Practice Problems, Exam II Solutions

STA-360/602, Spring 2018

February 28, 2018

1. 3.14, part d in Hoff. (Unit information prior).
2. (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 $Z \sim \mathcal{N}(0, \lambda^{-1})$.)
3. Work through section 10.3 in the Hoff book. (Metropolis).

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 $\hat{\theta} = \bar{y}$.

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

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

b)

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

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

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

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.

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

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 $Z \sim \mathcal{N}(0, \lambda^{-1})$.)

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