

MA40189 - Question Sheet Two

Simon Shaw, s.shaw@bath.ac.uk
<https://people.bath.ac.uk/masss/ma40189.html>

2021/22 Semester II

Set: Problems class, Thursday 24th February 2022.

Due in: Problems class, Thursday 3rd March 2022. Paper copies may be submitted in the problems class or directly to me either in lectures or my office, 4W4.10. Pdf copies may be submitted to the portal available on the [Moodle page](#).

Task: Attempt questions 1-2; questions 3-4 are extra questions which may be discussed in the problems class.

1. Let X_1, \dots, X_n be conditionally independent given λ so that $f(x|\lambda) = \prod_{i=1}^n f(x_i|\lambda)$ where $x = (x_1, \dots, x_n)$. Suppose that $\lambda \sim \text{Gamma}(\alpha, \beta)$ and $X_i|\lambda \sim \text{Exp}(\lambda)$ where λ represents the rate so that $E(X_i|\lambda) = \lambda^{-1}$.
 - (a) Show that $\lambda|x \sim \text{Gamma}(\alpha + n, \beta + n\bar{x})$.
 - (b) Show that the posterior mean for the failure rate λ can be written as a weighted average of the prior mean of λ and the maximum likelihood estimate, \bar{x}^{-1} , of λ .

A water company is interested in the failure rate of water pipes. They ask two groups of engineers about their prior beliefs about the failure rate. The first group believe the mean failure rate is $\frac{1}{8}$ with coefficient of variation $\frac{1}{\sqrt{11}}$, whilst the second group believe the mean is $\frac{1}{11}$ with coefficient of variation $\frac{1}{2}$.

[Note: The coefficient of variation is the standard deviation divided by the mean.]

Let X_i be the time until water pipe i fails and assume that the X_i follow the exponential likelihood model described above. A sample of five of pipes is taken and the following times to failure were observed: 8.2, 9.2, 11.2, 9.8, 10.1.

- (c) Find the appropriate members of the Gamma families the prior statements of the two groups of engineers represent. In each case find the posterior mean and variance. Approximating the posterior by $N(E(\lambda|x), \text{Var}(\lambda|x))$, where $x = (x_1, \dots, x_5)$, estimate, in each case, the probability that the failure rate is less than 0.1.
 - (d) How do you expect any differences between the engineers to be reconciled as more data becomes available?
2. Let x be the number of successes in n independent Bernoulli trials, each one having unknown probability θ of success. It is judged that θ may be modelled by a $\text{Unif}(0, 1)$ distribution so

$$f(\theta) = 1, \quad 0 < \theta < 1.$$

An extra trial, z is performed, independent of the first n given θ , but with probability $\frac{\theta}{2}$ of success. The full data is thus (x, z) where $z = 1$ if the extra trial is a success and 0 otherwise.

(a) Show that

$$f(\theta | x, z = 0) = c\{\theta^{\alpha-1}(1-\theta)^{\beta-1} + \theta^{\alpha-1}(1-\theta)^\beta\}$$

where $\alpha = x + 1$, $\beta = n - x + 1$ and $c = \frac{1}{B(\alpha, \beta) + B(\alpha, \beta + 1)}$.

(b) Hence show that

$$E(\theta | X, Z = 0) = \frac{(x+1)(2n-x+4)}{(n+3)(2n-x+3)}.$$

[Hint: Show that $c = \frac{\alpha+\beta}{B(\alpha, \beta)(\alpha+2\beta)}$ and work with α and β .]

(c) Show that, for all x , $E(\theta | X, Z = 0)$ is less than $E(\theta | X, Z = 1)$.

3. Let X_1, \dots, X_n be conditionally independent given θ , so $f(x | \theta) = \prod_{i=1}^n f(x_i | \theta)$ where $x = (x_1, \dots, x_n)$, with each $X_i | \theta \sim N(\mu, \theta)$ where μ is known.

(a) Let $s(x) = \sum_{i=1}^n (x_i - \mu)^2$. Show that we can write

$$f(x | \theta) = g(s, \theta)h(x)$$

where $g(s, \theta)$ depends upon $s(x)$ and θ and $h(x)$ does not depend upon θ but may depend upon x . The equation shows that $s(X) = \sum_{i=1}^n (X_i - \mu)^2$ is sufficient for X_1, \dots, X_n for learning about θ .

(b) An inverse-gamma distribution with known parameters $\alpha, \beta > 0$ is judged to be the prior distribution for θ . So,

$$f(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{-(\alpha+1)} e^{-\beta/\theta}, \quad \theta > 0.$$

i. Show that the distribution of the precision $\tau = \frac{1}{\theta}$ is *Gamma*(α, β).

ii. Find the posterior distribution of θ given $x = (x_1, \dots, x_n)$.

iii. Show that the posterior mean for θ can be written as a weighted average of the prior mean of θ and the maximum likelihood estimate, $s(x)/n$, of θ .

4. Suppose that X_1, \dots, X_n are identically distributed discrete random variables taking k possible values with probabilities $\theta_1, \dots, \theta_k$. Inference is required about $\theta = (\theta_1, \dots, \theta_k)$ where $\sum_{j=1}^k \theta_j = 1$.

(a) Assuming that the X_i s are independent given θ , explain why

$$f(x | \theta) \propto \prod_{j=1}^k \theta_j^{n_j}$$

where $x = (x_1, \dots, x_n)$ and n_j is the number of x_i s observed to take the j th possible value.

- (b) Suppose that the prior for θ is Dirichlet distributed with known parameters $a = (a_1, \dots, a_k)$ so

$$f(\theta) = \frac{1}{B(a)} \prod_{j=1}^k \theta_j^{a_j-1}$$

where $B(a) = B(a_1, \dots, a_k) = \frac{\prod_{j=1}^k \Gamma(a_j)}{\Gamma(\sum_{j=1}^k a_j)}$. Show that the posterior for θ given x is Dirichlet with parameters $a + n = (a_1 + n_1, \dots, a_k + n_k)$.