Practice Problems, Exam II Solutions

STA-360/602, Spring 2018

February 19, 2019

1 Solutions

1. (15 points)

   3.14, part d. (Unit information prior).

   Similarly to how we write the likelihood up to proportionality, I will define the log-likelihood up to an additive constant that doesn't contain the parameter.

   a)

$$
\begin{aligned}
p(y_1, ..., y_n) &\propto \prod_{i=1}^{n} \theta^{y_i} e^{-\theta} \\
&= \theta^{n\,\bar{y}} e^{-n\,\theta} \\
l(\theta|y) &= n\,\bar{y} \log \theta - n\,\theta \\
\frac{d}{d\theta} l(\theta|y) &= \frac{n\,\bar{y}}{\theta} - n \\
\\
\frac{d^2}{d\theta^2} l(\theta|y) &= -\frac{n\,\bar{y}}{\theta^2} < 0
\end{aligned}
$$

   Setting the derivative of the log-likelihood equal to zero gives us that $\hat{\theta} = \bar{y}$.

$$
\begin{aligned}
J(\theta) &= -\frac{d}{d\theta}\left(\frac{n\,\bar{y}}{\theta} - n\right) \\
&= \frac{n\,\bar{y}}{\theta^2}
\end{aligned}
$$

   so $J(\hat{\theta}) = \frac{n}{\bar{y}}$.

b)

$$\begin{aligned}\log p_U(\theta) &= \frac{l(\theta|y)}{n} + c \\ &= \frac{n\,\bar{y}\log\theta - n\,\theta}{n} + c \\ &= \bar{y}\log\theta - \theta + c\end{aligned}$$

which implies that

$$p_U(\theta) \propto \theta^{\bar{y}} e^{-\theta}$$

which implies that p_U is Gamma(θ; $\bar{y}+1$, 1).
We then get that

$$\begin{aligned}-\frac{\partial^2}{\partial\theta^2}\log p_U(\theta) &= -\frac{\partial^2}{\partial\theta^2}\left(\bar{y}\log\theta - \theta + c\right) \\ &= -\frac{\partial}{\partial\theta}\left(\frac{\bar{y}}{\theta} - 1\right) \\ &= \frac{\bar{y}}{\theta^2}\end{aligned}$$

Notice that this has $\frac{1}{n}$ times the information in the likelihood. In other words, if we think of the likelihood as having n units of information (1 for each observation), then p_U has 1 unit of information. Also note that we arrived at p_U by raising the likelihood to the power $\frac{1}{n}$.

c) I'll use notation as though it is a posterior just for convenience.

$$\begin{aligned}p(\theta\,|\,y_1,...,y_n) &\propto p(y_1,...,y_n\,|\,\theta)\,p_U(\theta) \\ &\propto \theta^{n\,\bar{y}}\,e^{-n\,\theta}\,\theta^{\bar{y}}\,e^{\theta} \\ &\propto \theta^{(n+1)\,\bar{y}}\,e^{-(n+1)\,\theta}\end{aligned}$$

So $\theta\,|\,y_1,...,y_n \sim$ Gamma($(n{+}1)\,\bar{y}{+}1$, $n{+}1$). One could argue that we shouldn't call this a posterior distribution because the construction of the prior involved the observed data and thus isn't technically a prior.

2. (15 points) (Normal-Normal) Derive the posterior predictive density $p(x_{n+1}|x_{1:n})$ for the Normal–Normal model covered in lecture. Hint: There is an easy way to do this and a hard way. To make the problem easier, consider writing $X_{n+1} = \boldsymbol{\theta} + Z$ given $x_{1:n}$, where $Z \sim \mathcal{N}(0, \lambda^{-1})$.)

$$
\begin{aligned}
E(X_{n+1}|X_{1_n}, \lambda^{-1}) &= E(\theta|X_{1_n}, \lambda^{-1}) + E(Z|X_{1_n}, \lambda^{-1}) = M \\
V(X_{n+1}|X_{1_n}, \lambda^{-1}) &= V(\theta|X_{1_n}, \lambda^{-1}) + V(Z|X_{1_n}, \lambda^{-1}) = L^{-1} + \lambda^{-1} \\
X_{n+1}|X_{1_n}, \lambda^{-1} &\sim N(M, L^{-1} + \lambda^{-1})
\end{aligned}
$$

3. For labs 4 – 6, see the solutions.

4. **Approach 1 (Simple, but not great)**

   To draw a sample Z from the distribution of $X \mid X < c$,

   (a) sample $U \sim \text{Uniform}(0, 1)$,
   (b) set $X = F^{-1}(U)$,
   (c) if $X \geq c$ then return to step 1 (reject), otherwise, output $Z = X$ as a sample (accept).

   Why does it work? By the inverse c.d.f. method, we know

   $$X = F^{-1}(U) \sim \mathcal{N}(0, 1).$$

   By the rejection principle, if we reject any samples X such that

   $$X \geq c,$$

   then what remains has the conditional distribution given

   $$X < c.$$

   This approach is not ideal, however, since the rejection rate may be very high, especially when $c \ll 0$.

   Approach 2 (Better)

   To draw a sample Z from the distribution of $X \mid X < c$,

   (a) sample $U \sim \text{Uniform}(0, 1)$,
   (b) set $V = F(c)U$, and
   (c) set $Z = F^{-1}(V)$.

   Why does this work? Note that in Approach 1, rejecting when

   $$X \geq c$$

   is identical to rejecting when

   $$U \geq F(c),$$

and by the rejection principle, we know that the distribution of the U's that remain after rejection is

$$U \mid U < F(c),$$

in other words, $\text{Uniform}(0, F(c))$.

But that means that the rejection step can be by-passed completely by just sampling

$$V \sim \text{Uniform}(0, F(c))$$

and setting

$$Z = F^{-1}(V).$$

Thus, we can directly sample

$$V \sim \text{Uniform}(0, F(c)),$$

by drawing

$$U \sim \text{Uniform}(0, 1)$$

and setting

$$V = F(c)U.$$