MATH 559: Bayesian Theory and Methods

December 13th, 2023
9.00am – 12.00pm.

Solutions

This page intentionally left blank.

1. Suppose $\{Y_n\}$ is an infinitely exchangeable sequence. For $n \geq 1$, suppose $Y_1, \ldots, Y_n$ is a finite number of elements from this sequence. Then if $f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n)$ is the *joint pdf* for $Y_1, \ldots, Y_n$, it can be deduced that

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = \int \prod_{i=1}^{n} f(y_i; \theta)\, \pi_0(\theta)\, d\theta. \tag{1}$$

where $\pi_0(\theta)$ is a *prior density* for the unknown finite-dimensional parameter θ, and $f(y; \theta)$ is a conditional pdf in y.

(a) Using (1), derive an expression for the *posterior predictive distribution* for Y_{n+1} (another element of the infinitely exchangeable sequence) conditional on $Y_1 = y_1, \ldots, Y_n = y_n$, and explain how this defines the *posterior density*, $\pi_n(\theta)$, as an updated version of $\pi_0(\theta)$.

6 MARKS

Q1(a): Denote the left hand side of (1) by $p_0(\mathbf{y})$. Then the posterior predictive is

$$p_n(y_{n+1}) = \frac{p_0(\mathbf{y}, y_{n+1})}{p_0(\mathbf{y})} = \int f_Y(y_{n+1}; \theta) \pi_n(\theta)\, d\theta$$

as

$$p_0(\mathbf{y}, y_{n+1}) = \int \prod_{i=1}^{n+1} f(y_i; \theta)\, \pi_0(\theta) d\theta = \int f_Y(y_{n+1}; \theta) \left\{ \prod_{i=1}^{n} f(y_i; \theta)\, \pi_0(\theta) \right\} d\theta$$

where

$$\pi_n(\theta) = \frac{1}{p_0(\mathbf{y})} \prod_{i=1}^{n} f(y_i; \theta)\, \pi_0(\theta)$$

is the posterior. We note that this has the same form as (1), but with $\pi_0(\theta)$ updated to $\pi_n(\theta)$.

(b) Suppose that in (1), it is specified that

$$f(y;\theta) = \mathbb{1}_{(0,\infty)}(y)\ \theta \exp\{-\theta y\} \quad y \in \mathbb{R}$$

where $\theta \in \Theta \equiv \mathbb{R}^+$. Suppose that data $y_1, \ldots, y_n$ are observed.

(i) Find the form of the posterior distribution if a *conjugate prior* is specified. 5 MARKS

Q1(b)(i): Likelihood: if $s_n = \sum_{i=1}^n y_i$

$$\mathcal{L}_n(\theta) = \theta^n \exp\{-\theta s_n\} \qquad \theta > 0$$

so a conjugate prior is $\pi_0(\theta) = Gamma(a_0, b_0)$, and then

$$\pi_n(\theta) \equiv Gamma(a_n, b_n)$$

where $a_n = a_0 + n$, $b_n = b_0 + s_n$.

(ii) Find the *Jeffreys prior* for this model. 4 MARKS

Q1(b)(ii): We have that

$$\ell(y;\theta) = \log f_Y(y;\theta = \log\theta - \theta y$$
$$\dot{\ell}(y;\theta) = \theta^{-1} - y$$
$$\ddot{\ell}(y;\theta) = -\theta^{-2}$$

so the Jeffreys prior is

$$\pi_0(\theta) = \frac{1}{\theta}$$

as is always the case for scale models.

(iii) Find the posterior predictive density for Y_{n+1} for this model under a conjugate prior.

5 MARKS

Q1(b)(iii): We compute

$$\begin{aligned}
p_n(y_{n+1}) &= \int_0^\infty f_Y(y_{n+1};\theta)\pi_n(\theta)\,d\theta \\
&= \int_0^\infty \mathbb{1}_{(0,\infty)}(y_{n+1})\,\theta\exp\{-\theta y_{n+1}\}\frac{b_n^{a_n}}{\Gamma(a_n)}\theta^{a_n-1}\exp\{-b_n\theta\}\,d\theta \\
&= \mathbb{1}_{(0,\infty)}(y_{n+1})\frac{b_n^{a_n}}{\Gamma(a_n)}\int_0^\infty \theta^{(a_n+1)-1}\exp\{-(b_n+y_{n+1})\theta\}\,d\theta \\
&= \mathbb{1}_{(0,\infty)}(y_{n+1})\frac{b_n^{a_n}}{\Gamma(a_n)}\frac{\Gamma(a_n+1)}{(b_n+y_{n+1})^{a_n+1}} \\
&= \mathbb{1}_{(0,\infty)}(y_{n+1})a_n\left(\frac{b_n}{b_n+y_{n+1}}\right)^{a_n}\left(\frac{1}{b_n+y_{n+1}}\right)
\end{aligned}$$

Hence

$$p_n(y_{n+1}) = \mathbb{1}_{(0,\infty)}(y_{n+1})(a_0+n)\left(\frac{b_0+s_n}{b_0+s_{n+1}}\right)^{a_0+n}\left(\frac{1}{b_0+s_{n+1}}\right)$$

where

$$s_{n+1} = \sum_{i=1}^{n+1} y_i.$$

2. Suppose that $Y_1, \ldots, Y_n$ are presumed conditionally independent and distributed as $Uniform(0, \theta)$ for some $\theta \in \Theta \equiv \mathbb{R}^+$, that is

$$f_Y(y; \theta) = \frac{\mathbb{1}_{(0,\theta)}(y)}{\theta} \qquad y \in \mathbb{R}$$

Bayesian inference for θ is to be carried out.

(a) Find the form of likelihood

$$\mathcal{L}_n(\theta) = \prod_{i=1}^{n} f_Y(y_i; \theta)$$

for observed data $y_1, \ldots, y_n$. 6 MARKS

Q2(a): Evidently

$$\mathcal{L}_n(\theta) = \left\{ \prod_{i=1}^{n} \mathbb{1}_{(0,\theta)}(y_i) \right\} \left(\frac{1}{\theta} \right)^n = \mathbb{1}_{(y_{\max},\infty)}(\theta) \left(\frac{1}{\theta} \right)^n$$

where $y_{\max} = \max\{y_1, \ldots, y_n\}$.

(b) Show that the statistic

$$T \equiv T(Y_1, \ldots, Y_n) = \max\{Y_1, \ldots, Y_n\}$$

is *sufficient in the Bayesian sense*, that is, that the posterior distribution depends on the data only through the observed value of T. 4 MARKS

Q2(b): The posterior distribution is proportional to likelihood times prior, but provided that the prior places non-zero probability on the interval (t, ∞), the posterior depends on the data only through t, due to the previous answer.

(c) Let $\psi = 1/\theta$, and suppose the prior density for ψ takes the form

$$\pi_0(\psi) = c\psi \qquad 0 < \psi < 1$$

and zero otherwise, for some constant $c > 0$. Find the posterior for ψ. 6 MARKS

Leave the normalizing constant for the posterior in the form of an integral if necessary.

Q2(c): We have

$$\begin{aligned}\pi_n(\psi) &\propto \{\mathbb{1}_{(0,1/t)}(\psi)\psi^n\} \times \{\mathbb{1}_{(0,1)}(\psi)\psi\} \\ &\propto \mathbb{1}_{(0,u_n)}(\psi)\psi^{n+1}\end{aligned}$$

where

$$u_n = \min\{1/t, 1\}.$$

We have that the normalizing constant is

$$\int_0^{u_n} \psi^{n+1}\, d\psi = \frac{u_n^{n+2}}{n+2}$$

so

$$\pi_n(\psi) = \mathbb{1}_{(0,u_n)}(\psi) \left(\frac{n+2}{u_n^{n+2}}\right) \psi^{n+1}.$$

(d) For the posterior in (c), find the *posterior mode*. 4 MARKS

Q2(d): The posterior is monotonically increasing on $(0, u_n)$, so the posterior mode is u_n.

3. Suppose that, for $n \geq 1$.

$$Y_{11}, \ldots, Y_{1n} \sim Normal(\mu_1, 1)$$
$$Y_{21}, \ldots, Y_{2n} \sim Normal(\mu_2, 1)$$

are conditionally independent random variables. Observed data $y_{11}, \ldots, y_{1n}$ and $y_{21}, \ldots, y_{2n}$ are to be used for Bayesian inference.

(a) Suppose that $\pi_0(\mu_1) \equiv Normal(\eta, 1)$. Show that the posterior for μ_1 is also a Normal distribution. 6 MARKS

Q3(a): We have that the likelihood is

$$\mathcal{L}_n(\mu_1) \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(y_{1i} - \mu_1)^2\right\}$$

and the prior is

$$\pi_0(\mu_1) \propto \exp\left\{-\frac{1}{2}(\mu_1 - \eta)^2\right\}$$

so the posterior is

$$\pi_n(\mu_1) \propto \exp\left\{-\frac{1}{2}\left[n(\mu_1 - \overline{y}_{1n})^2 + (\mu_1 - \eta)^2\right]\right\}.$$

Using the complete the square formula, we have

$$\pi_n(\mu_1) \propto \exp\left\{-\frac{(n+1)}{2}(\mu_1 - m_{1n})^2\right\}$$

where $m_{1n} = (n\overline{y}_{1n} + \eta)/(n+1)$. Hence

$$\pi_n(\mu_1) \equiv Normal(m_{1n}, 1/(n+1)).$$

(b) Suppose that

$$\pi_0(\mu_1, \mu_2) = \pi_0(\mu_1)\pi_0(\mu_2)$$

where $\pi_0(\mu_1) \equiv \pi_0(\mu_2) \equiv Normal(\eta, 1)$. Find the posterior for

$$\phi = \mu_2 - \mu_1$$

based on the observed data. 8 MARKS

Q3(b): We have that

$$\pi_n(\mu_1) \equiv Normal(m_{1n}, 1/(n+1)) \qquad \pi_n(\mu_2) \equiv Normal(m_{2n}, 1/(n+1))$$

with the parameters *a posteriori* independent, so by properties of the Normal distribution

$$\pi_n(\phi) \equiv Normal(m_{2n} - m_{1n}, 2/(n+1))$$

(c) Suppose that only the data $z_1, \ldots, z_n$

$$z_i = y_{2i} - y_{1i} \qquad i = 1, \ldots, n$$

are observed. Find the posterior for ϕ as defined in (b) based on $z_1, \ldots, z_n$ under the implied prior for ϕ specified in (b). 6 MARKS

Q3(c): It is evident that under the proposed sampling model

$$Z_i \sim Normal(\phi, 2)$$

so that the likelihood is proportional to

$$\exp\left\{-\frac{1}{4}(z_i - \phi)^2\right\}$$

We have that the implied prior for ϕ is $Normal(0, 2)$. Hence the new posterior for ϕ based on the z data is given by completing the square

$$\frac{n}{2}(\phi - \overline{z}_n)^2 + \frac{1}{2}\phi^2$$

where $\overline{z}_n = \overline{y}_{2n} - \overline{y}_{1n}$. Thus

$$\pi_n(\phi) \propto \exp\left\{-\frac{(n+1)}{4}(\phi - m_n)^2\right\}$$

where

$$m_n = \frac{(n/2)\overline{z}_n}{(n+1)/2} = \frac{n\overline{z}_n}{n+1}$$

so that

$$\pi_n(\phi) \equiv Normal(m_n, 2/(n+1)).$$

as in (b).

4. Suppose that in a Bayesian model, we have that

$$f_Y(y;\theta) = \mathbb{1}_{(0,\infty)}(y)\frac{y}{\theta}\exp\left\{-\frac{y^2}{2\theta}\right\} \quad y \in \mathbb{R}$$

for $\theta \in \Theta \equiv \mathbb{R}^+$.

(a) Find the posterior, $\pi_n(\theta)$ based on a sample $y_1, \ldots, y_n$, drawn conditionally independently from $f_Y(y;\theta)$, if

$$\pi_0(\theta) \equiv InvGamma(a_0, b_0/2).$$

for hyperparameters $a_0, b_0 > 0$. 8 MARKS

Q4(a): Up to a constant, we have that the likelihood is

$$\mathcal{L}_n(\theta) \propto \left(\frac{1}{\theta}\right)^n \exp\left\{-\frac{v_n}{2}\frac{1}{\theta}\right\}$$

where

$$v_n = \sum_{i=1}^n y_i^2.$$

Therefore the posterior is

$$\pi_n(\theta) \propto \left(\frac{1}{\theta}\right)^{n+a_0-1} \exp\left\{-\frac{(v_n+b_0)}{2}\frac{1}{\theta}\right\}$$

that is

$$\pi_n(\theta) \equiv InvGamma(a_n, b_n/2)$$

where

$$a_n = a_0 + n \qquad b_n = b_0 + v_n.$$

(b) Suppose $\lambda = 1/\theta$ is to be estimated using Bayesian estimation under *quadratic loss*, that is

$$L(t, \lambda) = (t - \lambda)^2.$$

Find the Bayesian estimate of λ. 6 MARKS

Q4(b): The posterior for λ is $\Gamma(a_n, b_n/2)$, and under quadratic loss, the estimate is the posterior mean

$$\widehat{\lambda} = 2\frac{a_n}{b_n} = 2\frac{a_0 + n}{b_0 + v_n}$$

(c) Describe one procedure to derive a 95% *credible interval* for θ. 6 MARKS

Q4(c): Simply look up the 0.025 and 0.975 quantiles of the Inverse Gamma pdf.

5. (a) Describe how to carry out Monte Carlo sampling from the pdf

$$f(x) = c\mathbb{1}_{(0,\infty)}(x)\sqrt{x}\exp\left\{-\frac{1}{2}x^2\right\} \qquad x \in R$$

using the following approaches.

(i) *Rejection sampling:* give a specific recommendation for the proposal distribution $f_0(x)$. 5 MARKS

Q5(a)(i): Recommend using the $Gamma(3/2, 1/2)$ as then

$$\frac{f(x)}{f_0(x)} = c\frac{\sqrt{x}\exp\left\{-x^2/2\right\}}{\sqrt{x}\exp\left\{-x/2\right\}} = c\exp\{(x - x^2)/2\}.$$

and as $x - x^2$ is maximized when $x = 1/2$, the bound on the ratio of the unnormalized densities is

$$M = \exp\{1/8\}.$$

The algorithm proceeds as follows:

- Sample (X, U) where $X \sim f_0(x)$ and $U \sim Uniform(0, 1)$
- Accept X if

  $$U < \frac{g(x)}{Mg_0(x)}$$

  otherwise return to step 1.

Note: cannot use $f_0(x) \equiv Normal(0, 1)$ as this leaves the density ratio unbounded.

(ii) *The Metropolis-Hastings algorithm:* give a specific recommendation for the proposal density $q(x, \cdot)$.
5 MARKS

Q5(a)(ii): Here the Normal random walk is adequate: the algorithm proceeds as follows: set $x_0 = 1$, and for $t = 1$

- sample $z \sim Normal(x_{t-1}, \sigma_q^2)$ (ie $q(x, z) \equiv Normal(x, \sigma_q^2)$);
- accept z and set $x_t = z$ with probability

$$\alpha(x, z) = \min\left\{1, \frac{f(z)}{f(x)}\right\}$$

  otherwise $x_t = x_{t-1}$.

Here $f(x)$ is quite like the $Normal(0, 1)$, so choosing $\sigma_q = 1$ should be reasonable.

(b) The Bayesian analysis of the Generalized Linear Model with $Y_i \sim Poisson(\mu_i)$ and

$$\log \mu_i = \beta_0 + \beta_1 x_i$$

for outcomes $Y_1, \ldots, Y_n$ which are conditionally independent given predictor values $x_1, \ldots, x_n$ and parameters (β_0, β_1) is to be considered.

(i) Write down the joint posterior distribution $\pi_n(\beta_0, \beta_1)$ up to proportionality for a suitably chosen prior distribution. 5 MARKS

Q5(b)(i): We first need the likelihood

$$Y_i | \mathbf{X}_i = \mathbf{x}_i \sim Poisson(\mu_i)$$

where

$$\log \mu_i = \beta_0 + \beta_1 x_i = \mathbf{x}_i \beta$$

say, so that

$$\mathcal{L}_n(\beta) = \prod_{i=1}^{n} \frac{\exp\{y_i \log \mu_i - \mu_i\}}{y_i!} = \prod_{i=1}^{n} \frac{\exp\{y_i \mathbf{x}_i \beta - \exp\{\mathbf{x}_i \beta\}\}}{y_i!}$$

For the prior, we ideally need a joint prior on the whole of $\mathbb{R}^2$, so one may choose the multivariate Normal, or the special case of independent Normal priors. Then

$$\pi_n(\beta_0, \beta_1) \propto \prod_{i=1}^{n} \exp\{y_i \mathbf{x}_i \beta - \exp\{\mathbf{x}_i \beta\}\} \times \exp\left\{-\frac{1}{2}(\beta - \mathbf{m})^\top \mathbf{M}^{-1}(\beta - \mathbf{m})\right\}$$

which is not a known distribution, but is readily computed pointwise for (β_0, β_1).

(ii) Describe one method for performing Monte Carlo sampling from $\pi_n(\beta_0, \beta_1)$, giving details of each step of the approach. 5 MARKS

Q5(c)(ii): There are several options

- Metropolis-Hastings on (β_0, β_1) jointly
- Metropolis-within-Gibbs, sampling the full conditionals

$$\pi_n(\beta_0|\beta_1) \qquad \pi_n(\beta_1|\beta_0)$$

  in turn with updating
- Rejection sampling, after choosing a suitable proposal (eg bivariate Student-t)
- Sampling-importance-resampling

Answers should give step-by-step instructions, as per lecture notes (ie bookwork).