UNIVERSITETET I OSLO

Matematisk Institutt

Exam in:	**STK 4021/9021 – Applied Bayesian Analysis and Numerical Methods**
	Part I of two parts: The project
With:	**Nils Lid Hjort**
Time for exam:	**30.xi.–12.xii.2017**

This is the exam project set for STK 4021/9021, autumn semester 2017. It is made available on the course website as of *Thursday 30 November 12:00*, and candidates must submit their written reports by *Tuesday 12 December 9:59* (or earlier), to the reception office at the Department of Mathematics (at Ullevål Stadion, Sognsveien 77B, 2nd floor), in duplicate. The supplementary four-hour written examination takes place *Wednesday 13 December* (practical details concerning this are provided elsewhere). Reports may be written in nynorsk, bokmål, riksmål, English, Latin, or verse, and should preferably be text-processed (TeX, LaTeX, Word), but may also be hand-processed. Give your *StudentWeb number* on the first page. Write concisely (краткость – сестра таланта; brevity is the soul of wit; in der Beschränkung zeigt sich erst der Meister). 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 should 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 seven pages (where the last page is a brief Appendix). Note that *the STK 9021 students need to answer also Exercise 3*, whereas the STK 4021 students can confine their attention to Exercises 1 and 2.

Exercise 1

Some kids are cooler than others, perhaps, and in certain dramatic cases cooling a newborn can save its life. More specifically, without going into the drastic physiological details, in some rare cases newborns are being critically deprived of oxygen to the brain as a consequence of a difficult birth. Pioneering research, involving in particular Professor of Systems Physiology Marianne Thoresen from the University of Oslo, has demonstrated that a form of 'cooling', where the little body has its temperature lowered to 33° Celsius during a certain period just after birth, can save its life, and with no loss of later mental or motoric capacities (you may check out `www.tv2.no/a/8732302/`). There is ongoing

research and controversy, however, regarding *the time window* where the cooling operation is helpful, or useless, or too late.

The questions dealt with in this exercise are meant to shed light on one of these discussions, following a study very recently reported on in *Journal of the American Medical Association* (Laptook et al., October 2017). They analysed various aspects of cooled and non-cooled newborns, where each case involved oxygen deprivation during birth, and where the cooling action, if taken, was initiated inside the time window 6 hours to 24 hours after birth (as opposed to starting earlier, which has been the general recommendation, so far). Luckily acidemic or moderate encephalopathy cases happen very rarely, so over a period of eight years, across a wide network of US hospitals, there are only a few hundred such cases. We shall avoid various extra details here and concentrate on a simple 2×2 table for two binomials,

$$\begin{aligned} y_0 &\sim \text{bin}(m_0, p_0), \quad \text{for a group of non-cooled newborns,} \\ y_1 &\sim \text{bin}(m_1, p_1), \quad \text{for a group of cooled newborns.} \end{aligned}$$

The event in question is 'death or disability', with a precise definition of disability, assessed when the child is about 18 months old. The actual numbers, to be used below, are 22 of 79 cases for group 0 (non-cooled) and 19 of 78 cases for group 1 (cooled). The emphasis is on the parameter

$$\rho = \frac{p_1}{p_0} \quad \text{(called rr, for relative risk, or risk ratio).}$$

(a) For the general binomial model (m, p), show that the Jeffreys prior is proportional to $1/\sqrt{p(1-p)}$, and that this is a Beta$(\frac{1}{2}, \frac{1}{2})$.

(b) With the Jeffreys prior for the two binomial experiments above, find the posterior distributions for p_0 and for p_1. Simulate 10^5 realisations of $\rho = p_1/p_0$ from the posterior distribution. Show a histogram for this posterior, and record the 0.025, 0.50, 0.975 quantiles. Comment on your findings.

(c) For a binomial $y \sim \text{bin}(m, p)$, with the usual estimator $\widehat{p} = y/m$, it is well known that

$$\sqrt{m}(\widehat{p} - p) \to_d \text{N}(0, p(1-p))$$

as m increases, with $\to_d$ denoting convergence in distribution (this famous result goes back to de Moivre, 1738). Use the delta method (see the Appendix of this exam set) to show that $\log \widehat{p}$ is approximately unbiased for $\log p$, approximately normal, with

$$\text{Var} \log \widehat{p} \approx \frac{1}{m}\Big(\frac{1}{p} - 1\Big).$$

Show also that a natural estimator for this variance is $1/y - 1/m$.

(d) For the cooling and non-cooling data, compute $\widehat{\rho} = \widehat{p}_1/\widehat{p}_0$, and also

$$\widehat{\gamma} = \log \widehat{\rho}$$

on the log-scale of $\gamma = \log \rho$. Give a formula for estimating the variance in the approximate normal distribution of $\widehat{\gamma}$, and compute the estimated standard deviation here. (You should find a number close to 0.269.)

(e) Use the above to put up an approximate 95% confidence interval ('classic', 'frequentist') for first γ, and then ρ. Comment on any similarity (or not) with the Jeffreys prior based Bayesian interval from question (b).

(f) Suppose now in general terms that γ is an unknown parameter of interest, and that an estimator $\widehat{\gamma}$ is such that $\widehat{\gamma} \,|\, \gamma \sim \mathrm{N}(\gamma, \sigma^2)$, with a precision σ taken to be known or very well estimated. With a prior distribution $\gamma \sim \mathrm{N}(\gamma_0, \sigma_0^2)$, show that $(\gamma, \widehat{\gamma})$ has a joint binormal distribution,

$$\begin{pmatrix} \gamma \\ \widehat{\gamma} \end{pmatrix} \sim \mathrm{N}_2(\begin{pmatrix} \gamma_0 \\ \gamma_0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_0^2 \\ \sigma_0^2 & \sigma_0^2 + \sigma^2 \end{pmatrix}).$$

Using established formulae for conditional distributions for the multinormal family, as e.g. given in the relevant Nils Exercises, show that this leads to

$$\gamma \,|\, \widehat{\gamma} \sim \mathrm{N}(\widehat{\gamma}_B, \sigma_B^2),$$

where

$$\widehat{\gamma}_B = \gamma_0 + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}(\widehat{\gamma} - \gamma_0) \quad \text{and} \quad \sigma_B^2 = \frac{\sigma_0^2 \sigma^2}{\sigma_0^2 + \sigma^2}.$$

(g) Consider the 'neutral prior' which takes $\gamma \sim \mathrm{N}(0, \sigma_0^2)$, with $\sigma_0 = 0.35$. Compute the associated 95% prior credibility interval for γ, on the scale of γ, and for the ρ scale.

(h) Compute the posterior distribution for ρ with the above prior. Attempt to construct and display a figure with three probability densities for the focus parameter ρ – the neutral prior; the posterior stemming from this neutral prior; and the posterior associated with the Jeffreys priors (i.e. the one worked with in question (b)).

(i) So what's the chance that the cooling apparatus is beneficial, even when it starts after 6 hours? A natural quantity to consider is the posterior probability that $p_1 < p_0$. With the Jeffreys based prior, compute

$$Q(19) = \Pr\{p_1 < p_0 \,|\, \text{data}\} = \Pr\{p_1 < p_0 \,|\, y_0 = 22, y_1 = 19\}$$

(with group sizes $m_0 = 79$ and $m_1 = 78$ considered fixed and given). For what range of potential other outcomes y_1 (which could have occurred), is

$$Q(y_1) = \Pr\{p_1 < p_0 \,|\, y_0 = 22, y_1\}$$

0.95 or higher?

(j) In the Laptook et al. (2017) journal article, they work with what they term a 'neutral prior', specified as above, with $\gamma = \log(p_1/p_0) \sim \mathrm{N}(0, 0.35^2)$. Find Beta priors $p_0 \sim \mathrm{Beta}(a_0, b_0)$, $p_1 \sim \mathrm{Beta}(a_1, b_1)$ matchings this, i.e. with the properties that the implied $\gamma = \log(p_1/p_0)$ has mean zero and standard deviation 0.35.

(k) Suppose NRK Dagsrevyen takes an interest in the cooling of neonates and in the issues pointed to above. Since they know you've now spent time and efforts analysing the data, looking into both Bayesian and non-Bayesian angles, they point a camera at you, wishing to know your conclusions. Please answer them – write a couple of paragraphs about this, related in particular to the subquestions below. The context is as in the Laptook et al. article and by implication in this exercise, relating to newborns with the described troubles during birth (with hypoxic-ischemic encephalopathy), and with the cooling machinery initiated inside the 6 hour to 24 hour time window.

[i] Is cooling better than not cooling?

[ii] What place and role does Bayesian analysis have in such discussions?

Exercise 2

SOLA DOSIS FACIT VENENUM, says Paracelsus, pointing to a tentative principle of toxiology (more or less that 'all things are poison, and nothing is without poison, the dosage alone makes it so a thing is not a poison'), and, in turn, to logistic regression type models. Assume for the present purposes that a certain compound is administered, at certain dosage levels, to Ratti norvegici laboratory rats. One then records whether Event A occurs or not, where Event A could be that the animal is seen as having become ill or worse (in which case it is sent back to better life and gentle healing). The uses of such experiments include finding pharmacologic and toxicologic information and practical rules relative to worker protection, how much coffeine humans can tolerate, discovering and testing physiological theories, etc.

The data to be analysed are as follows. At each of ten dosage levels $x_1, \ldots, x_{10}$, equal to

$$0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0,$$

$m = 5$ rats are exposed to the compound at that level, and the number of these five rats that experience Event A is

$$0, 2, 1, 3, 1, 2, 3, 3, 3, 2$$

respectively. We hence have ten binomial experiments, say

$$y_j \sim \mathrm{bin}(m, p_j) \quad \text{for } j = 1, \ldots, 10,$$

and these are modelled as

$$p_j = \Pr(A \,|\, x_j) = H(a + bx_j) = \frac{\exp(a + bx_j)}{1 + \exp(a + bx_j)},$$

with $H(u) = \exp(u)/\{1 + \exp(u)\}$ the logistic transform.

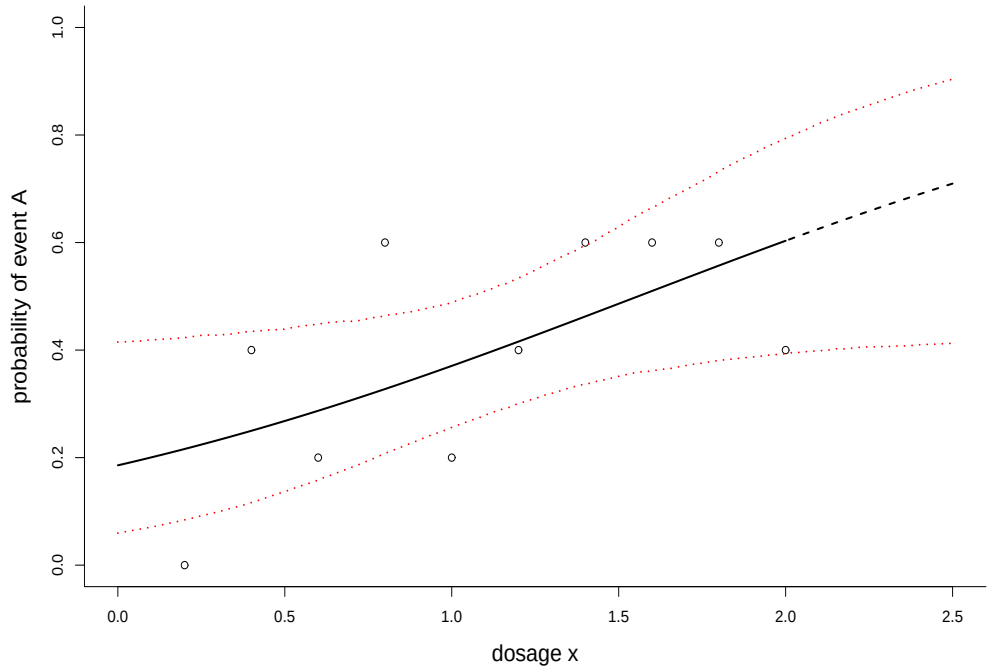
(a) Show that the log-likelihood function is

$$\ell(a,b) = \sum_{j=1}^{10} \Big[y_j \log p_j(a,b) + (m - y_j) \log\{1 - p_j(a,b)\} + \log \binom{m}{y_j} \Big].$$

Programme this log-likelihood function and find its maximisers. I find $\widehat{a} = -1.479$ and $\widehat{b} = 0.949$; the black full curve in the figure below represents the estimated probability

$$\widehat{p}(x) = H(\widehat{a} + \widehat{b}x),$$

plotted alongside the raw data estimates y_j/m.



(b) With a flat prior for (a,b) on $[-10,10] \times [-10,10]$, set up a Markov Chain Monte Carlo scheme to assess the posterior distribution of (a,b). Record the posterior means and posterior standard deviations, for the two model parameters, and compare to values obtained by the 'Lazy Bayesian' strategy, that of normal approximations from maximum likelihood theory.

(c) Use your simulations to produce a 90% pointwise credibility band around the estimated curve $\widehat{p}(x)$, with $\text{low}(x)$ the 0.05 quantile and $\text{up}(x)$ the 0.95 quantile of the posterior distribution for $H(a + bx)$; these are the red dotted curves I've plotted in the figure above.

(d) We also take an interest in $p(x) = \Pr(A \,|\, x)$ for higher dosage levels than for the range 0.2 to 2.0. Give the posterior distribution of $p(x_{\text{new}})$ for the high dosage level $x_{\text{new}} = 2.50$, in terms of a histogram or estimated density. Discuss briefly the assumptions underlying your analysis.

(e) In addition to finding the posterior distribution for $p(x_{\text{new}})$ above, find the predictive distribution for y_{new}, the number of $m = 5$ Ratti norvegici experiencing Event A when the dosage is $x_{\text{new}} = 2.50$.

(f) Above your Bayesian computations have been carried out with a flat prior for (a, b) on $[-10, 10] \times [-10, 10]$. In this particular context it is natural to assume that b cannot be negative, however, as more poison should increase the probability for Event A. Set up a second MCMC to compute the posterior distribution for (a, b) when the prior is flat on $[-10, 10] \times [0, 10]$. With this prior, compute by simulation the 0.05, 0.50, 0.95 quantile points of the posterior distribution for the point so-called LD50 parameter, or 'lethal dose 50-percent', the dosage level x_0 where 50% of the objects are expected to experience Event A.

Exercise 3 – for the PhD students taking STK 9021 only

The number of PhD candidates in the kingdom of Norway has increased with a factor of about 2.5 over the past fifteen years (from 4123 in 2002 via 7883 in 2008 to 9586 in 2014 and even to 10208 in 2016, actually). This is spellbindingly momentous.

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 than those taking the STK 4021 version. We solve this here by demanding that the STK 9021 candidates work also with the present Exercise 3 (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 Petra M. Kuhnert's article *Four case studies in using expert opinion to inform priors*, published in *Environmetrics* in 2011, to the course website. Your task is:

(a) read through the paper (without having to go into all details);

(b) give your own brief summary, perhaps two or three paragraphs, highlighting what *you* find of importance here;

(c) give some more details regarding what goes on, and what is important, for *one* of the four case studies of the paper;

(d) briefly describe something that interests *you*, with some characteristics perhaps similar to cases treated in the paper, and where *you* say something about how to get hold of expert opinions, along with *your* opinion regarding how relevant or how important such expert opinions could be for your models and analyses.

In total you could write perhaps three or four pages regarding this Exercise 3.

Appendix: the delta method

The delta method is a classical way of approximating or estimating variances for parameter estimators, through derivation and linearisation. One version of this is as follows, with $\to_d$ denoting convergence in distribution. If

$$\sqrt{n}(X_n - a) \to_d \mathrm{N}(0, \tau^2),$$

and $g(x)$ is a smooth function with derivative g', then

$$\sqrt{n}\{g(X_n) - g(a)\} \to_d \mathrm{N}(0, (\tau')^2), \quad \text{with} \quad \tau' = |g'(a)|\tau.$$

A translation of this precise limit result is as follows: Suppose X_n is unbiased or approximately unbiased for a, approximately normally distributed, and with variance approximately τ^2/n. Then $g(X_n)$ is approximately unbiased for $g(a)$, also approximately normal, and with variance approximately $g'(a)^2\tau^2/n$.

There are various useful generalisations of the statements above, e.g. to the case of g being a function of several different components, but we do not need these extensions here.