Stats 579 – Intermediate Bayesian Modeling

Assignment # 2 Solutions

1. Let $w \equiv G(y)$ with y a vector having density $f(y \mid \theta)$ and G having a differentiable inverse function. Find the density of w in general and show that the likelihoods satisfy $L(\theta \mid y) \propto L(\theta \mid w)$. (HINT: Proposition B.4 from the textbook may be useful.)

   Proposition B.4 states that the density of $w = G(y)$ is given by

   $$f_w(u) = f_y\left(G^{-1}(u)\right)\left|\det\left(dG^{-1}(u)\right)\right|,$$

   where $dG^{-1}(u)$ is the derivative (or matrix of partial derivatives) of G^{-1} evaluated at u. In the problem, then, if $y \sim f(y \mid \theta)$ and G^{-1} differentiable, then the density of w is

   $$f_w(u \mid \theta) = f_y\left(G^{-1}(u) \mid \theta\right)\left|\det\left(dG^{-1}(u)\right)\right|.$$

   Then the likelihoods are given by

   $$\begin{aligned} L(\theta \mid y) &= f_y(y \mid \theta) \\ L(\theta \mid w) &= f_w(w \mid \theta) \\ &= f_y\left(G^{-1}(w) \mid \theta\right)\left|\det\left(dG^{-1}(w)\right)\right| \\ &= f_y(y \mid \theta)\left|\det\left(dG^{-1}(w)\right)\right| \\ &\propto f_y(y \mid \theta) \end{aligned}$$

   □

2. Let $y_i \sim f(y_i \mid \theta)$ where $i \in \{1, ..., n\}$, and let $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. Prove the following statement:

$$\sum_{i=1}^{n}(y_i - \theta)^2 = \sum_{i=1}^{n}(y_i - \overline{y})^2 + n(\overline{y} - \theta)^2$$

$$\begin{aligned}
\sum_{i=1}^{n}(y_i - \theta)^2 &= \sum_{i=1}^{n}(y_i - \overline{y} + \overline{y} - \theta)^2 \\
&= \sum_{i=1}^{n}\left[(y_i - \overline{y})^2 + (\overline{y} - \theta)^2 + 2(y_i - \overline{y})(\overline{y} - \theta)\right] \\
&= \sum_{i=1}^{n}(y_i - \overline{y})^2 + \sum_{i=1}^{n}(\overline{y} - \theta)^2 + 2\sum_{i=1}^{n}(y_i - \overline{y})(\overline{y} - \theta) \\
&= \sum_{i=1}^{n}(y_i - \overline{y})^2 + (\overline{y} - \theta)^2\sum_{i=1}^{n} i + 2(\overline{y} - \theta)\sum_{i=1}^{n}(y_i - \overline{y}) \\
&= \sum_{i=1}^{n}(y_i - \overline{y})^2 + n(\overline{y} - \theta)^2 + 2(\overline{y} - \theta)\left[\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} \overline{y}\right)\right] \\
&= \sum_{i=1}^{n}(y_i - \overline{y})^2 + n(\overline{y} - \theta)^2 + 2(\overline{y} - \theta)\left[\left(\sum_{i=1}^{n} y_i\right) - n\overline{y}\right] \\
&= \sum_{i=1}^{n}(y_i - \overline{y})^2 + n(\overline{y} - \theta)^2 + 2(\overline{y} - \theta)\left[\left(\sum_{i=1}^{n} y_i\right) - n\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)\right] \\
&= \sum_{i=1}^{n}(y_i - \overline{y})^2 + n(\overline{y} - \theta)^2 + [2(\overline{y} - \theta) \times 0] \\
&= \sum_{i=1}^{n}(y_i - \overline{y})^2 + n(\overline{y} - \theta)^2
\end{aligned}$$

$\square$

3. Let $y_i|\theta \overset{\text{iid}}{\sim} \text{Exp}(\theta)$ where $i \in \{1, ..., n+1\}$, and let $p(\theta) = e^{-\theta}$.

(a) Given $y = \{y_1, ..., y_n\}$, obtain the predictive probability that $y_{n+1} > t_0$ using calculus. Argue that this probability is the posterior mean of a particular function of θ.

First, observe that by conjugacy $\theta \mid y \sim \text{Gamma}(1+n, 1+\sum y)$. Then the predictive probability that $y_{n+1} > t_0$ is given by:

$$
\begin{aligned}
&\Pr[y_{n+1} > t_0 \mid y] \\
&= \int_{t_0}^{\infty} f(y_{n+1} \mid y) dy_{n+1} \\
&= \int_{t_0}^{\infty} \int_{\Theta} f(y_{n+1} \mid \theta) p(\theta \mid y) d\theta dy_{n+1} \\
&= \int_{t_0}^{\infty} \int_{\Theta} \theta e^{-\theta y_{n+1}} \frac{(1+\sum y)^{1+n}}{\Gamma(1+n)} \theta^n e^{-\theta(1+\sum y)} d\theta dy_{n+1} \\
&= \int_{t_0}^{\infty} \left(\frac{(1+\sum y)^{1+n}}{\Gamma(1+n)} \int_{\Theta} \theta^{n+1} e^{-\theta\left(1+\sum_{i=1}^{n+1} y_i\right)} d\theta \right) dy_{n+1} \\
&= \int_{t_0}^{\infty} \left(\frac{(1+\sum y)^{n+1}}{\Gamma(n+1)} \frac{\Gamma(n+2)}{\left(1+\sum_{i=1}^{n+1} y_i\right)^{n+2}} \int_{\Theta} \frac{\left(1+\sum_{i=1}^{n+1} y_i\right)^{n+2}}{\Gamma(n+2)} \theta^{n+1} e^{-\theta\left(1+\sum_{i=1}^{n+1} y_i\right)} d\theta \right) dy_{n+1} \\
&= \int_{t_0}^{\infty} \frac{(1+\sum y)^{1+n}}{\Gamma(1+n)} \frac{\Gamma(n+2)}{\left(1+\sum_{i=1}^{n+1} y_i\right)^{n+2}} dy_{n+1} \\
&= \int_{t_0}^{\infty} \frac{(1+\sum_{i=1}^{n} y_i)^{n+1}}{\left(1+\sum_{i=1}^{n+1} y_i\right)^{n+2}} \frac{\Gamma(n+2)}{\Gamma(n+1)} dy_{n+1} \\
&= (n+1)\left(1+\sum y\right)^{n+1} \int_{t_0}^{\infty} \frac{1}{(y_{n+1} + [1+\sum y])^{n+2}} dy_{n+1} \\
&= (n+1)\left(1+\sum y\right)^{n+1} \left[-\frac{1}{(n+1)\left(y_{n+1} + [1+\sum y]\right)^{n+1}} \right]_{t_0}^{\infty} \\
&= (n+1)\left(1+\sum y\right)^{n+1} \left[0 - \left(-\frac{1}{(n+1)\left(t_0 + [1+\sum y]\right)^{n+1}} \right) \right] \\
&= \left(\frac{1+\sum y}{t_0 + 1 + \sum y} \right)^{n+1}
\end{aligned}
$$

Now let

$$g(y_{n+1} \mid \theta) = \int_{t_0}^{\infty} \theta e^{-\theta y_{n+1}} dy_{n+1},$$

and observe that

$$\begin{aligned}
\mathrm{E}_{\theta|y}[g(y_{n+1} \mid \theta)] &= \int_{\Theta} g(y_{n+1} \mid \theta) p(\theta \mid y) d\theta \\
&= \int_{\Theta} \left(\int_{t_0}^{\infty} \theta e^{-\theta y_{n+1}} dy_{n+1} \right) \frac{(1 + \sum y)^{1+n}}{\Gamma(1+n)} \theta^n e^{-\theta(1+\sum y)} d\theta
\end{aligned}$$

Then if the order of integration can be reversed, this is identical to our expression for $\Pr[y_{n+1} > t_0]$. Since $f(y_{n+1} \mid \theta)$ and $p(\theta \mid y)$ are both nonnegative and we know, from the first part, that the joint integral exists, by Fubini's theorem we therefore know that the order of integration can be reversed. Therefore, the probability of interest is equivalent to the posterior expectation of the function $g(y_{n+1} \mid \theta)$ as defined above.

(b) How would you interpret the difference between the meaning of these two quantities (the predictive probability and the posterior mean of a function of θ), despite the fact that the values are identical?

If we consider $g(y_{n+1} \mid \theta)$, we recognize that it's the probability that an observation y_{n+1} will be greater than t_0 conditional on some parameter value θ. Then the posterior expectation of $g(y_{n+1} \mid \theta)$ is the expectation of this probability under the posterior distribution for the parameter θ.

Conversely, $\Pr[y_{n+1} > t_0 \mid y]$ is a marginal probability statement: the probability that y_{n+1} will be greater than t_0 given some data y. But to leave the distinction here is to do little more than restate the question. Given that the values are identical (corresponding only to a change of integrals via Fubini's theorem), what meaningful difference is there between the posterior expectation of the probability and the (marginal) predictive probability?

The difference is in where we focus our attention and where we elide discussion of the Bayesian mechanisms at work. The posterior expectation interpretation focuses our attention on the parameter θ, which is irrelevant for the probability. It elides discussion of the observed data themselves, which factor into the creation of a posterior distribution for θ but do not appear in the function for which an expected value is desired.

The predictive interpretation focuses our attention on the data—which are still relevant to the probability, as demonstrated by the fact that they appear in the equation we obtained. This is good. It keeps us more closely tied to the objects we know and care about in Bayesian analysis. This form elides discussion of the parameter θ, which is fine since the parameter does not appear in, and is largely irrelevant to, the final calculation.

However, both forms elide discussion of the parametric model for the data, which is still very much present in the calculation. Properly, both should be thought of as conditional on a proposed model as well as their other conditioned elements. In both forms, in the end, the parameters of that model are being marginalized out—but its structure remains. We must stay alert to whether or not the model itself is a good fit to data of this type. I do not believe either method is preferable in terms of foregrounding the assumed model. I find the predictive version otherwise preferable, however, for keeping the focus on observables and away from elements which will be marginalized out and play no important part in discussion of the probability.

4. Let $y|\theta \overset{\text{iid}}{\sim} \text{Pois}(\theta)$ where $i \in \{1, ..., n+1\}$, and let $p(\theta) = e^{-\theta}$. Given $y = \{y_1, ..., y_n\}$, obtain the predictive probability that $y_{n+1} = 0$ using calculus. Argue that this probability is the posterior mean of a particular function of θ.

   First, observe that by conjugacy $\theta \mid y \sim \text{Gamma}(1 + \sum y, 1 + n)$. Then the predictive probability that $y_{n+1} > t_0$ is given by:

$$
\begin{aligned}
&\Pr[y_{n+1} = 0 \mid y] \\
&\quad = f(y_{n+1} = 0 \mid y) \\
&\quad = \int_\Theta f(y_{n+1} = 0 \mid \theta) p(\theta \mid y) d\theta \\
&\quad = \int_\Theta \frac{\theta^0 e^{-\theta}}{0!} \frac{(1+n)^{1+\sum y}}{\Gamma(1+\sum y)} \theta^{\sum y} e^{-\theta(1+n)} d\theta \\
&\quad = \frac{(1+n)^{1+\sum y}}{\Gamma(1+\sum y)} \int_\Theta \theta^{\sum y} e^{-\theta(n+2)} d\theta \\
&\quad = \frac{(1+n)^{1+\sum y}}{\Gamma(1+\sum y)} \frac{\Gamma(1+\sum y)}{(n+2)^{1+\sum y}} \int_\Theta \frac{(n+2)^{1+\sum y}}{\Gamma(1+\sum y)} \theta^{\sum y} e^{-\theta(n+2)} d\theta \\
&\quad = \frac{(1+n)^{1+\sum y}}{\Gamma(1+\sum y)} \frac{\Gamma(1+\sum y)}{(n+2)^{1+\sum y}} \\
&\quad = \frac{(n+1)^{1+\sum y}}{(n+2)^{1+\sum y}} \frac{\Gamma(1+\sum y)}{\Gamma(1+\sum y)} \\
&\quad = \left(\frac{n+1}{n+2}\right)^{1+\sum y}
\end{aligned}
$$

   Here consider $f(y_{n+1} = 0 \mid \theta)$, which we already know to be $\Pr[y_{n+1} = 0 \mid \theta]$. Observe that

$$
\mathrm{E}_{\theta|y}[f(y_{n+1} = 0 \mid \theta)] = \int_\Theta f(y_{n+1} = 0 \mid \theta) p(\theta \mid y) d\theta,
$$

   and we are immediately done. This is precisely how we calculated the marginal probability above. (NB: we don't need to worry about Fubini's theorem here because we are calculating the probability of a discrete random variable at a single point. There is only one integral to deal with; we thus do not need to worry about re-ordering.)