

Bayesian Inference: Assignment 1

Answers to the starred (*) questions should be submitted to Canvas by no later than **4pm, Friday 23rd February 2024**. Solutions to non-assessed questions will be made available in class and/or in Canvas.

Question 1*

Nasa's new *Earth Surface Mineral Dust Source Investigation* instrument investigates how airborne dust particles affect the climate. Since July 2020, the instrument has revealed that more than a billion metric tonnes – the weight of 10,000 aircraft carriers – of mineral dust are carried through the atmosphere from deserts and dry regions to other areas on earth. One area that has been particularly badly affected is southeastern Europe, in particular Greece, Cyprus and Turkey.

In a joint project, data scientists at the Cyprus Institute and Newcastle University are interested in the probability θ that particulate matter in the air, during and immediately following a dust storm, will result in an environmental health warning being issued for certain at-risk groups (i.e. people suffering from chronic lung conditions such as asthma).

- (a) After extensive discussions with a team of scientists – including meteorologists and epidemiologists – prior beliefs regarding θ were expressed according to the probability density function shown below in Figure 1. Look at it, and then answer the following questions.
- (i) In plain English, briefly describe what Figure 1 tells us in terms of the project team's prior beliefs about θ . **[2 marks]**
- (ii) The density shown in Figure 1 is that of a *Beta distribution*. Choose the most appropriate prior from the list below, and clearly explain your choice: **[2 marks]**
- $\theta \sim Be(20, 77)$ or $\theta \sim Be(17, 5)$ or $\theta \sim Be(20, 20)$ or $\theta \sim Be(5, 17)$.
- (b) Of the last ten dust storms to affect southeastern Europe, four resulted in environmental health warnings (in particular, the storm affecting Crete and Cyprus in summer 2021).
- (i) Assuming these storms occur independently, write down the likelihood function for θ . **[2 marks]**
- (ii) Find the posterior distribution for θ , and use plots/sketches and numerical summaries to help describe changes in beliefs about θ in light of this information. **[4 marks]**
- (c) Find the posterior distribution for θ assuming an appropriate ignorance prior. How does this compare to the posterior obtained using the informative prior? **[3 marks]**

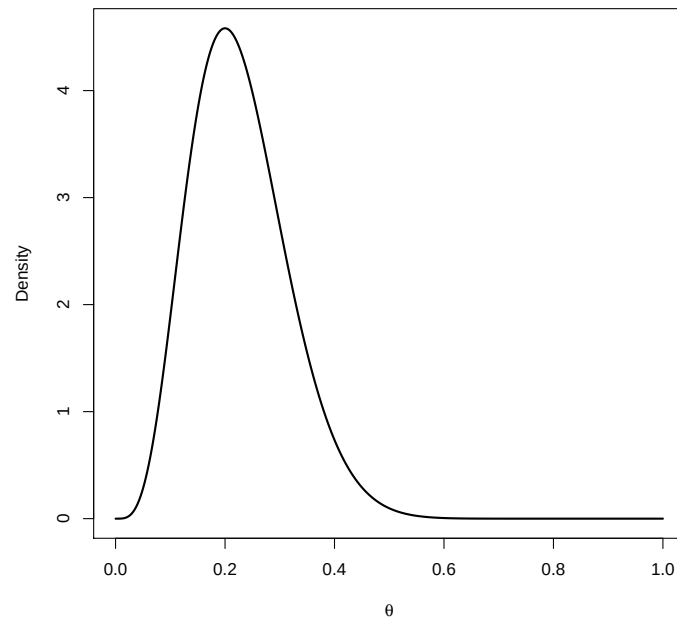


Figure 1: Elicited prior density representing beliefs about θ .

Question 2*

Suppose that a random sample $x_1, x_2, \dots, x_n$ is obtained from a truncated unit exponential distribution, with density

$$f(x|\theta) = \begin{cases} e^{\theta-x}, & x > \theta \\ 0, & \text{otherwise,} \end{cases}$$

where $\theta \in \mathbb{R}$. Derive the posterior density for θ assuming a normal $N(b, 1/d)$ prior distribution. [7 marks]

Hint:

$$\theta^2 - 2b\theta - \frac{2n\theta}{d} \propto \left(\theta - \frac{db + n}{d} \right)^2.$$

Question 3

Suppose you have a random sample $x_1, x_2, \dots, x_n$. In the following models, derive the posterior density and name the posterior distribution (and its parameters).

(i) $f(x|\theta) = \theta^{x-1}(1-\theta)$, $x = 1, 2, \dots$ with a $Beta(3, 2)$ prior distribution for θ .

(ii) $f(x|\theta) = \frac{e^{-\theta}\theta^x}{x!}$, $x = 0, 1, \dots$ with a $Exp(2)$ prior distribution for θ .

(iii) $f(x|\theta) = \binom{4}{x} \theta^x (1 - \theta)^{4-x}$, $x = 0, 1, 2, 3, 4$ with a $U(0, 1)$ prior distribution for θ .

Question 4

Suppose that a random sample $x_1, x_2, \dots, x_{10}$ is obtained from a uniform $U(0, \theta)$ distribution. Derive the posterior density for θ assuming a gamma $Ga(20, 1)$ prior distribution θ . Hint: this distribution will depend on the maximum observed x -value $x_{\max}$.

Question 5

The dimensions of a component from a long production run vary according to a $N(\mu, 1)$ distribution, and the mean dimension μ varies from production run to production run according to a $N(10, 1/4)$ distribution. From one production run 12 components are drawn at random and their average dimension is found to be $10\frac{1}{3}$. On this information what is the probability that the mean component dimension is at least 10?

Question 6

Suppose that the time in minutes required to serve a customer at a certain facility has an exponential distribution for which the parameter θ is unknown, and that the prior distribution of θ is a gamma distribution with mean 0.2 and standard deviation 1. If the average time required to serve a random sample of 20 customers is observed to be 3.8 minutes, determine the posterior distribution of θ .

Question 7

The data in Table 1 show the time intervals (in minutes) between eruptions of the "Old Faithful" geyser. Note that the average time between eruptions is 67.38 minutes, that is, 1.123 hours. If the eruptions occur randomly in time according to a Poisson process with a rate of θ eruptions per hour then the times between eruptions will form a random sample from an exponential distribution with rate θ . Making this assumption, establish a prior distribution for θ with $E(\theta) = 1$ and $Var(\theta) = 1$. Determine the posterior distribution for θ . Comment on your analysis.

70	64	72	76	80	48	88	53	71	56	69	72	76	54	76
65	54	86	40	87	49	76	51	77	49	71	78	80	51	82
49	80	43	83	49	75	47	78	71	69	63	64	82	68	71
71	63	79	66	75	56	83	67	65	77	72	79	73	53	69
53	78	55	67	68	73	53	70	69	66	79	48	90	49	78
52	79	49	75	75	50	87	40	76	57	71	70	69	72	51
84	43	73	73	70	84	71	79	58	73					

Table 1: Time intervals between eruptions of the “Old Faithful” geyser.