STK 4021-9021 Autumn 2021
Bayesian Analysis: the Oblig

This is The Oblig, the mandatory assignment, for STK 4021-9021, Bayesian Analysis, Autumn 2021. It is made available at the course website Tuesday October 19, and the submission deadline is Tuesday November 2, 13:58, *via the Canvas system.* Reports may be written in nynorsk, bokmål, riksmål, English, or Latin, should preferably be text-processed (for instance with TeX or LaTeX), and must be submitted as a single pdf file. The submission must contain your name, the course, and assignment number.

The Oblig set contains four exercises and comprises six pages (in addition to the present introduction page, 'page 0', and the last page is an Appendix).

It is expected that you give a clear presentation with all necessary explanations, but write concisely (in der Beschränkung zeigt sich erst der Meister; brevity is the soul of wit; краткость – сестра таланта). Remember to include all relevant plots and figures. These should preferably be placed inside the text, close to the relevant subquestion.

For a few of the questions setting up an appropriate computer programme might be part of your solution. The code ought to be handed in along with the rest of the written assignment; you might place the code in an appendix.

Students who fail the assignment but have made a genuine effort at solving the exercises are given a second attempt at revising their answers. All aids, including collaboration, are allowed, but the submission must be written by you and reflect your understanding of the subject. If we doubt that you have understood the content you have handed in, we may request that you give an oral account.

Application for postponed delivery: If you need to apply for a postponement of the submission deadline due to illness or other reasons, you have to contact the Student Administration at the Department of Mathematics (email: `studieinfo@math.uio.no`) well before the deadline.

The mandatory assignment in this course must be approved, in the same semester, before you are allowed to take the final examination.

Complete guidelines about delivery of mandatory assignments, along with a 'log on to Canvas', can be found here:

`www.uio.no/english/studies/admin/compulsory-activities/mn-math-mandatory.html`

Enjoy [imperative pluralis].

Nils Lid Hjort

1. The children of Odin

ODIN HAD SIX SONS (though sources are not entirely clear on the matter): Thor, Balder, Vitharr, Váli (cf. the Eddic poems and the Snorri Edda), Heimdallr, Bragi (cf. Snorri's kennings). But these sources are silent when it comes to any daughters. So how many children did he perhaps have, in total? We assume here that the if he had N children, then the number of boys y is binomial (N, p_0), with $p_0 = 0.515$, the same boy probability as in Scandinavia today.

(a) Put up the likelihood function, for the unknown N, with the given $y = 6$ boys. What is the maximum likelihood estimate?

(b) With the prior $p_0(N) \propto 1/(N+1)$, for $N = 0, 1, \ldots, 50$, find and portray the posterior distribution for the number of Odin's children. What is a 90 percent credibility interval for N?

(c) Think for at least three minutes in order to propose *your own prior*, $p_{\rm myown}(N)$. Compute and display the resulting posterior distribution, alongside the previous one.

(d) Comment on the assumptions underlying the calculations here.

2. Decision A, or B, or C?

CONSIDER THE SQUARE ROOT DISTRIBUTION for independent nonnegative observations $y_1, \ldots, y_n$, with density

$$f(y \,|\, \theta) = \frac{\theta}{2\sqrt{y}}\, e^{-\theta\sqrt{y}} \quad \text{for } y > 0,$$

where θ is an unknown positive parameter.

(a) Show that $f(y \,|\, \theta)$ indeed is a density, for each given θ. Write down the likelihood function for the observed data, and show that the maximum likelihood estimator is $\widehat\theta = 1/w_n$, where $w_n = (1/n)\sum_{i=1}^n \sqrt{y_i}$.

(b) Assume θ is given a Gamma prior, with parameters (a, b), i.e. with prior density

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

Find the posterior distribution for θ.

(c) Assume θ has a Gamma prior with carefully set parameters $(4.4, 2.2)$, and that twelve costly data points have been observed from the model:

0.771, 0.140, 0.135, 0.007, 0.088, 0.008, 0.268, 0.022, 0.131, 0.142, 0.421, 0.125

Display the prior and the posterior densities in a diagram. Compute also the probabilities p_1, p_2, p_3, that θ is in $(0, 1.50)$, or $(1.50, 3.00)$, or $(3.00, \infty)$, for the prior and then for the posterior.

(d) A certain institution needs to take a decision, in January 2022, related to the size of the parameter θ. The three possible decisions are A, business as usual; B, investing a certain high sum in some repair; C, investing a substantially higher sum in a more costly operation. The loss function, associated with future costs, in annual million kroner, is

$$L(\theta, A) = \begin{cases} 0 & \text{if } \theta \le 1.50, \\ 1 & \text{if } \theta > 1.50, \end{cases}$$
$$L(\theta, B) = \begin{cases} 0 & \text{if } \theta \in (1.50, 3.00), \\ 2 & \text{if } \theta \notin (1.50, 3.00), \end{cases}$$
$$L(\theta, C) = \begin{cases} 0 & \text{if } \theta > 3.00, \\ 3 & \text{if } \theta \le 3.00. \end{cases}$$

Which decision looked best, before the twelve data points were collected? Which decision is best, after having collected the data?

3. Bad-tempered and good-tempered men (and their wives)

Are bad-tempered men better at finding good-tempered women than the good-tempered men are? Or, to rephrase such a delicate and intricate question, do good-tempered women in their good-temperedness have a certain tendency to penetrate the shields of even bad-tempered men? Sir Francis Galton did not merely invent fingerprinting and correlation and regression and the two-dimensional normal distribution while working on anthropology and genetics and meteorology or exploring the tropics, but had a formidable appetite for even arcane psychometrics and for actually attempting to answer half-imprecise but good questions like the above in meaningful ways – by going out in the world to observe, to note, to think, to analyse (just as his perhaps even more famous cousin did).

On an inspired day in 1887 he therefore sat down and examined interview results pertaining to 111 married couples (see the Appendix), and classified the wives and husbands into 'bad-tempered' and 'good-tempered', reaching the following table:

		wife : good-tempered	wife : bad-tempered
husband :	good-tempered	24	27
	bad-tempered	34	26

He did not merely compute the proportions of relevance to the question raised above, but speculated about methods for answering whether the observed deviations from 'there is no difference' were statistically significant (some ten years before such concepts slowly began to take precise form in the statistics community). – Below it is your task to use the above simple dataset to help illustrate certain Bayesian techniques which might have interested Galton.

(a) Let us write

$$\begin{pmatrix} N_{0,0} & N_{0,1} \\ N_{1,0} & N_{1,1} \end{pmatrix} = \begin{pmatrix} 24 & 27 \\ 34 & 26 \end{pmatrix}$$

for the counts $N_{i,j} = \#\{X = i, Y = j\}$ for $i, j = 0, 1$, with X the good- (0) or bad-tempered (1) category for the husband and Y similarly the category for the wife. We shall take take the observed counts to be a random sample from the multinomial model with parameters $(n, p_{0,0}, p_{0,1}, p_{1,0}, p_{1,1})$, with $p_{i,j}$ interpreted as $\Pr\{X = i, Y = j\}$ for a randomly selected married couple (X, Y). Briefly discuss the validity of this assumption. Also give clear interpretations to the quantities

$$\alpha_i = p_{i,0} + p_{i,1} \quad \text{for } i = 0, 1,$$
$$\beta_j = p_{0,j} + p_{1,j} \quad \text{for } j = 0, 1.$$

(b) Show that the Jeffreys prior takes the form

$$\pi_{\text{jeff}}(p_{0,0}, p_{0,1}, p_{1,0}) \propto \frac{1}{\sqrt{p_{0,0}p_{0,1}p_{1,0}(1 - p_{0,0} - p_{0,1} - p_{1,0})}} = \frac{1}{\sqrt{p_{0,0}p_{0,1}p_{1,0}p_{1,1}}}$$

over the simplex where the $p_{i,j}$ are positive with $p_{0,0} + p_{0,1} + p_{1,0} < 1$. Also, show that this is the Dirichlet distribution for $p = (p_{0,0}, p_{0,1}, p_{1,0}, p_{1,1})$ with parameters $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$.

(c) We shall take an interest in the three parameters

$$\phi = \sum_{i,j} (p_{i,j} - \alpha_i \beta_j)^2 / p_{i,j}, \qquad \gamma = p_{0,0} + p_{1,1}, \qquad \delta = \frac{p_{0,1}p_{1,0}}{p_{0,0}p_{1,1}}.$$

Explain how these parameters may be interpreted in the present context. For the Jeffreys prior, use simulation to display the 0.05, 0.50, 0.95 quantiles of these parameters.

(d) Using Galton's data and the Jeffreys prior, derive the posterior distribution of the $(p_{0,0}, p_{0,1}, p_{1,0}, p_{1,1})$. Again via simulation, display the 0.05, 0.50, 0.95 quantiles for the posterior distribution of the three parameters ϕ, γ, δ. Sum up your findings.

(e) You're now invited to come up with *your own prior* for $(p_{0,0}, p_{0,1}, p_{1,0}, p_{1,1})$, which matches *your* prior beliefs concerning the world of married couples. For simplicity you are asked to choose your prior from the class of Dirichlet distributions, say $\text{Dir}(a_{0,0}, a_{0,1}, a_{1,0}, a_{1,1})$ with density

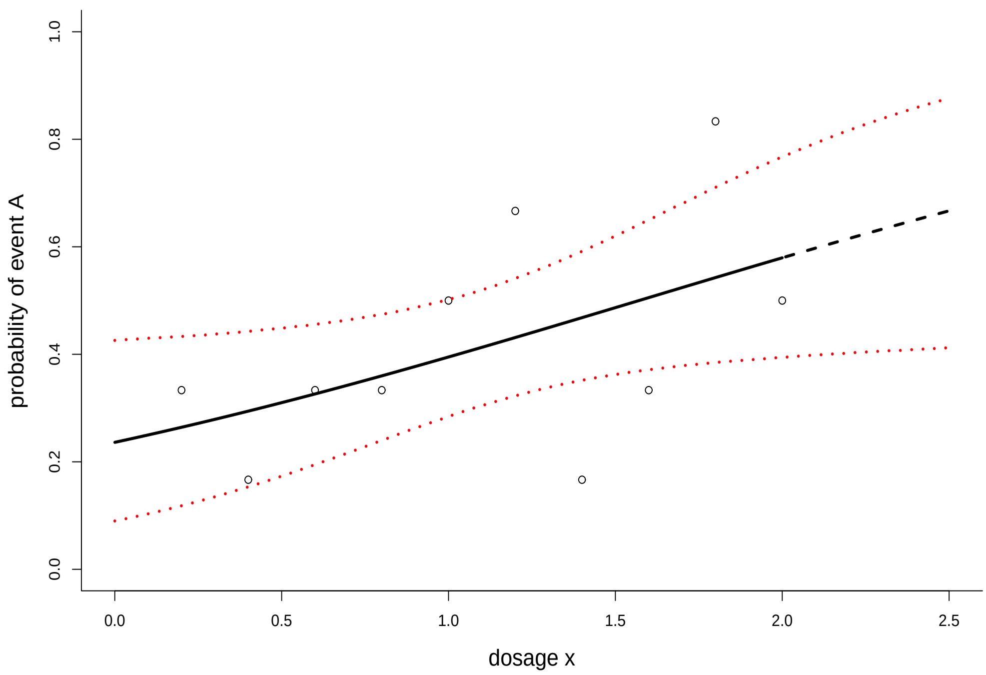
$$\frac{\Gamma(k)}{\Gamma(a_{0,0})\Gamma(a_{0,1})\Gamma(a_{1,0})\Gamma(a_{1,1})} p_{0,0}^{a_{0,0}-1} p_{0,1}^{a_{0,1}-1} p_{1,0}^{a_{1,0}-1} (1 - p_{0,0} - p_{0,1} - p_{1,0})^{a_{1,1}-1}$$

over the simplex where the $p_{i,j}$ are positive with $p_{0,0} + p_{0,1} + p_{1,0} < 1$; also, $k = a_{0,0} + a_{0,1} + a_{1,0} + a_{1,1}$. Discuss, but briefly, how you arrived at your prior.

(f) Please redo part of or all of the analysis using your own prior. Briefly discuss whether there are any noticeable discrepancies between the Jeffreys prior based analysis and that based on your own prior.

4. Rats

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 caffeine 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

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

$m = 6$ rats are exposed to the compound at that level, and the number of these six rats that experience Event A is

$$2,\ 1,\ 2,\ 2,\ 3,\ 4,\ 1,\ 2,\ 5,\ 3$$

respectively. We hence have ten binomial experiments, say

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

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 $\widehat{a} = -1.173$ and $\widehat{b} = 0.747$; the black full curve in the figure above 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 $[-8,8] \times [-8,8]$, 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 $\mathrm{low}(x)$ the 0.05 quantile and $\mathrm{up}(x)$ 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 $p(x_{\rm new})$ for the high dosage level $x_{\rm new} = 2.50$, in terms of a histogram or estimated density. Discuss briefly the assumptions underlying your analysis.

(e) In addition to finding the posterior distribution for $p(x_{\rm new})$ above, find the predictive distribution for $y_{\rm new}$, the number of $m = 6$ Ratti norvegici experiencing Event A when the dosage is $x_{\rm new} = 2.50$.

(f) Above your Bayesian computations have been carried out with a flat prior for (a,b) on $[-8,8] \times [-8,8]$. 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 $[-8,8] \times [0,8]$. 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.

Appendix: Measuring good-temperedness and bad-temperedness

To help assess or ascertain whether you or persons near you are good-tempered or bad-tempered, perhaps via interviews with members of your own family (across several generations), below are the criteria Galton instructed his data compilers to use. Matching a high enough number of epithets on the 'good' list makes you a good-tempered person, and correspondingly with the 'bad' list. Galton had such data for nearly two thousand individuals, and appears to have been specifically interested in the inheritance aspect, how and to what degree character traits are passed on to the next generation. The data set used for this project's Exercise 3 are extracted from these data, using the 111 married couples.

Good temper: amiable, buoyant, calm, cool, equable, forbearing, gentle, good, mild, placid, self-controlled, submissive, sunny, timid, yielding (15 epithets in all).

Bad temper: acrimonious, aggressive, arbitrary, bickering, capricious, captious, choleric, contentious, crotchety, decisive, despotic, domineering, easily offended, fiery, fits of anger, gloomy, grumpy, harsh, hasty, headstrong, huffy, impatient, imperative, impetuous, insane temper, irritable, morose, nagging, obstinate, odd-tempered, passionate, peevish, peppery, proud, pugnacious, quarrelsome, quick-tempered, scolding, short, sharp, sulky, sullen, surly, uncertain, vicious, vindictive (46 epithets in all).

Discussing these criteria at some length, Galton includes the following comment: 'We can hardly, too, help speculating uneasily upon the terms that our own relatives would select as most appropriate to our particular selves.'