

STK4021/9021 — Applied Bayesian Analysis

Mandatory assignment

Submission deadline

Thursday 24th October 2024, 12:00 (noon) on 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 1–4. Only students taking STK9021 should attempt Problem 5.

You can choose between scanning handwritten notes or typing the solution directly on a computer (for instance with L^AT_EX). 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. Let y be a Bernoulli random variable with parameter $\theta \in (0, 1)$, so y has the mass function $f_\theta(y) = \theta^y(1 - \theta)^{1-y}$, for $y = 0, 1$.

- (a) Find the mean and variance of y , as functions of θ .

We want to estimate θ using the loss function $L(t, \theta) = (t - \theta)^2 / \{\theta(1 - \theta)\}$.

- (b) Verify that the maximum likelihood estimator for θ is y itself, and that this estimator has unit (frequentist) risk. That is, show that $R(y, \theta) = 1$.

Now introduce the Bayesian machinery. Let $\theta \sim \text{Beta}(a, b)$ in the prior distribution, where $a, b > 0$. That is, θ has prior density

$$\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1}, \quad \text{for } 0 < \theta < 1.$$

- (c) Show that $\mathbb{E}[\theta] = a/(a+b)$ and

$$\mathbb{E}[1/\theta] = \begin{cases} (a+b-1)/(a-1) & \text{if } a > 1 \\ \infty & \text{otherwise.} \end{cases}$$

- (d) Find the distribution of $\gamma = 1 - \theta$.

- (e) Show that $\pi(\theta)$ is a conjugate prior, and find the posterior distribution $\pi(\theta | y)$.

- (f) From now on, assume that $a, b > 1$. Show that the Bayes estimator is given by

$$\hat{\theta} = (a + y - 1) / (a + b - 1).$$

- (g) Show that the minimal Bayes risk is $1/(a + b - 1)$.

Problem 2. *I sing of Artemis, whose shafts are of gold, who cheers on the hounds, the pure maiden, shooter of stags, who delights in archery, own sister to Apollo with the golden sword. Over the shadowy hills and windy peaks she draws her golden bow, rejoicing in the chase, and sends out grievous shafts.*¹

Artemis is said to be the greatest bow-and-arrow archer of all the Greek gods. One day, she discusses the art of archery with Dionysus, the god of wine. As Dionysus courteously listens, Artemis explains how she has perfected her control of the trajectory of a shooting arrow. If she shoots an arrow with initial velocity v_0 (measured in metres per second) at an angle $0 < \theta < \pi/2$ from the ground, then (ignoring air resistance), the arrow perfectly follows a trajectory in two-dimensional space parametrised by

$$\gamma(t) = \begin{pmatrix} v_0 \cos(\theta)t \\ v_0 \sin(\theta)t - \frac{1}{2}gt^2 \end{pmatrix}, \quad (1)$$

where t is time (in seconds) and $g = 9.81 \text{ ms}^{-2}$ is the gravitational constant on Earth. Artemis proudly proclaims that she is able to control the initial velocity v_0 and angle θ perfectly, such that in theory, her arrows will travel precisely the same distance every time.

¹From *Homeric Hymns and Homerica* (translated by H. G. Evelyn-White). Harvard University Press, 1914.

- (a) Show that if an arrow follows the parametrisation (1), then it will travel a distance $x = \{v_0^2/g\} \sin(2\theta)$ before it lands.

Dionysus is impressed, but skeptical. He is willing to believe that Artemis can perfectly control the initial velocity v_0 , but he doubts whether she is able to shoot at precisely the same angle every time. Being the god of wine, he is quite skeptical of the concept of aiming in general. To meet his skepticism, Artemis claims that yesterday, she fired $n = 10$ arrows with exactly the same initial velocity $v_0 = 20.5 \text{ ms}^{-1}$ and at precisely the same angle θ , whose precise value she cannot remember. She takes up her notebook and shows Dionysus how far (in metres) the arrows travelled:

43.50	41.50	43.00	41.82	44.55
42.86	41.77	43.22	41.63	44.02

Dionysus smiles cunningly as he believes these numbers clearly indicate that the angles from which Artemis has fired the arrows are not the same. Artemis, however, reminds him that in the real world, there is air resistance, and so they would expect the arrows to land in slightly different places even if the angle is constant. The two confer with the four gods of wind — Zephyros, Notos, Boreas, and Euros — who explain to them that yesterday, the wind was arranged precisely such that if an arrow was fired with initial velocity v_0 at an angle θ from the ground, then the distance it would travel follows a normal distribution with mean $\{v_0^2/g\} \sin(2\theta)$ and variance $\sigma^2 = 1.1^2$. Furthermore, they assure Artemis and Dionysus that the distances travelled by different arrows are independent.

- (b) Assume that the distances $x_1, \dots, x_n$ are distributed as the gods of wind have declared. Write down the log-likelihood function $\ell(\theta)$ and show that the equation $\ell'(\theta) = 0$ has the roots

$$\theta = \pi/4, \quad \theta = \frac{1}{2} \arcsin(g\bar{x}/v_0^2), \quad \theta = \pi/2 - \frac{1}{2} \arcsin(g\bar{x}/v_0^2),$$

where $\bar{x} = n^{-1} \sum_{i=1}^n x_i$.

- (c) Show further, with full mathematical detail, that the maximum likelihood estimator $\hat{\theta}$ is given by

$$\hat{\theta} = \begin{cases} \pi/4 & \text{if } \bar{x} > v_0^2/g \\ \pi/4 \pm \left\{ \pi/4 - \frac{1}{2} \arcsin(g\bar{x}/v_0^2) \right\} & \text{otherwise.} \end{cases}$$

What is the maximum likelihood estimate for the actual $n = 10$ datapoints given above?

- (d) In this case, we ended up with quite a strange maximum likelihood estimator. Maybe things become a bit clearer if we turn to a Bayesian formulation. Impose a uniform prior $\pi(\theta)$ on θ on the domain $(0, \pi/2)$. Find the Laplace approximation of the posterior distribution (where you will have to take into account that θ_{MAP} may be multivalued).

- (e) Although we have not covered Markov chain Monte Carlo theory in lectures yet, we will nevertheless apply such an algorithm here. Let S be some large number, say $S = 10^5$. Consider the following procedure:

```

1: Sample  $\theta_1 \sim \pi(\theta)$ 
2: for  $s = 1, \dots, S - 1$  do
3:    $\theta' \sim \pi(\theta)$  ▷ Generate proposal
4:    $\alpha \leftarrow \exp\{\ell(\theta') - \ell(\theta_s)\}$ 
5:    $u \sim \text{Uniform}[0, 1]$ 
6:   if  $u \leq \alpha$  then
7:      $\theta_{s+1} \leftarrow \theta'$  ▷ Accept proposal
8:   else
9:      $\theta_{s+1} \leftarrow \theta_s$  ▷ Reject proposal
10:  end if
11: end for

```

It turns out that we can essentially treat the samples $\theta_1, \dots, \theta_S$ as a sample from the true posterior distribution $\pi(\theta | x_1, \dots, x_n)$. Implement this algorithm in your code, and plot the posterior density (either using histograms or an empirical density).

- (f) Include a plot of the density yielded by the Laplace approximation. How well does it approximate the true posterior? Can you think of any minor adjustments to improve it? Are there any further points of criticism towards the Laplace approximation you can think of in this problem?

In order to test the hypothesis of whether θ can be treated as constant, Artemis and Dionysus first need to know which angle Dionysus was attempting to aim at. They confer with the Mnemosyne, the god of memories, who informs them that Artemis was attempting to aim at an angle of $\theta_0 = \pi/4 + 0.05$ from the ground. Dionysus therefore suggests using the prior

$$\theta \sim \frac{1}{2}\text{Uniform}(0, \pi/2) + \frac{1}{2}\delta_{\theta_0},$$

where δ_x means the degenerate prior taking all its mass at the point x .

- (g) Using this prior, find the posterior probability that $\theta = \theta_0$, given the data. Should Dionysus believe Artemis' claim?
- (h) In the analysis above, the first prior component $\text{Uniform}(0, \pi/2)$ is perhaps a bit too broad. Therefore, repeat part (g), but with the prior

$$\theta \sim \frac{1}{2}\text{Uniform}(\theta_0 - 0.1, \theta_0 + 0.1) + \frac{1}{2}\delta_{\theta_0}.$$

Does your conclusion change?

The remaining parts of Problem 2 are optional.

- (i) Ignore air resistance. Show that the trajectory of an arrow fired with initial velocity v_0 at an angle θ from the ground follows the curve

$$y = f_\theta(x) = (\tan \theta)x - \frac{1}{2} \frac{g}{v_0^2} (1 + \tan^2 \theta)x^2.$$

- (j) Let $\mathcal{S}$ be the union of all possible trajectories. That is,

$$\mathcal{S} = \left\{ (x, y) \in \mathbb{R}^2 : x, y > 0 \text{ and } y = f_\theta(x) \text{ for, some } \theta \in (0, \pi/2) \right\}.$$

Show that the boundary of the set $\mathcal{S}$ is a parabola, and find its equation.

Problem 3. In this question, let $y_1, \dots, y_n$ be independent one-dimensional Gaussian random variables with mean θ and variance σ^2 .

- (a) Suppose that σ^2 is a known parameter, and that we wish to estimate θ . Write down the log-likelihood $\ell(\theta)$ and verify that the maximum likelihood estimator is given by $\hat{\theta} = \bar{y} = n^{-1} \sum_{i=1}^n y_i$.
- (b) Impose a Gaussian prior distribution on θ , say $\theta \sim N(\mu, \tau^2)$. Show (either from scratch or via tools from lectures), that the posterior distribution of $\theta | y_1, \dots, y_n$ is also a Gaussian distribution, and find its mean μ_n and variance σ_n^2 .
- (c) Show that we can write $\mu_n = \lambda_n \mu + (1 - \lambda_n) \hat{\theta}$, where $0 < \lambda_n < 1$ is a constant depending on n . Find the limit of λ_n as $n \rightarrow \infty$, and interpret what this result means.
- (d) Let y' be another Gaussian observation, with mean θ and variance σ^2 , which is independent of $y_1, \dots, y_n$. Show that predictive distribution of $y' | y_1, \dots, y_n$ is yet another Gaussian distribution, with mean μ_n and variance $\sigma^2 + \sigma_n^2$.
- (e) Let $\gamma = 1/\sigma^2$ be the precision of the data model, so that $y_i \sim N(\theta, \gamma^{-1})$, independently for $i = 1, \dots, n$. Now suppose θ is known, and that we instead wish to estimate γ . Write down the log-likelihood function $h(\gamma)$ and verify that the maximum likelihood estimator is given by $\hat{\gamma}^{-1} = n^{-1} \sum_{i=1}^n (y_i - \theta)^2$.
- (f) Impose a Gamma prior on γ , say $\gamma \sim \text{Gamma}(a, b)$, with $a, b > 0$, so that

$$\pi(\gamma) = \frac{b^a}{\Gamma(a)} \gamma^{a-1} e^{-b\gamma}, \quad \gamma > 0.$$

Verify that $\mathbb{E}[\gamma] = a/b$ in the prior.

- (g) Show that the posterior distribution of $\gamma | y_1, \dots, y_n$ is also a Gamma distribution, and find its parameters a_n and b_n .
- (h) Like you did in part (c), show that the posterior mean of γ can also be written as a convex combination of the prior mean $\mathbb{E}[\gamma]$ and the maximum likelihood estimator $\hat{\gamma}$.
- (i) Let $n = 1$, so we only have a single Gaussian random variable $y \sim N(\theta, \gamma^{-1})$. Show that marginally, y has a Student's t-distribution,

$$\pi(y) = \frac{\Gamma(\nu/2 + 1/2)}{\Gamma(\nu/2)} \left(\frac{\delta}{\pi\nu} \right)^{1/2} \left\{ 1 + \frac{\delta(y - \theta)^2}{\nu} \right\}^{-\nu/2 - 1/2},$$

where $\delta = a/b$ and $\nu = 2a$. [Hint: use the substitution $z = \gamma[b + (y - \theta)^2/2]$ when integrating.]

- (j) Show that as $\pi(y)$ converges to a Gaussian density with mean θ and precision δ as $\nu \rightarrow \infty$. [You are allowed to assume that $\lim_{\nu \rightarrow \infty} \pi(y)$ is also a probability density.]

Problem 4. Let θ be a one-dimensional parameter with prior distribution $\pi(\theta)$, and suppose we have data y from a generative model $\pi(y | \theta)$.

- (a) Assume that θ is a continuous variable (so $\pi(\theta)$ denotes its density). For $0 < p < 1$, define the loss function L by

$$L(t, \theta) = \{p\mathbb{1}_{t \leq \theta} + (1-p)\mathbb{1}_{t > \theta}\} |t - \theta| = \begin{cases} p(\theta - t) & \text{if } t \leq \theta, \\ (1-p)(t - \theta) & \text{if } t > \theta. \end{cases}$$

Show that the Bayes estimator is given by the posterior p -quantile. What happens when $p = 1/2$?

[Hint: Use the Leibniz integral rule.]

- (b) Now assume that θ is a discrete variable (so $\pi(\theta)$ denotes its probability mass function). Define the loss function L by

$$L(t, \theta) = 1 - \mathbb{1}_{\theta=t} = \begin{cases} 0 & \text{if } \theta = t, \\ 1 & \text{otherwise.} \end{cases}$$

Show that the Bayes estimator is given by

$$\hat{\theta} = \arg \max_t \mathbb{P}(\theta = t | y).$$

Problem 5. This problem is for STK9021 students only. Write a short essay (between two and five 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.