UC Berkeley
Department of Statistics

STAT 210A: Introduction to Mathematical Statistics

Problem Set 5- Solutions
Fall 2014

Issued: Thursday, Oct 2
Due: Thursday, Oct 9

Problem 5.1

For $\theta \sim \mathcal{U}(0,1)$, we have the following:

$$P(X_{n+1}|X_1,...,X_n) = \int_\theta P(X_{n+1},\theta|X_1,...,X_n)d\theta \propto \int_\theta P(X_{n+1}=1|\theta)P(\theta|X_1,...,X_n)d\theta.$$

Note $X_i|\theta \overset{i.i.d}{\sim} Bernoulli(\theta)$, the posterior distribution of $\theta|X_1,...,X_n$ is Beta(a,b) where $a = \sum_{i=1}^n X_i + 1, b = n - \sum_{i=1}^n X_i + 1$. Therefore:

$$P(X_{n+1}=1|X_1,...,X_n) = \int_\theta \theta \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}d\theta = \frac{a}{a+b} = \frac{\sum_{i=1}^n X_i + 1}{n+2}$$

the second last equality is because of the expected value of Beta distribution.

Problem 5.2

(1) If μ and σ^2 are independent and $\sigma^2 \sim IG(a,b),\ \mu \sim \mathcal{N}(\mu_0,\tau^2)$, then the prior and posterior distribution can be written as:

$$\begin{aligned} P(\sigma^2,\mu) &\propto (\sigma^2)^{-a-1}\exp\{-\frac{b}{\sigma^2}\}\exp\{-\frac{(u-u_0)^2}{2\tau^2}\} \\ P(\sigma^2,\mu|X) &\propto P(\sigma^2,\mu)\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}\} \\ &\propto (\sigma^2)^{-a-1-\frac{n}{2}}\exp\{-\frac{(u-u_0)^2}{2\tau^2} - \frac{\sum(x_i-\mu)^2+2b}{2\sigma^2}\}. \end{aligned}$$

The joint posterior has nested term of μ and σ^2 hence does not belong to the same family of the prior.

(2) Do a coupling between μ and σ^2, let $\sigma^2 \sim IG(a,b)$ and $\mu|\sigma^2 \sim \mathcal{N}(\mu_0, (\frac{\sigma}{\tau})^2)$, then:

$$\begin{aligned}
P(\sigma^2,\mu) &\propto (\sigma^2)^{-a-1}\exp\{-\frac{b}{\sigma^2}-\frac{\tau^2(u-u_0)^2}{2\sigma^2}\} \\
P(\sigma^2,\mu|X) &\propto P(\sigma^2,\mu)\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}\} \\
&\propto (\sigma^2)^{-a-1-\frac{n}{2}}\exp\{-\frac{b}{\sigma^2}-\frac{\tau^2(u-u_0)^2}{2\sigma^2}-\frac{\sum(x_i-\mu)^2}{2\sigma^2}\} \\
&\propto (\sigma^2)^{-a'-1}\exp\{-\frac{b'}{\sigma^2}-\frac{\tau'^2(u-u_0')^2}{2\sigma^2}\}
\end{aligned}$$

where

$$\begin{aligned}
a' &= a+\frac{n}{2} \\
u_0' &= \frac{\tau^2\mu_0+\sum x_i}{\tau^2+n} \\
b' &= b+\frac{1}{2}\left\{\sum x_i^2-\frac{(\tau^2\mu_0+\sum x_i)^2}{\tau^2+n}\right\} \\
\tau' &= \sqrt{\tau^2+n}.
\end{aligned}$$

Therefore this is a conjugate prior.

(3) In application, conjugate prior is preferred when the posterior distribution is desired for further use or analysis. For example, in online learning, we estimate the parameter iteratively, in each step, the posterior of the previous step acts as a new prior. In this case, a conjugate prior is easier to implement since we only need to update hyper-parameters instead of calculating the posterior each time. However there are cases where the independence of mean and variance is needed to interpret the model, we would prefer using non-conjugate version.

Problem 5.3

By Keener's book(4.18), if $\eta = h(\theta)$ then

$$I(\theta) = [h'(\theta)]^2 I(\eta).$$

So the transformation we need is the solution of $h'(\theta) = c\sqrt{I(\theta)}$.

- for Binomial distribution, $I(\theta) = \frac{n}{\theta(1-\theta)}$ therefore requires $h'(\theta) \propto \frac{\sqrt{n}}{\sqrt{\theta(1-\theta)}}$. Integrate on both sides yields $\eta = h(\theta) = 2C_1\sqrt{n}\arcsin(\sqrt{\theta}) + C_2$.
- for gamma distribution Gamma(a,θ), $I(\theta) = \frac{a}{\theta^2}$. It requires $h'(\theta) \propto \frac{\sqrt{a}}{\theta}$ thus: $h(\theta) = \sqrt{a}C_1 \log\theta + C_2$.

- for Maxwell distribution, $I(\theta) = \frac{3}{2\theta^2}$, thus $h(\theta) = C_1 \log\theta + C_2$.

Problem 5.4

(1) Write the posterior distribution of β:

$$
\begin{aligned}
P(\beta|y) &\propto \frac{1}{\sqrt{(2\pi)^p|g^{-1}(X'X)^{-1}|}} \exp\{-\frac{1}{2}\beta' g(X'X)\beta\} \frac{1}{\sqrt{(2\pi)^n}} \exp\{-\frac{1}{2}(y - X\beta)'(y - X\beta)\} \\
&\propto \exp\Big\{ -\frac{g+1}{2}\Big(\beta - \frac{1}{g+1}(X'X)^{-1}X'y\Big)' X'X \Big(\beta - \frac{1}{g+1}(X'X)^{-1}X'y\Big)\Big\}.
\end{aligned}
$$

Therefore it satisfies $\mathcal{N}(\frac{1}{g+1}(X'X)^{-1}X'y, (g+1)^{-1}(X'X)^{-1})$

(2) $E[\beta|y] = \frac{1}{g+1}(X'X)^{-1}X'y = \frac{1}{g+1}\hat{\beta}$.

(3) $E[\mu|y] = E[X\beta|y] = \frac{1}{g+1}X(X'X)^{-1}X'y = \frac{1}{g+1}\hat{\beta}$.

(4) $Var[\mu|y] = XVar\beta|yX' = \frac{1}{g+1}X(X'X)^{-1}X'$

(5) $\mu = X\beta$ is normal distributed, $Var[\mu|y] = \frac{1}{g+1}XX'$ under orthogonal design. $X'X = I_p$ can't ensure that XX' is zero off-diagonal then $Cov(\mu_i, \mu_k)$ is not necessary to be zero. Therefore it can't guarantee μ_j and μ_k are independent.

Problem 5.5

(1) X_i are iid Bernoulli(θ) given θ, therefore:

$$
P((X_1, ..., X_n) = (x_1, ...x_n)) = \int_\theta P((x_1, ..., x_n)|\theta)p(\theta)d\theta = \int_\theta \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}p(\theta)d\theta
$$

symmetric w.r.t $x_1, ..., x_n$ thus $(X_{\pi(1)}, ..., X_{\pi(n)}) \stackrel{d}{=} (X_1, ..., X_n)$.

(2) Using tower property:

$$
\begin{aligned}
Cov[X_1, X_2] &= E[X_1X_2] - E[X_1]E[X_2] = E[E[X_1X_2|\theta]] - E[E[X_1\theta]]E[E[X_2|\theta]] \\
&= E[\theta]^2 - (E[\theta])^2 = Var(\theta) \geq 0
\end{aligned}
$$

The equality is achieved when $Var(\theta) = 0$ which implies $\theta = \mathit{constant}$.