STA261 (Summer 2024) - Assignment 6

These problems are meant to test your understanding of the concepts in Module 6. They are *not* to be handed in. Some of these have been modified (or in some cases taken directly) from questions in the *Additional Resources* listed in the course syllabus, and no claims of originality are made.

---

1. Prove the following:
   (a) Suppose $A \subseteq \Theta$. If $\Pi(A) = 0$, then $\Pi(A \mid \mathbf{x}) = 0$.
   (b) If $\pi(\theta)$ is proper, then so is $\pi(\theta \mid \mathbf{x})$. The converse does not hold.
   (c) $\pi(\theta)$ is improper if and only if $c \cdot \pi(\theta)$ is improper for any $c > 0$.
   (d) When using an improper prior $\pi(\theta)$, the posterior under $\pi(\theta)$ is proper if and only if the posterior under $c \cdot \pi(\theta)$ is proper for any $c > 0$, and then the posteriors are identical.
2. Prove Theorem 6.1.
3. Prove Theorem 6.2, and determine the updated parameters of the posterior.
4. For the posterior mean estimator in Example 6.16, calculate its MSE (in the freqentist sense) and show that the MSE constant if we choose the hyperparameters $\alpha = \beta = \sqrt{n/4}$.
5. Suppose that $\mathbf{x}_{1:n} := (x_1, \ldots, x_n)$ is a sample from $\{f_\theta : \theta \in \Theta\}$ and that we have a prior $\pi(\theta)$ on θ. If we also have an additional sample $\mathbf{x}_{n+1:n+m} := (x_{n+1}, \ldots, x_{n+m})$, show that using the posterior $\pi(\theta \mid \mathbf{x}_{1:n})$ as a prior and then conditioning on $\mathbf{x}_{n+1:n+m}$ is the same as the posterior obtained by using the prior $\pi(\theta)$ and conditioning on $\mathbf{x}_{1:n+m} = (x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+m})$.
6. Suppose that π_1 and π_2 are pmfs on Θ. Show that when we choose a finite mixture prior of the form $\pi(\theta) = \alpha\pi_1(\theta) + (1-\alpha)\pi_2(\theta)$ for some $\alpha \in (0,1)$, then the prior predictive distribution is given by $f(\mathbf{x}) = \alpha f_1(\mathbf{x}) + (1-\alpha) f_2(\mathbf{x})$, and the posterior distribution is also a mixture of the form
$$\pi(\theta \mid \mathbf{x}) = \alpha' \pi_1(\theta \mid \mathbf{x}) + (1-\alpha')\pi_2(\theta \mid \mathbf{x})$$
for some constant $\alpha' \in (0,1)$.
7. Suppose $\mathcal{X} = \{1,2\}$ and $\Theta = \{1,2,3\}$. Three pmfs on $\mathcal{X}$ – one for each value of $\theta \in \Theta$ – are specified in the following table:

	$x = 1$	$x = 2$
$f_1(x)$	1/2	1/2
$f_2(x)$	1/3	2/3
$f_3(x)$	3/4	1/4

Suppose we use the prior $\pi(\theta)$ specified in the following table:

	$\theta = 1$	$\theta = 2$	$\theta = 3$
$\pi(\theta)$	1/5	2/5	2/5

   (a) Determine the posterior distribution of θ for each possible sample of size 2.

(b) Suppose we want to estimate θ based on having observed $x = 1$. Determine the MAP and the posterior mean estimate. Which would you prefer in this situation? Explain why.

(c) Determine a 0.8 HPD region for θ having observed $x = 1$.[1]

(d) Suppose instead interest was in $\tau(\theta) = \mathbb{1}_{\theta \in \{1,2\}}$. Identify the prior distribution of τ. Identify the posterior distribution of τ based on having observed $x = 1$. Determine a 0.5 HPD region for τ.

8. Show that the Gamma (α, β) family is conjugate for the Pareto(θ) family, which features pdfs

$$f_\theta(x) = \theta x_0^\theta \cdot x^{-\theta-1}, \quad x \geq x_0, \quad \theta > 0.$$

9. Show that the family

$$\left\{ \pi_{\alpha,\beta}(\theta) = \frac{\theta^{-\alpha}}{(\alpha-1)\beta^{\alpha-1}} \mathbb{1}_{\theta \geq \beta} \mid \alpha > 1, \beta > 0 \right\}$$

is conjugate for the Unif $(0, \theta)$ family.

10. Let $\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n \overset{iid}{\sim}$ Multinomial$(N, k; \theta_1, \theta_2, \ldots, \theta_k)$, where $\theta_1, \ldots, \theta_k > 0$ and $\sum_i \theta_i = 1$ and N and k are known integers. That is, each vector $\vec{X}_i = (X_{i,1}, \ldots, X_{i,k})$ has pmf

$$f_\theta(\vec{x}) = \frac{N!}{x_1! \cdots x_k!} \cdot \prod_{j=1}^{k} \theta_j^{x_j}.$$

(a) Show that the Dir$(\alpha_1, \ldots, \alpha_k)$ family is conjugate for the model.[2]

(b) What's the interpretation of the Dirichlet prior when $\alpha_1 = \cdots = \alpha_k = 1$?

(c) Show that if $(Y_1, \ldots, Y_k) \sim$ Dir$(\alpha_1, \ldots, \alpha_k)$, then $Y_j \sim$ Beta $\left(\alpha_j, \sum_{i=1}^k \alpha_i - \alpha_j\right)$.[3] Use this result to write down the marginal posterior $\pi(\theta_1 \mid \mathbf{x})$.

(d) Assuming each $\alpha_1, \ldots, \alpha_k \geq 1$, determine the MAP estimator and the posterior mean estimator of θ_1.

11. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ where μ is known. If we place a Gamma (α, β) prior on $1/\sigma^2$, determine the posterior distribution $\pi(\sigma^2 \mid \mathbf{x})$, as well as the MAP and posterior mean estimators for σ^2.

12. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ where μ is known. Determine Jeffreys' prior for σ^2.

13. Suppose that a manufacturer wants to construct a 0.95-credible interval for the mean lifetime θ of an item sold by the company. A consulting engineer is 99% certain that the mean lifetime is less than 50 months. If we put an Exp (λ) prior on θ, determine λ based on this information.

---

[1] In lecture, we only defined an HPD *interval*, but the concept extends to more general regions when Θ is not necessarily an interval. The way to find an HPD region is to find the "smallest possible" subset $A \subseteq \Theta$ such that $\Pi(A \mid \mathbf{x}) \geq 1 - \alpha$ and $\pi(\theta \mid \mathbf{x}) \geq \pi(\theta' \mid \mathbf{x})$ for all $\theta \in A$ and $\theta' \in A^c$. When Θ is finite, "smallest" simply means "the fewest number of elements".

[2] If you haven't seen the Dirichlet distribution before, it's a continuous distribution supported on $\mathcal{X} = \{\mathbf{x} \in (0,1)^k : \sum_{i=1}^k x_i = 1\}$ (which is called a *standard simplex*) with density given by

$$f_{\boldsymbol{\alpha}}(\mathbf{x}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_k)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_k)} \prod_{i=1}^{k} x_i^{\alpha_i - 1}, \qquad \alpha_1, \ldots, \alpha_k > 0.$$

[3] This one is pretty tricky. First show that $(Y_1, \ldots, Y_{k-2})$ has a Dir$(\alpha_1, \ldots, \alpha_{k-2}, \alpha_{k-1} + \alpha_k)$ distribution using a u-substitution with $u = \frac{y_{k-1}}{1 - y_1 - \cdots - y_{k-2}}$ and use induction to get down to the distribution of Y_1, and view that as a certain beta distribution. Then observe that nothing changes in what you just did if you swap around the indices of $(Y_1, \ldots, Y_k)$ so that the first one is j.

14. Show that the parameters of a normal distribution are completely determined as soon as we specify two quantiles of the distribution. Why is this useful for prior elicitation?

15. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ where σ^2 is known. Suppose we place a $\mathcal{N}\left(\theta, \sigma_0^2\right)$ prior on μ. Let $\tau = \sigma/\mu$ be the coefficient of variation.

    (a) Determine the posterior $\pi(\tau \mid \mathbf{x})$.

    (b) Show that the posterior mean estimator for τ does not exist.

    (c) Determine the MAP estimator for τ.

16. Let $A \subseteq \Theta$. What's the relationship between the Bayes factor in favour of A and the Bayes factor in favour of A^c?

17. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Unif}\,(0, \theta)$, where $\theta > 0$. If we place an improper prior on θ of the form $\pi(\theta) \propto 1$, determine the form of a $(1-\alpha)$ HPD interval for θ.

18. In the setup of Example 6.8, determine the Bayes factor for testing $H_0 : \mu = \mu_0$ versus $H_A : \mu \neq \mu_0$. I get

$$BF_{H_0} = \sqrt{1 + \frac{n\tau^2}{\sigma^2}} \cdot \exp\left(-\frac{n(\bar{x} - \mu_0)^2}{2\sigma^2} - \frac{\left(\frac{\mu}{\tau^2} + \frac{n\bar{x}}{\sigma^2}\right)^2}{2\left(\frac{1}{\tau^2} + \frac{n}{\sigma^2}\right)} + \frac{\theta^2}{2\tau^2} + \frac{n\bar{x}^2}{2\sigma^2} \right).$$

19. Recall the *simple linear regression* setup from previous assignments. To keep things (relatively) clean, let's assume we're doing regression through the origin as in Example 2.17, so that there are two unknown parameters in the model: β and σ^2. It's standard practice to use a Bayesian hierarchical model in this situation, where we treat σ^2 as a hyperparameter for β and give it its own hyperprior.

    (a) Justify the expression $\pi(\beta, \sigma^2) = \pi(\sigma^2) \cdot \pi(\beta \mid \sigma^2)$.

    (b) Explain why $\pi(\beta, \sigma^2 \mid \mathbf{y}) \propto \pi(\sigma^2) \cdot \pi(\beta \mid \sigma^2) \cdot f_{\beta,\sigma^2}(\mathbf{y})$.

    (c) If we choose a $\mathcal{N}\left(\theta, \sigma^2\right)$ prior for β and a $\text{Gamma}\,(a, b)$ hyperprior for $1/\sigma^2$, determine the posterior distribution $\pi(\beta, \sigma^2 \mid \mathbf{y})$. It should be another hierarchical model which is "conjugate" to the original one.

20. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ with both parameters unknown. Suppose we choose a $\mathcal{N}\left(\theta, \sigma^2\right)$ prior for μ and a $\text{Gamma}\,(a, b)$ hyperprior for $1/\sigma^2$.

    (a) Determine the posterior distribution $\pi(\mu, \sigma^2 \mid \mathbf{x})$. (*Hint*: you've already done much – but not all – of the work in a previous question...).

    (b) Derive the conditional posterior $\pi(\mu \mid \sigma^2, \mathbf{x})$.

    (c) Derive the conditional posterior $\pi(\sigma^2 \mid \mathbf{x})$.

    (d) The naïve way to determine the conditional posterior $\pi(\mu \mid \mathbf{x})$ is to marginalize σ^2 out of $\pi(\mu \mid \sigma^2, \mathbf{x})$ by working out the integral $\int_0^\infty \pi(\mu \mid \sigma^2, \mathbf{x})\, d\sigma^2$, which is absolutely horrible (as if the computations aren't long enough already). There's a clever way to avoid that here, however.

        i. Use your $\pi(\mu \mid \sigma^2, \mathbf{x})$ to find a function of μ and σ^2 which follows a $\mathcal{N}\,(0, 1)$ distribution.

        ii. Use your $\pi(\sigma^2 \mid \mathbf{x})$ to find a function of σ^2 which follows a $\chi^2_{(2a+n)}$ distribution. (*Hint*: if $Y \sim \text{Gamma}\,(\alpha, \beta)$, then $2\beta Y \sim \chi^2_{(2\alpha)}$).

        iii. Argue that the two functions you've found are independent, and find a function of them that follows a t_{2a+n} distribution using Theorem 3.4.

        iv. Finally, show that conditional on $\mathbf{x}$, we have $\mu \overset{d}{=} c + d \cdot T$ for some $c \in \mathbb{R}$ and $d > 0$, where $T \sim t_{2a+n}$. This implies that $\pi(\mu \mid \mathbf{x})$ is a *non-central t-distribution*.