10th October, 2023

STK4021/9021 — Applied Bayesian Analysis

Mandatory assignment

Submission deadline

Thursday 26th October 2023, 14:30 in Canvas (canvas.uio.no).

Instructions

Note that you have **one attempt** to pass the assignment. This means that there are no second attempts.

All students should attempt Problems 1–3. Only students taking STK9021 should attempt Problem 4.

You can choose between scanning handwritten notes or typing the solution directly on a computer (for instance with LaTeX). The assignment must be submitted as a single PDF file. Scanned pages must be clearly legible. The submission must contain your name, course and assignment number.

It is expected that you give a clear presentation with all necessary explanations. Remember to include all relevant plots and figures. 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.

In exercises where you are asked to write a computer program, you need to hand in the code along with the rest of the assignment. It is important that the submitted program contains a trial run, so that it is easy to see the result of the code.

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 (e-mail: studieinfo@math.uio.no) no later than the same day as the deadline.

All mandatory assignments 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:

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

GOOD LUCK!

Problem 1. This problem starts with two general questions about Bayesian inference, followed by a more specific problem.

(a) Is it true that if two model parameters are independent in the prior, then they are also independent in the posterior? Give a proof or a counterexample.

(b) Suppose we sample $\theta \sim \pi(\cdot)$ from the prior, $y \sim \pi(\cdot \mid \theta)$ from the forward model and then $\theta' \sim \pi(\cdot \mid y)$ from the posterior, given y. What is the distribution of θ'?

We now turn to a more specific problem. Consider the square root distribution for independent nonnegative observations $y_1, \ldots, y_n$, with density

$$f(y, \theta) = \frac{\theta}{2\sqrt{y}} e^{-\theta\sqrt{y}} \quad \text{for } y > 0, \tag{1}$$

where $\theta > 0$ is an unknown parameter.

(c) Show that $f(y, \theta)$ indeed defines a probability density function. Show that its mean is $2/\theta^2$, and find also its median.

(d) Suppose there are independent observations $y_1, \ldots, y_n$ from the density above. Write down an expression for the log-likelihood function and find a formula for the maximum likelihood estimator $\widehat{\theta}$. Also find expressions for the exact and/or approximate distribution for this estimator.

(e) Assume next that a Gamma prior distribution is elicited for θ, with some parameters (a, b), i.e. of the form $\{b^a/\Gamma(a)\}\theta^{a-1}\exp(-b\theta)$ for θ positive. Find the posterior distribution for θ given the observations $y_1, \ldots, y_n$.

(f) To learn more about the consequences of having a Gamma type prior with the square root distribution, find a formula for the marginal density of y, when $y \mid \theta$ has the density (1) and $\theta \sim \text{Gamma}(a, b)$. Find in particular such a formula for both the density and the cumulative distribution function of y in the case of $(a, b) = (1, 1)$.

(g) Assume now that 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.

(h) Give a formula for the predictive density for y_{13}, and find the 0.10, 0.50, 0.90 quantiles of this distribution (perhaps by simulation).

(i) A certain institution needs to take a decision, in January 2024, 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

$$L(\theta, \mathrm{A}) = \begin{cases} 0 & \text{if } \theta \leq 1.50, \\ 1 & \text{if } \theta > 1.50, \end{cases}$$

$$L(\theta, \mathrm{B}) = \begin{cases} 0 & \text{if } \theta \in (1.50, 3.00), \\ 2 & \text{if } \theta \notin (1.50, 3.00), \end{cases}$$

$$L(\theta, \mathrm{C}) = \begin{cases} 0 & \text{if } \theta > 3.00, \\ 3 & \text{if } \theta \leq 3.00. \end{cases}$$

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

Problem 2. Let $y = (y_1, \ldots, y_n)$ be observed data and suppose we have k competing models $m_1, \ldots, m_k$. For $j = 1, \ldots, k$, the parameters of model m_j are θ_j, living in the parameter space Θ_j. Furthermore, write $\pi_j(\theta_j)$ and $\pi_j(y \mid \theta_j)$ for the m_j prior density and likelihood, respectively. Finally, let P_j be the prior probability assigned to model m_j. The full parameter space is

$$\Omega = \bigcup_{j=1}^{k} \bigcup_{\theta \in \Theta_j} \{(\theta, m_j)\}.$$

(a) Show that the marginal likelihood for the full model takes the form

$$\pi(y) = P_1 \pi_1(y) + \cdots + P_k \pi_k(y),$$

where

$$\pi_j(y) = \int_{\Theta_j} \pi_j(y \mid \theta_j) \pi_j(\theta_j) \, \mathrm{d}\theta_j$$

is the model m_j marginal likelihood.

(b) Show also that the posterior probabilities assigned to the models can be written as

$$P_j^\star = \pi(m_j \mid y) = \frac{P_j \pi_j(y)}{P_1 \pi_1(y) + \cdots + P_k \pi_k(y)} = \frac{P_j \pi_j(y)}{\pi(y)},$$

for $j = 1, \ldots, k$.

Is there a decline in the number of skiing days per year in Nordmarka? Download the `bjornholt.csv` data set, available on the course webpage. These data contain the number of days per year at which there were more than 25 cm of snow at Bjørnholt, Nordmarka. The measurements started in 1897 and have continued until 2022, apart from a gap between 1938 and 1954, so the total number of measurements is $n = 109$. We let $\mathbf{x} = (x_1, \ldots, x_n)^\top$ denote the years and $\mathbf{y} = (y_1, \ldots, y_n)^\top$ denote the number of skiing days, respectively. The data are displayed in Figure 1.

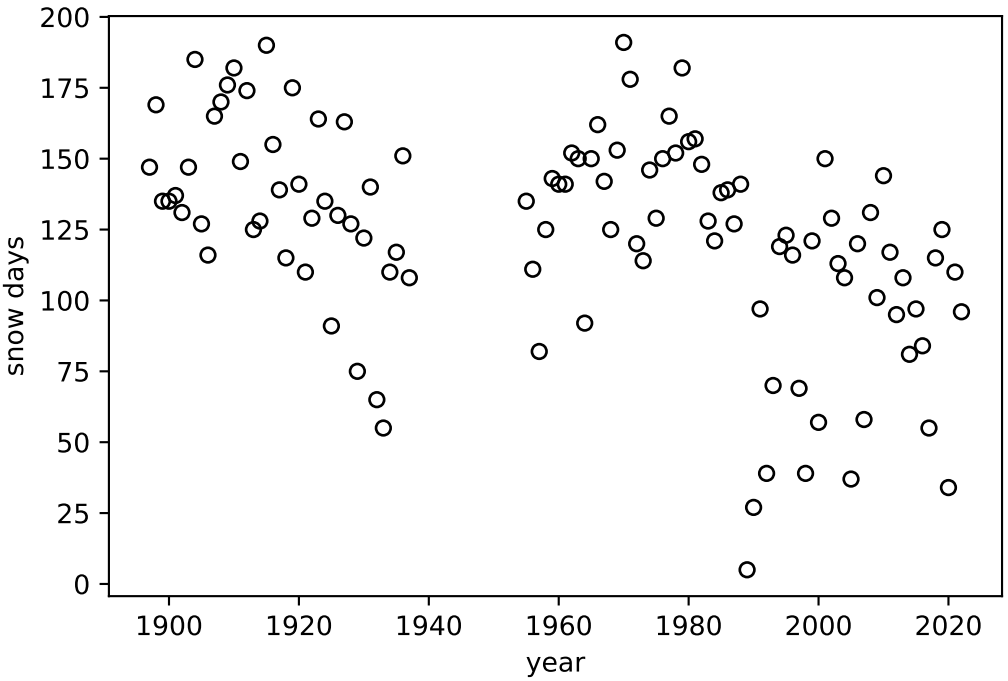

Figure 1: The wondrously falling snow flakes [you may sing Schubert's Der Leiermann here, or simply listen to the recording by Ian Bostridge and Leif Ove Andsnes] represent skiing days per winter season at Bjørnholt, from 1897 to 2022, but with no data for the seasons 1938 to 1954.

(c) First consider linear regression,

$$y_i = w_0 + w_1(x_i - 1900) + \varepsilon_i,$$

where $\varepsilon_1, \ldots, \varepsilon_n \sim N(0, \beta^{-1})$, independently. We can think of this as polynomial regression with $p = 2$ and feature maps

$$\phi_j(x) = (x - 1900)^j,$$

for $j = 0, 1, \ldots, p-1$. Writing $\mathbf{w} = (w_0, w_1)^\top$, use

$$\pi(\mathbf{w}) = \mathrm{N}(\mathbf{w}; \mathbf{0}, \alpha^{-1} I)$$

as prior, and recall that the likelihood takes the form

$$\pi(\mathbf{y} \mid \mathbf{w}) = N(\mathbf{y}; \Phi\mathbf{w}, \beta^{-1} I),$$

where

$$\Phi = \begin{pmatrix} \phi_0(x_1) & \phi_1(x_1) & \cdots & \phi_{p-1}(x_1) \\ \phi_0(x_2) & \phi_1(x_2) & \cdots & \phi_{p-1}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(x_n) & \phi_1(x_n) & \cdots & \phi_{p-1}(x_n) \end{pmatrix}$$

is the design matrix.

Use empirical Bayes to find optimal values of the hyperparameters α, β. What is the resulting value of the log marginal likelihood $\log \pi(\mathbf{y})$?

(d) Find the posterior distribution for $\mathbf{w} = (w_0, w_1)^\top$. Let $x' = 2023$, the current year, and let y' be the number of skiing days. Find the predictive distribution $\pi(y' \mid \mathbf{y})$, and give a 95% credibility interval for y'. In a diagram, plot the predictive mean for all years between 1897 and 2023, along with bands spanning $\pm$ one predictive standard deviation.

(e) As an alternative hypothesis to a declining trend, there might just be a constant trend, so $p = 1$ and $\mathbf{w} = (w_0)$. Repeat the analysis from (c) and (d) using the constant model, including the plot of the predictive distribution.

(f) Now we will compare the two models. Let $m = 0$ and $m = 1$ denote the constant and linear models, respectively. Use $P_0 = P_1 = 1/2$ as prior model probabilities.

Write down the full parameter space Ω.

What is the full log marginal likelihood? What are the posterior model probabilities? Which model should we choose? *Hint: To evaluate the full log marginal likelihood, use the* `LogSumExp` *trick.*

(g) Discuss briefly strengths and weaknesses with the analysis you have performed in this exercise.

Let us return to regression with a fixed value of p (the first level of inference, in the terminology of MacKay (1992)).

(h) Consider the predictive distribution $\pi(y' \mid \mathbf{y})$ of the output y' of a new input x'. Recall that this distribution has variance

$$\sigma_n^2(x') = \frac{1}{\beta} + \phi(x') S_n \phi(x'),$$

where

$$S_n^{-1} = \alpha I + \Phi^\top \Phi$$

is the posterior precision matrix. Use the Sherman-Morrison formula for a nonsingular $p \times p$ matrix M and p-dimensional vectors $\mathbf{u}$ and $\mathbf{v}$,

$$(M + \mathbf{u}\mathbf{v}^\top)^{-1} = M^{-1} - \frac{M^{-1}\mathbf{u}\mathbf{v}^\top M^{-1}}{1 + \mathbf{v}^\top M^{-1}\mathbf{u}},$$

to show that

$$\sigma_{n+1}^2(x') \leq \sigma_n^2(x').$$

What does this result tell us?

Problem 3. Let $\alpha > 0$ be a fixed constant, and consider a sequence $X_1, X_2, \ldots$ of random variables generated as follows. For the first variable, we have that $X_1 \sim P$, where P is some fixed continuous probability distribution. Suppose now that we have generated $X_1 = x_1, \ldots, X_n = x_n$. Then X_{n+1} is generated by

$$X_{n+1} \mid \{X_1 = x_1, \ldots, X_n = x_n\} \sim \frac{\alpha}{\alpha + n} P + \frac{1}{\alpha + n} \sum_{i=1}^{n} \delta_{x_i}, \tag{2}$$

where δ_x is the degenerate distribution at the point x. That is, with probability $\alpha/(\alpha+n)$, $X_{n+1} \sim P$, and with probability $n/(\alpha+n)$, $X_{n+1} \sim \text{Uniform}\{x_1, \ldots, x_n\}$.

(a) What is the probability that $X_1 = X_2$? What is the probability that $X_1 = X_2 = \cdots = X_n$?

(b) Use the results from (a) to show that the X_i are not independent.

(c) For the values $\alpha \in \{1, 10, 100\}$ generate a sample of size $n = 1000$, using $P = \text{Beta}(1, 2)$. Count the number of unique values in each sample. How is the value of α related to this number?

(d) Show that the number K_n of unique values in the sample $X_1, \ldots, X_n$ can be written as

$$K_n = \sum_{i=1}^{n} B_i,$$

where $B_i \sim \text{Bernoulli}(\alpha/(\alpha + i - 1))$, independently, for $i = 1, \ldots, n$.

(e) In addition to counting the number of unique values K_n in the sample $X_1, \ldots, X_n$, we are also interested in the *clustering* of the sample. For example, for $n = 8$, we could have that

$$X_1 = X_3 = X_8, \quad X_2 = X_5, \quad X_4 = X_6 = X_7,$$

with three distinct values in each cluster (so $X_1 \neq X_2$ etc.) We write this event as

$$C = \{\{1, 3, 8\}, \{2, 7\}, \{4, 5, 6\}\}.$$

Find $\mathbb{P}(C)$, the probability of observing C as the clustering of $X_1, \ldots, X_8$.

(f) Now let

$$D = \{\{1, 2, 5\}, \{3, 7, 8\}, \{4, 6\}\}.$$

Show that $\mathbb{P}(C) = \mathbb{P}(D)$.

(g) Deduce that the sequence $X_1, X_2, \ldots$ is exchangeable.

(h) With $P = \text{Beta}(1, 2)$ again, what is the probability that $X_{2023} \in [1/3, 2/3]$?

(i) Show that for any (measurable) sets A and B, we have

$$\mathbb{P}(X_1 \in A, X_2 \in B) = \frac{\alpha}{\alpha + 1} P(A)P(B) + \frac{1}{\alpha + 1} P(A \cap B).$$

Problem 4. This problem is for STK9021 students only. Write a short essay (between three and six pages excluding references) on a part of the syllabus that you particularly enjoyed, or found particularly interesting. You are highly encouraged to explain how some of the themes of the syllabus relate to your own research project, if this is the case. In addition, you are encouraged to reflect on some of the broader themes of the course, including:

- Frequentist versus Bayesian inference (you may find Gelman (2008) an interesting read)
- The first and second levels of inference, (MacKay, 1992),
- The two cultures (Breiman, 2001).

Bibliography

Leo Breiman. Statistical modeling: the two cultures. *Statistical Science*, 16:199–231, 2001.

Andrew Gelman. Objections to Bayesian statistics. *Bayesian Analysis*, 3:445–449, 2008.

David J. C. MacKay. Bayesian interpolation. *Neural Computation*, 4:415–447, 1992.