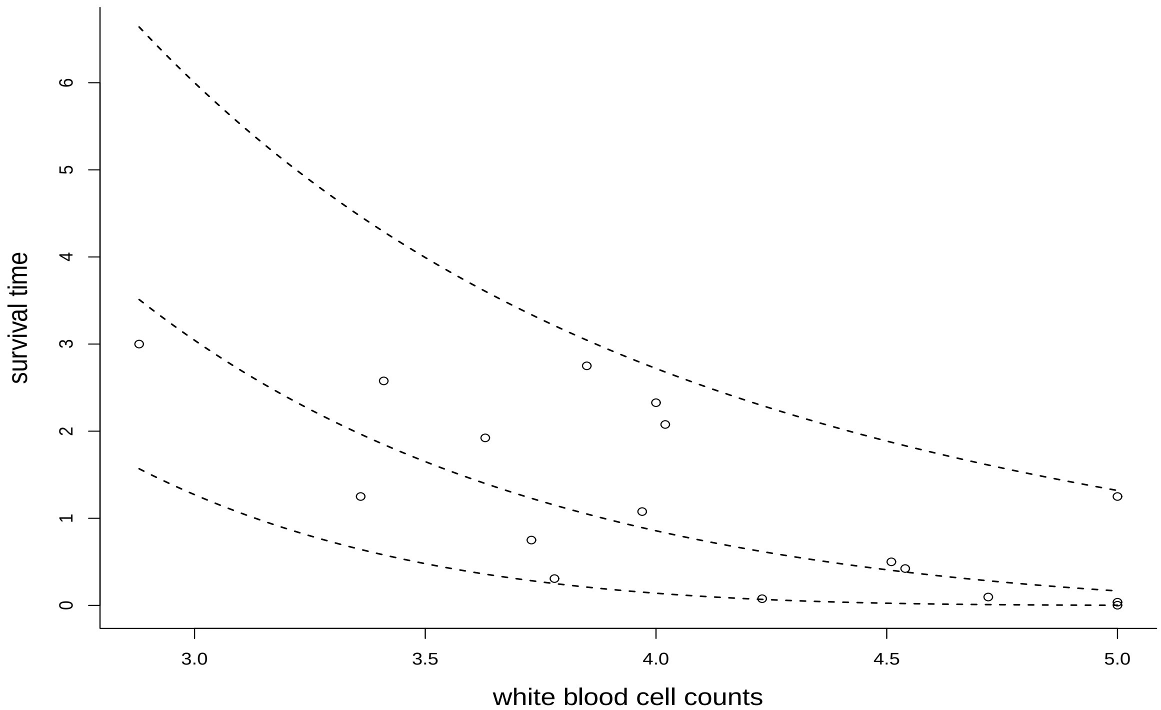

Figure B: The plot shows the data pairs (x_0, y) for the $n = 17$ patients, with x_0 the $\log_{10}$ white blood cell count and y the survival time after diagnosis, in years. Also displayed are the 0.05, 0.50, 0.95 quantiles of the estimated survival distribution, as a function of x_0.