Petter Mostad
Matematisk Statistik
Chalmers

Suggested solution for exam in MSA100/MVE185Computer Intensive Statistical Methods, October 2009

1. (a) The posterior is

$$\pi(\lambda \mid y) \propto \pi(y \mid \lambda)\pi(\lambda) \propto_\lambda \lambda^y \exp(-t\lambda)\pi(\lambda)$$

For the 4 possible values of λ, we get

$$\begin{array}{lcl} 0.1 & : & 0.1^2 \exp(-7 \cdot 0.1) \cdot 0.1 = 0.000497 \\ 0.2 & : & 0.2^2 \exp(-7 \cdot 0.2) \cdot 0.3 = 0.002959 \\ 0.3 & : & 0.3^2 \exp(-7 \cdot 0.3) \cdot 0.5 = 0.005511 \\ 0.4 & : & 0.4^2 \exp(-7 \cdot 0.4) \cdot 0.1 = 0.000973 \end{array}$$

The sum of these values is 0.00994. Adjusting each value by dividing by the sum we get the posterior

λ	0.1	0.2	0.3	0.4
Probability	0.05	0.30	0.55	0.10

(b) The prior predictive distribution at zero is

$$\begin{aligned} \pi(0) &= \sum_{\lambda=0.1}^{0.4} \pi(0 \mid \lambda)\pi(\lambda) \\ &= \sum_{\lambda=0.1}^{0.4} \exp(-7\lambda)\pi(\lambda) \\ &= \exp(-7 \cdot 0.1) \cdot 0.1 + \exp(-7 \cdot 0.2) \cdot 0.3 \\ &\quad + \exp(-7 \cdot 0.3) \cdot 0.5 + \exp(-7 \cdot 0.4) \cdot 0.1 \\ &= 0.19 \end{aligned}$$

2. (a) We get

$$\begin{aligned} \pi(\mu \mid \text{data}) &\propto \pi(\text{data} \mid \mu) \\ &\propto \Phi(165 \mid \mu, 10)^2 \cdot [\Phi(175 \mid \mu, 10) - \Phi(165 \mid \mu, 10)]^5 \\ &\quad \cdot (1 - \Phi(175 \mid \mu, 10))^3 \end{aligned}$$

(b) Several suggestions are acceptable. The simplest is to numerically compute the function above on a dense set of evenly spaced values in some reasonable interval, say 0 - 300, divide by the sum, and use the resulting discrete distribution as an approximation for the posterior.

(c) As there are 5 lengths in the interval between 165 and 175, the posterior probability of very large σ must be small. Similarly, as there are both lengths below 165 and lengths above 175, the posterior probability of very small σ must also be small. Regarding

μ, the data clearly would give small posterior probabilities to values far away from the interval $[165, 175]$. Together, these observations would lead to a guess that the posterior should be proper, as the answer to the first question.

To make such an argument more precise, one could for example observe that, for fixed μ and large σ, the factors in the likelihood are of the order $\frac{1}{2}^2$, $\frac{1}{\sigma}^5$, and $\frac{1}{2}^3$, respectively, so that the posterior would be of the order σ^{-6}. For small σ, one could show that the posterior would approach 0 as σ approaches zero.

The intuitive answer to the last question is that the posterior is improper. The reason is that the data in this form would not contain any information about the spread of the actual lengths. In particular, the given data would not limit the possibility of very large σ. More mathematically, for fixed μ, the likelihood will approach a positive constant when σ grows large. Thus the posterior would behave like $1/\sigma$ for such large σ, and the posterior would be improper.

3. (a) We get that

$$\begin{aligned}
\pi(p_A, p_B \mid data) &\propto_{p_A,p_B} \pi(data \mid p_A, p_B) \\
&\propto_{p_A,p_B} p_A^{22}(1-p_A)^{78} p_B^{12}(1-p_B)^{88} \\
&\propto_{p_A,p_B} \frac{\Gamma(102)}{\Gamma(23)\Gamma(79)} p_A^{22}(1-p_A)^{78} \frac{\Gamma(102)}{\Gamma(13)\Gamma(89)} p_B^{12}(1-p_B)^{88}.
\end{aligned}$$

Thus the posterior is proportional to the product of two Beta distributions, with

$$\begin{aligned}
p_A \mid data &\sim \text{Gamma}(23, 79) \\
p_B \mid data &\sim \text{Gamma}(13, 89).
\end{aligned}$$

As the product of these distributions clearly integrate to 1, we see that the posterior is in fact equal to the product, and not just proportional to it.

(b) Simulate 10000 values from the Beta$(23, 79)$ distribution, and put them in a vector P_A. Independently, simulate 10000 values from the Beta$(13, 89)$ distribution, and put them in a vector P_B. Compute the 10000 quotients P_B/P_A, and use the sample 2.5% and 97.5% quantiles of this set of numbers as the boundaries of the credibility interval. In R,

```
> PA <- rbeta(10000, 23, 79)
> PB <- rbeta(10000, 13, 89)
> quantile(PB/PA, c(0.025, 0.975))
     2.5%     97.5%
0.2914252 1.0266977
```

(c) Similar as above, except that now we find the empirical proportion when $P_A > P_B$. In R,

```
> sum(PA>PB)/10000
[1] 0.9688
```

4. The code computes the approximate expected value of func(X), where X is a random variable with a distributioni given by the `posterior` function. The algorithm implemented is called Importance Sampling.

5. The purpose of the Metropolis-Hastings algorithm is to simulate from a distribution for a parameter θ based on $f(\theta)$, a function proportional to the density function. The algorithm produces a chain of values $\theta_1, \ldots, \theta_t, \ldots$ with stationary distribution equal to the distribution in question. The main ingredient is a proposal function $J(\theta_* \mid \theta_t)$ providing a probability distribution for θ_* given the value of θ_t at each step. The new value θ_{t+1} is set equal to θ_* with probability

$$r = \min\left(1, \frac{f(\theta_*)J(\theta_t \mid \theta_*)}{f(\theta_t)J(\theta_* \mid \theta_t)}\right)$$

or otherwise set equal to θ_t. (More can be written, but such a short explanation is enough here).

6. (a) The prior can be re-written as a normal distribution:

$$\begin{aligned} f(\theta) &= \exp(-2\theta^2 + \theta - 3) \\ &\propto_\theta \exp\left(-2(\theta - \frac{1}{4})^2\right) \\ &\propto_\theta \frac{1}{\sqrt{2\pi\frac{1}{4}}}\exp\left(-\frac{1}{2\cdot\frac{1}{4}}(\theta - \frac{1}{4})^2\right). \end{aligned}$$

Thus the prior is normal with expectation $\frac{1}{4}$ and variance $\frac{1}{4}$. The two observations $y_1 = 1$ and $y_2 = 4$ are normally distributed with expectation θ and variance 1. Using standard formulas, we then get that

$$\begin{aligned} \theta \mid y_1 &\sim \mathcal{N}\left(\frac{1\cdot 1 + 4\cdot\frac{1}{4}}{1+4}, \frac{1}{1+4}\right) \sim \mathcal{N}\left(\frac{2}{5}, \frac{1}{5}\right) \\ \theta \mid y_1, y_2 &\sim \mathcal{N}\left(\frac{1\cdot 4 + 5\cdot\frac{2}{5}}{1+5}, \frac{1}{1+5}\right) \sim \mathcal{N}\left(1, \frac{1}{6}\right). \end{aligned}$$

Alternatively, and faster, one can observe that the likelihood of the data is the same as for a single observation of $\frac{5}{2}$ with a normal distribution with expectation θ and variance $\frac{1}{2}$. Then we get directly

$$\theta \mid y_1, y_2 \sim \mathcal{N}\left(\frac{2\cdot\frac{5}{2} + 4\cdot\frac{1}{4}}{2+4}, \frac{1}{2+4}\right) \sim \mathcal{N}\left(1, \frac{1}{6}\right).$$

(b) As the prior for θ is a normal distribution $\mathcal{N}(\frac{1}{4}, \frac{1}{4})$, and observations y have a normal distribution with expectation θ and variance 1, we know that y has a normal distribution, and the expectation and variance of its distribution can be found as

$$\begin{aligned} \mathrm{E}(y) &= \mathrm{E}(\mathrm{E}(y \mid \theta)) = \mathrm{E}(\theta) = \frac{1}{4} \\ \mathrm{Var}(y) &= \mathrm{Var}(\mathrm{E}(y \mid \theta)) + \mathrm{E}(\mathrm{Var}(y \mid \theta)) = \mathrm{Var}(\theta) + \mathrm{E}(1) = \frac{1}{4} + 1 = \frac{5}{4}. \end{aligned}$$

So the prior predictive distribution is $\mathcal{N}(\frac{1}{4}, \frac{5}{4})$.

7. (a) First of all, as Mary-Ann is doing a permutation test, she should do sampling without replacement, meaning that the first and fourth suggested code are incorrect. Secondly, her null-hypothesis is that Y is independent of X *given the value of F*. Thus, when doing the permutation, Mary-Ann should only permute within groups that have the same value for F. Thus the third suggested code is the correct one.

   (b) The p-value.

8. (a) We see that the conditional prior of β given a fixed α is a Gamma distribution:

$$\begin{aligned}\pi(\beta \mid \alpha) &\propto_\beta \pi(\alpha, \beta) \\ &\propto_\beta \beta^2 \exp(-\alpha^2\beta) \\ &\propto_\beta \frac{(\alpha^2)^3}{\Gamma(3)}\beta^2 \exp(-\alpha^2\beta).\end{aligned}$$

So

$$\beta \mid \alpha \sim \text{Gamma}(3, \alpha^2).$$

In general, the Gamma distribution is a conjugate distribution to the Gamma likelihood when α is fixed: Assume that $\beta \mid \alpha \sim \text{Gamma}(\alpha_0, \beta_0)$. Then

$$\begin{aligned}\pi(\beta \mid y, \alpha) &\propto_\beta \pi(y \mid \beta, \alpha)\pi(\beta \mid \alpha) \\ &\propto_\beta \beta^\alpha \exp(-\beta y)\beta^{\alpha_0 - 1}\exp(-\beta_0\beta) \\ &\propto_\beta \beta^{\alpha+\alpha_0-1}\exp(-(y+\beta_0)\beta).\end{aligned}$$

Thus in this case

$$\beta \mid y, \alpha \sim \text{Gamma}(\alpha_0 + \alpha, \beta_0 + y),$$

and the prior is semi-conjugate. In our particular case, $\alpha_0 = 3$ and $\beta_0 = \alpha^2$, so

$$\beta \mid y, \alpha \sim \text{Gamma}(\alpha + 3, \alpha^2 + y).$$

   (b) Taking advantage of the semi-conjugacy, we get

$$\begin{aligned}\pi(y \mid \alpha) &= \frac{\pi(y \mid \beta, \alpha)\pi(\beta \mid \alpha)}{\pi(\beta \mid y, \alpha)} \\ &= \frac{\frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1}\exp(-\beta y)\frac{(\alpha^2)^3}{\Gamma(3)}\beta^\alpha \exp(-\alpha^2\beta)}{\frac{(\alpha^2+y)^{\alpha+3}}{\Gamma(\alpha+3)}\beta^{\alpha+2}\exp(-(\alpha^2+y)\beta)} \\ &= \frac{\Gamma(\alpha+3)}{\Gamma(3)\Gamma(\alpha)}\alpha^6 \frac{y^{\alpha-1}}{(\alpha^2+y)^{\alpha+3}}.\end{aligned}$$

For the marginal prior for α, we get

$$\begin{aligned}\pi(\alpha) &= \frac{\pi(\alpha, \beta)}{\pi(\beta \mid \alpha)} \\ &\propto_\alpha \frac{\beta^2 \exp(-\alpha^2\beta)}{\frac{(\alpha^2)^3}{\Gamma(3)}\beta^2\exp(-\alpha^2\beta)} \\ &\propto_\alpha \alpha^{-6}.\end{aligned}$$

Thus the marginal posterior for α becomes

$$\begin{aligned}
\pi(\alpha \mid y) &\propto_\alpha \pi(y \mid \alpha)\pi(\alpha) \\
&\propto_\alpha \frac{\Gamma(\alpha+3)}{\Gamma(\alpha)}\alpha^6 \frac{y^{\alpha-1}}{(\alpha^2+y)^{\alpha+3}}\alpha^{-6} \\
&\propto_\alpha \frac{\Gamma(\alpha+3)}{\Gamma(\alpha)} \frac{y^{\alpha-1}}{(\alpha^2+y)^{\alpha+3}}.
\end{aligned}$$

For those who like to integrate, one may instead compute

$$\begin{aligned}
\pi(\alpha \mid y) &= \int_0^\infty \pi(\alpha,\beta \mid y)\, d\beta \\
&\propto_\alpha \int_0^\infty \pi(y \mid \alpha,\beta)\pi(\alpha,\beta)\, d\beta \\
&\propto_\alpha \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp(-\beta y)\beta^2 \exp(-\alpha^2\beta)\, d\beta \\
&\propto_\alpha \frac{y^{\alpha-1}}{\Gamma(\alpha)} \int_0^\infty \beta^{\alpha+2} \exp(-(y+\alpha^2)\beta)\, d\beta \\
&\propto_\alpha \frac{y^{\alpha-1}}{\Gamma(\alpha)} \frac{\Gamma(\alpha+3)}{(y+\alpha^2)^{\alpha+3}}
\end{aligned}$$