Stats 579 – Intermediate Bayesian Modeling

Assignment # 1 Solutions

1. Let $y_i|\theta \overset{\text{iid}}{\sim} \text{Bern}(\theta)$ for $i \in \{1, ..., n\}$, and let $p(\theta) = I_{[0,1]}(\theta)$, i.e. θ is uniform on $[0, 1]$.

   (a) Obtain the marginal density of $(y_1, ..., y_n)$.

$$
\begin{aligned}
&f(y_1, ..., y_n) \\
&\quad = \int f(y_1, ..., y_n \mid \theta) p(\theta) d\theta \\
&\quad = \int \theta^{\sum_{i=1}^n y_i} (1-\theta)^{n - \sum_{i=1}^n y_i} d\theta \\
&\quad = \frac{\Gamma\left(1 + \sum_{i=1}^n y_i\right) \Gamma\left(1 + n - \sum_{i=1}^n y_i\right)}{\Gamma(2+n)} \\
&\qquad \times \int \frac{\Gamma(2+n)}{\Gamma\left(1 + \sum_{i=1}^n y_i\right) \Gamma\left(1 + n - \sum_{i=1}^n y_i\right)} \theta^{\sum_{i=1}^n y_i} (1-\theta)^{n - \sum_{i=1}^n y_i} d\theta \\
&\quad = \frac{\Gamma\left(1 + \sum_{i=1}^n y_i\right) \Gamma\left(1 + n - \sum_{i=1}^n y_i\right)}{\Gamma(2+n)}
\end{aligned}
$$

(b) Calculate the predictive probability that $y_{n+1} = 1$ given that $y_1 = ... = y_n = 1$. Simplify the formula you get using the fact that $\Gamma(a+1) = a\Gamma(a)$ and thus establish that for $n = 1000$, $\Pr[y_{n+1} = 1] = \frac{1001}{1002}$.

$$
\begin{aligned}
&f(y_{n+1} \mid y_1, ..., y_n) \\
&= \int f(y_{n+1} \mid \theta) p(\theta \mid y_1, ..., y_n) d\theta \\
&= \int \theta^{y_{n+1}} (1-\theta)^{1-y_{n+1}} \\
&\quad \times \frac{\Gamma(2+n)}{\Gamma(1+\sum_{i=1}^n y_i)\Gamma(1+n-\sum_{i=1}^n y_i)} \theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i} d\theta \\
&= \frac{\Gamma(2+n)}{\Gamma(1+\sum_{i=1}^n y_i)\Gamma(1+n-\sum_{i=1}^n y_i)} \int \theta^{\sum_{i=1}^{n+1} y_i}(1-\theta)^{(n+1)-\sum_{i=1}^{n+1} y_i} d\theta \\
&= \frac{\Gamma(2+n)}{\Gamma(1+\sum_{i=1}^n y_i)\Gamma(1+n-\sum_{i=1}^n y_i)} \frac{\Gamma\left(1+\sum_{i=1}^{n+1} y_i\right)\Gamma\left(1+(n+1)-\sum_{i=1}^{n+1} y_i\right)}{\Gamma(2+(n+1))} \\
&\quad \times \int \frac{\Gamma(2+(n+1))}{\Gamma\left(1+\sum_{i=1}^{n+1} y_i\right)\Gamma\left(1+(n+1)-\sum_{i=1}^{n+1} y_i\right)} \theta^{\sum_{i=1}^{n+1} y_i}(1-\theta)^{(n+1)-\sum_{i=1}^{n+1} y_i} d\theta \\
&= \frac{\Gamma(2+n)}{\Gamma(1+\sum_{i=1}^n y_i)\Gamma(1+n-\sum_{i=1}^n y_i)} \frac{\Gamma\left(1+\sum_{i=1}^{n+1} y_i\right)\Gamma\left(1+(n+1)-\sum_{i=1}^{n+1} y_i\right)}{\Gamma(2+(n+1))} \\
&= \frac{\Gamma(2+n)}{\Gamma(2+(n+1))} \frac{\Gamma\left(1+\sum_{i=1}^{n+1} y_i\right)\Gamma\left(1+(n+1)-\sum_{i=1}^{n+1} y_i\right)}{\Gamma(1+\sum_{i=1}^n y_i)\Gamma(1+n-\sum_{i=1}^n y_i)} \\
&= \frac{1}{n+2} \frac{\Gamma\left(1+\sum_{i=1}^{n+1} y_i\right)}{\Gamma(1+\sum_{i=1}^n y_i)} \frac{\Gamma\left(1+(n+1)-\sum_{i=1}^{n+1} y_i\right)}{\Gamma(1+n-\sum_{i=1}^n y_i)}
\end{aligned}
$$

Observe that when $y_{n+1} = 0$, the second of these fractions is unity since $1 + \sum_{i=1}^n y_i = 1 + \sum_{i=1}^{n+1} y_i$. In this case, the third fraction reduces to

$$
\frac{\Gamma\left(1+(n+1)-\sum_{i=1}^{n+1} y_i\right)}{\Gamma(1+n-\sum_{i=1}^n y_i)} = \frac{(1+n-\sum_{i=1}^n y_i)\Gamma(1+n-\sum_{i=1}^n y_i)}{\Gamma(1+n-\sum_{i=1}^n y_i)} = \left(1+n-\sum_{i=1}^n y_i\right).
$$

This gives a predictive distribution of

$$
f(y_{n+1} = 0 \mid y_1, ..., y_n) = \frac{1+n-\sum_{i=1}^n y_i}{n+2}.
$$

By an analogous chain of arguments, we will find that

$$
f(y_{n+1} = 1 \mid y_1, ..., y_n) = \frac{1+\sum_{i=1}^n y_i}{n+2}.
$$

Applying this to the case where $y_1 = ... = y_n = 1$ with $n = 1000$ gives us

$$
f(y_{1001} = 1 \mid y_1, ..., y_{1000}) = \frac{1001}{1002}.
$$

2. Suppose in a random sample of 10 transportation workers, all were found to be on drugs. Find Box's marginal p-value and perform Bayesian significance tests to evaluate whether such data are consistent with the following models:

   (a) $y_1, ..., y_{10} \mid \theta \stackrel{\text{iid}}{\sim} \text{Bern}\,(0.10)$

   Recall that Box's marginal p-value is given by

$$p = \Pr\left[f(y) \leq f(y_{obs})\right] = \int I_{(-\infty, f(y_{obs}))}[f(y)]f(y)dy.$$

   Further, since our prior for θ is a degenerate point-mass at 0.10, we have

$$\begin{aligned} f(y) &= \int f(y \mid \theta)p(\theta)d\theta \\ &= 0.1^{\sum_{i=1}^{10} y_i} 0.9^{10-\sum_{i=1}^{10} y_i} \end{aligned}$$

   In this particular application, $y_1 = ... = y_{10} = 1$, so $f(y_{obs}) = 0.1^{10}0.9^0 = {}^1/_{10^{10}}$. Consideration of the problem should indicate that no event is less likely under this model than ten successes in ten trials, so $f(y)$ is minimized at y_{obs}. Then Box's marginal p-value for this model with these data gives $p = 0.0000000001$. Clearly, a significance test for this model indicates that these data are not consistent with the model suggested.

(b) $y_1, ..., y_{10} \mid \theta \overset{\text{iid}}{\sim} \text{Bern}(\theta); \qquad \theta \sim \text{Beta}(0.05, 0.45)$

Here, our prior leads to a more interesting marginal distribution for y.

$$
\begin{aligned}
f(y) &= \int f(y \mid \theta) p(\theta) d\theta \\
&= \int \theta^{\sum_{i=1}^{10} y_i} (1-\theta)^{10-\sum_{i=1}^{10} y_i} \, \frac{\Gamma(0.5)}{\Gamma(0.05)\Gamma(0.45)} \, \theta^{-0.95} (1-\theta)^{-0.55} d\theta \\
&= \frac{\Gamma(0.5)}{\Gamma(0.05)\Gamma(0.45)} \, \frac{\Gamma(0.05 + \sum_{i=1}^{10} y_i)\Gamma(10.45 - \sum_{i=1}^{10} y_i)}{\Gamma(10.5)} \\
&\quad \times \int \frac{\Gamma(10.5)}{\Gamma(0.05 + \sum_{i=1}^{10} y_i)\Gamma(10.45 - \sum_{i=1}^{10} y_i)} \theta^{-0.95 + \sum_{i=1}^{10} y_i} (1-\theta)^{9.45 - \sum_{i=1}^{10} y_i} d\theta \\
&= \frac{\Gamma(0.5)}{\Gamma(0.05)\Gamma(0.45)} \, \frac{\Gamma(0.05 + \sum_{i=1}^{10} y_i)\Gamma(10.45 - \sum_{i=1}^{10} y_i)}{\Gamma(10.5)}
\end{aligned}
$$

Again, successes are less likely than failures, so $f(y)$ is minimized at y_{obs}. In this case,

$$
\begin{aligned}
p &= f(y_{obs}) \\
&= \frac{\Gamma(0.5)}{\Gamma(0.05)\Gamma(0.45)} \, \frac{\Gamma(10.05)\Gamma(0.45)}{\Gamma(10.5)} \\
&= \frac{\Gamma(0.5)}{\Gamma(10.5)} \, \frac{\Gamma(10.05)}{\Gamma(0.05)} \\
&= \frac{\Gamma(0.5)}{\Gamma(0.5)\prod_{i=0}^{9}(i + 0.5)} \, \frac{\Gamma(0.05)\prod_{i=0}^{9}(i + 0.05)}{\Gamma(0.05)} \\
&= \frac{\prod_{i=0}^{9}(i + 0.05)}{\prod_{i=0}^{9}(i + 0.5)} \\
&\simeq 0.0326
\end{aligned}
$$

A significance test based on this p-value again suggests that the data aren't consistent with the model suggested (although they're much more consistent with this model than with the preceding one).

(c) $y_1, ..., y_{10} \mid \theta \overset{\text{iid}}{\sim} \text{Bern}(\theta); \qquad \theta \sim \text{Beta}(1,1)$

The form of our answer to this problem is much like the form of our answer to the last problem. Here, a slight rework of the marginal gives us

$$\begin{aligned} f(y) &= \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} \frac{\Gamma(1+\sum_{i=1}^{10} y_i)\Gamma(11-\sum_{i=1}^{10} y_i)}{\Gamma(12)} \\ &= \frac{\Gamma(1+\sum_{i=1}^{10} y_i)\Gamma(11-\sum_{i=1}^{10} y_i)}{\Gamma(12)}, \end{aligned}$$

where the final equality holds because $\Gamma(2) = 1\Gamma(1)$ and $\Gamma(1) = 1$.

Here, however, we observe that both ten successes and ten failures will be equally likely under the model (with all other outcomes being more likely). Then

$$\begin{aligned} p &= \Pr\left[f(y) \le f(y_{obs})\right] \\ f(y_{obs}) &= \frac{\Gamma(11)\Gamma(1)}{\Gamma(12)} \\ &= \frac{1}{11} \\ f(y_1 = ... = y_{10} = 0) &= \frac{\Gamma(1)\Gamma(11)}{\Gamma(12)} \\ &= \frac{1}{11} \end{aligned}$$

So $p = \frac{1}{11} + \frac{1}{11} \simeq 0.1818$. In this case, a significance test indicates that the data are somewhat consistent with this model.

3. Let $y|\theta \sim \text{Pois}(\theta)$. Conduct the following Bayesian hypothesis tests:

   (a) Assume a prior on θ of $\text{Gamma}(1,1)$. For $H_0 : \theta \leq 1$ versus $H_1 : \theta > 1$ obtain the formula for the posterior probability that H_0 is true. Use software (e.g. R, WinBUGS) to calculate the probability for $y = 3$, 5, and 7.

To begin, note that the stated prior induces a prior probability of $1 - e^1$ on H_0, and equivalently a prior probability of e^{-1} on H_1. To perform the hypothesis test, we simply find the posterior distribution for θ given the data y and calculate the posterior probability that $\theta \leq 1$.

We observe that the $\text{Pois}(\theta)$ distribution is conjugate with the $\text{Gamma}(1,1)$ distribution, yielding a $\text{Gamma}(y+1,2)$ posterior. Then in general, we can say that

$$\Pr[H_0 \mid y] = \int_0^1 \frac{2^{y+1}}{\Gamma(y+1)} \theta^y e^{-2\theta} d\theta.$$

This integral is not, in general, analytically tractable. We use the *pgamma* function in R to find the posterior probability for H_0 for the three listed values of y, obtaining

y	$\Pr[H_0 \mid y]$
3	0.1429
5	0.0166
7	0.0011

Comparing this to the prior probability for H_0, $\Pr[H_0] \simeq 0.6321$, we can see that for any of these data, we have shifted our beliefs considerably toward favoring H_1.

(b) For $H_0 : \theta = 1$ versus $H_1 : \theta \neq 1$ with $q_0 = 0.5$ and $p_1(\theta) = e^{-\theta}$, obtain the analytical formula for the posterior probability that H_0 is true. Use software (e.g. WinBUGS, R) to calculate the exact probabilities for $y = 3$, 5, and 7.

Following the textbook, we define a prior for θ as

$$p(\theta) = q_0 I_{\{1\}}(\theta) + (1 - q_0)e^{-\theta} I_{(0,\infty)}(\theta).$$

Then the posterior probability for H_0 is given by

$$\Pr[\theta = 1 \mid y] = \frac{q_0 f_0(y)}{q_0 f_0(y) + (1 - q_0) f_1(y)},$$

where

$$\begin{aligned} f_0(y) &= \int f(y \mid \theta) dp_0(\theta) \\ &= f(y \mid \theta = 1) \\ &= \frac{e^{-1}}{y!} \end{aligned}$$

and

$$\begin{aligned} f_1(y) &= \int f(y \mid \theta) dp_1(\theta) \\ &= \int \frac{\theta^y e^{-\theta}}{y!} e^{-\theta} d\theta \\ &= \frac{1}{y!} \int \theta^y e^{-2\theta} d\theta \\ &= \frac{1}{\Gamma(y+1)} \frac{\Gamma(y+1)}{2^{y+1}} \int \frac{2^{y+1}}{\Gamma(y+1)} \theta^y e^{-2\theta} d\theta \\ &= \frac{1}{2^{y+1}} \end{aligned}$$

A little calculus shows us that

$$\sum_{y=0}^{\infty} \frac{e^{-1}}{y!} = \sum_{y=0}^{\infty} \frac{1}{2^{y+1}} = 1,$$

so these are indeed probability density functions for discrete random variables. Then the posterior probability that $\theta = 1$ is relatively simple to calculate using the formula above, for any given choice of y. In particular, when $q_0 = 0.5$,

y	$\Pr[H_0 \mid y]$
3	0.4952
5	0.1640
7	0.0183

4. For $h > 0$, let $x_1, ..., x_h$ be exchangeable random quantities and let

$$y_h = \frac{1}{h} \sum_{i=1}^{h} x_i.$$

Assume that for each random quantity x_i, all moments exist—that is,

$$\mathrm{E}[x_i] = \mu_1, \quad \mathrm{E}\left[x_i^2\right] = \mu_2, \quad ....$$

Further assume that $n \leq h$ and let j_i for $i \in \{1, ..., h\}$ represent any of the $h!$ re-orderings of the indicies $1, ..., h$. Then

$$\mathrm{E}\left[\prod_{i=1}^{n} x_{j_i}\right] = m_n.$$

Prove that for all n,

$$\lim_{h \to \infty} \mathrm{E}[y_h^n] = m_n$$

By the multinomial theorem, we know that

$$y_h^n = \left(\frac{1}{h}\right)^n \sum_{k_1+...+k_h=n} \frac{n!}{k_1!...k_h!} \prod_{i=1}^{h} x_i^{k_i}.$$

Then by linearity of expectation,

$$\begin{aligned}
\mathrm{E}[y_h^n] &= \mathrm{E}\left[\left(\frac{1}{h}\right)^n \sum_{k_1+...+k_h=n} \frac{n!}{k_1!...k_h!} \prod_{i=1}^{h} x_i^{k_i}\right] \\
&= \left(\frac{1}{h}\right)^n \sum_{k_1+...+k_h=n} \frac{n!}{k_1!...k_h!} \mathrm{E}\left[\prod_{i=1}^{h} x_i^{k_i}\right].
\end{aligned}$$

Since $n \leq h$, we can consider the elements of this sum according to how many of $k_1, ..., k_h$ are equal to zero. When $h - n$ of the k_i's are equal to zero, since $\sum_{i=1}^{h} k_i = n$, it is necessarily true that the non-zero k_i's are all equal to 1. There are $\binom{h}{n} = \frac{h!}{n!(h-n)!}$ ways that the k_i's can be chosen so that n of them equal 1 and $h - n$ of them equal 0. Further,

$$\begin{aligned}
\sum_{\substack{k_1+...+k_h=n \\ k_i \in \{0,1\}}} \frac{n!}{k_1!...k_h!} \mathrm{E}\left[\prod_{i=1}^{h} x_i^{k_i}\right] &= \frac{h!}{n!(h-n)!} \frac{n!}{k_1!...k_h!} m_n \\
&= \frac{h!}{(h-n)!} m_n
\end{aligned}$$

In general, for any combination of k_i's, we recognize that $s \leq n$ of them will be positive and $n - s$ of them will be repeats among those s terms. A general expression for the number of ways we can choose these coefficients is then given by $\binom{h}{s} s^{n-s}$. This term comes about from drawing s distinct random quantities out of $\{x_1, ..., x_h\}$ and then re-drawing with replacement from these s quantities a total of $n - s$ times. Then for general s, we will have

$$\sum_{\substack{k_1+...+k_h=n \\ \sum_i I_{k_i>0}=s}} \frac{n!}{k_1!...k_h!} \mathrm{E}\left[\prod_{i=1}^{h} x_i^{k_i}\right] = \frac{h!}{s!(h-s)!} s^{n-s} \frac{n!}{k_1!...k_h!} \nu_{k_1,...,k_h},$$

where $\nu_{k_1,...,k_h}$ is the expectation associated with products of the random quantities sharing the same k_i indices, potentially reordered.

We do not know the value of $\nu_{k_1,...,k_h}$ in general, but we can inductively use the Cauchy-Schwarz inequality to prove that it is finite based on our assumptions. Consider the random quantity $t_1^{u_1} t_2^{u_2} ... t_s^{u_s}$ with $u \in \mathcal{N}$ and let $T_{s-1} = t_1^{u_1} t_2^{u_2} ... t_{s-1}^{u_{s-1}}$. Then by Cauchy-Schwarz,

$$\begin{aligned} \prod_{i=1}^{s} t_i^{u_i} &= T_{s-1} t_s^{u_s} \\ \mathrm{E}\left[\prod_{i=1}^{s} t_i^{u_i}\right] &= \mathrm{E}[T_{s-1} t_s^{u_s}] \\ &\leq \sqrt{\mathrm{E}\left[T_{s-1}^2\right] \mathrm{E}\left[t_s^{2u_s}\right]} \end{aligned}$$

If we know that all the moments of $t_1, ..., t_s$ exist, then clearly $\mathrm{E}\left[t_s^{2u_s}\right]$ exists and is one of these moments. Then if we let $s = 2$, we know that $\mathrm{E}\left[T_{s-1}^2\right] = \mathrm{E}\left[t_1^{2u_1}\right]$ and $\mathrm{E}\left[t_2^{2u_2}\right]$ both exist. This means that

$$\begin{aligned} T_2 &= \prod_{i=1}^{2} t_i^{u_i} \\ \mathrm{E}[T_2] &\leq \sqrt{\mathrm{E}\left[t_1^{2u_1}\right] \mathrm{E}\left[t_2^{2u_2}\right]} \end{aligned}$$

Then by induction, we can see that for any s and any collecton of u_i's,

$$\mathrm{E}\left[\prod_{i=1}^{s} t_i^{u_i}\right] \leq \sqrt{\prod_{i=1}^{s} \mathrm{E}\left[t_i^{2u_i}\right]}.$$

Applying this to our scenario, since the moments of the x_i's all exist, we know that $\nu_{k_1,...,k_h}$ is finite for all choices of $k_1, ..., k_h$.

Now, consider

$$\lim_{h\to\infty} \frac{\frac{h!}{s!(h-s)!} s^{n-s} \frac{n!}{k_1!...k_h!}}{h_n}.$$

When $s < n$, this limit is 0 because the numerator is of order h^s and the denominator is of order h^n. When $s = n$, however, both numerator and denominator are of the order h^n and we have, in particular,

$$\lim_{h\to\infty} \frac{h!/(h-n)!}{h^n} = 1.$$

Then we have

$$
\begin{aligned}
&\lim_{h\to\infty} \mathrm{E}[y_h^n] \\
&= \lim_{h\to\infty} \left(\frac{1}{h}\right)^n \sum_{s=1}^{n} \sum_{\substack{k_1+\ldots+k_h=n \\ \sum_i I_{k_i>0}=s}} \frac{n!}{k_1!\ldots k_h!}\ \mathrm{E}\left[\prod_{i=1}^{h} x_i^{k_i}\right] \\
&= \sum_{s=1}^{n} \lim_{h\to\infty} \left(\frac{1}{h}\right)^n \sum_{\substack{k_1+\ldots+k_h=n \\ \sum_i I_{k_i>0}=s}} \frac{n!}{k_1!\ldots k_h!}\ \nu_{k_1,\ldots,k_h} \\
&= \sum_{s=1}^{n-1} 0 + \lim_{h\to\infty} \frac{h!(h-1)!\ldots(h-n+1)!}{h^n} m_n \\
&= m_n
\end{aligned}
$$

□