STAT 210A - Homework 5

Hoang Duong

October 8, 2014

Problem 1. Bernoulii

Proof. We have:

$$\begin{aligned}
\mathbb{P}\left[X_{n+1}=1 \mid X_1=k_1, X_2=k_2, ..., X_n=k_n\right] =& \frac{\mathbb{P}\left[X_{n+1}=1, X_1=k_1, X_2=k_2, ..., X_n=k_n\right]}{\mathbb{P}\left[X_1=k_1, X_2=k_2, ..., X_n=k_n\right]}\\
=& \frac{\int_0^1 \mathbb{P}\left[X_{n+1}=1, X_1=k_1, X_2=k_2, ..., X_n=k_n \mid \theta\right] d\theta}{\int_0^1 \mathbb{P}\left[X=k_1, X_2=k_2, ..., X_n=k_n \mid \theta\right] d\theta}\\
=& \frac{\int_0^1 \theta \prod_{i=1}^n \left(\theta^{k_i}(1-\theta)^{1-k_i}\right) d\theta}{\int_0^1 \prod_{i=1}^n \left(\theta^{k_i}(1-\theta)^{1-k_i}\right) d\theta}\\
=& \frac{\int_0^1 \theta^{(\sum k_i)+1}(1-\theta)^{n-\sum k_i} d\theta}{\int_0^1 \theta^{(\sum k_i)}(1-\theta)^{n-\sum k_i} d\theta}
\end{aligned}$$

Now consider the sequence of integral $a_{m,l} = \int_0^1 \theta^m(1-\theta)^l d\theta$. Notice that first $a_{m,l} = a_{l,m}$, second $a_{m,0} = a_{0,m} = \frac{1}{m+1}$. Using integration by part we have:

$$\begin{aligned}
a_{m,l} =& \int_0^1 \theta^m (1-\theta)^l d\theta\\
=& \frac{1}{2}\int_0^1 \theta^{m-1}(1-\theta)^l d\theta^2\\
=& \frac{1}{2}\theta^{m+1}(1-\theta)^t\Big|_0^1 - \frac{1}{2}\int_0^1 \theta^2 d\left(\theta^{m-1}(1-\theta)^l\right)\\
=& -\frac{1}{2}\int_0^1 (m-1)\theta^m(1-\theta)^l - l\theta^{m+1}(1-\theta)^{l-1} d\theta\\
=& -\frac{m-1}{2}a_{m,l} + \frac{l}{2}a_{m+1,l-1}\\
\Rightarrow \frac{m+1}{2}a_{m,l} =& \frac{l}{2}a_{m+1,l-1}\\
\Rightarrow a_{m,l} =& \frac{l}{m+1}a_{m+1,l-1}\\
\Rightarrow a_{m,l} =& \frac{l}{m+1}\frac{l-1}{m+2}\cdots\frac{1}{m+l}a_{m+l,0}\\
=& \frac{l!m!}{(m+l+1)!}
\end{aligned}$$

So we have the ratio:

$$\frac{a_{m+1,l}}{a_{m,l}} = \frac{l!(m+1)!}{(m+l+2)!}\frac{(m+l+1)!}{l!m!} = \frac{m+1}{m+l+2}$$

Thus our probability is:

$$\mathbb{P}\left[X_{n+1}=1 \mid X_1=k_1, X_2=k_2, ..., X_n=k_n\right]=\frac{1+\sum k_i}{n+2}$$

So the probability is quite close to the ratio of 1 from the Bernoulli trial, which is $\frac{\sum k_i}{n}$. □

Problem 2. Gaussian setting

$$\begin{aligned}
& X_i \mid \mu, \sigma^2 \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma^2) \\
\Rightarrow p(X \mid \mu, \sigma^2) \sim & \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum(x-\mu)^2\right\} \\
\prec & (\sigma^2)^{-n/2}\exp\left(-c/\sigma^2\right)
\end{aligned}$$

Proof. (a) From the class, if we fix σ^2, the conjugate prior for μ is Gaussian $\mathcal{N}(c, d^2)=\frac{1}{\sqrt{2\pi d}}\exp\left\{-\frac{1}{2d^2}(x-c)^2\right\}$. If we fix μ, the conjugate prior for σ^2 is inverse Gamma $IG(a,b) \sim \frac{b^a}{\Gamma(a)}x^{-(a-1)}\exp(-\frac{b}{x})$.

Now if we assume σ^2 and μ independent, and taking the product of the priors, which is:

$$\begin{aligned}
f_{\mu,\sigma^2}(y,z) = & \frac{1}{\sqrt{2\pi d}}\exp\left\{-\frac{1}{2d^2}(y-c)^2\right\}\frac{b^a}{\Gamma(a)}z^{-(a-1)}\exp(-\frac{b}{z}) \\
\prec & z^{-(a-1)}\exp\left\{-\frac{1}{2d^2}(y-c)^2-\frac{b}{z}\right\}
\end{aligned}$$

Thus the posterior distribution is:

$$\begin{aligned}
f_{\mu,\sigma^2|X}(y,z) \prec & z^{-n/2}\exp\left\{-\frac{1}{2z}\left[ny^2-2y\sum x+\sum x^2\right]\right\}z^{-(a-1)}\exp\left\{-\frac{1}{2d^2}(y-c)^2-\frac{b}{z}\right\} \\
= & z^{-(a-1+n/2)}\exp\left\{\frac{Ay^2+By+C}{z}+Dy^2+Ey+F\right\}
\end{aligned}$$

where A, B, C, D, E, F are constant with respect to y, z. The posterior distribution is different with the prior distribution, since in the posterior we have the term $y/z, y^2/z$, while the prior does not.

(b) A conjugate prior would be: $\sigma^2 \sim IG(a,b), \mu \mid \sigma^2 \sim \mathcal{N}(c, d^2\sigma^2)$. Indeed, the prior distribution for μ, σ^2 is:

$$\begin{aligned}
f_{\mu,\sigma^2}(y,z) = & f_{\mu|\sigma^2}(y)g_{\sigma^2}(z) \\
\prec & \frac{1}{\sqrt{2\pi z}}\exp\left\{-\frac{1}{2d^2z}(y-c)^2\right\}z^{-(a-1)}\exp\left(-\frac{b}{z}\right) \\
\prec & z^{-(a-1/2)}\exp\left\{-\frac{1}{2d^2z}(y-c)^2-\frac{b}{z}\right\}
\end{aligned}$$

The posterior distribution is then:

$$\begin{aligned}
f_{\mu,\sigma|X}(y,z) \prec & f_{X|\mu=y,\sigma^2=z}(X)f_{\mu,\sigma}(y,z) \\
\prec & z^{-n/2}\exp\left\{-\frac{1}{2z}\left(\sum_{i=1}^{n}(x_i-y)^2\right)\right\}z^{-(a-1/2)}\exp\left\{-\frac{1}{2d^2z}(y-c)^2-\frac{b}{z}\right\} \\
\prec & z^{-(a-1/2-n/2)}\exp\left\{-\frac{1}{2}(y-A)^2-\frac{B}{z}\right\}
\end{aligned}$$

For some A, B not depending on y, z. Thus the posterior distribution has the same form as the prior distribution. The prior mentioned at the beginning of (b) is a conjugate prior.

Conjugate prior is nice computationally as we aggregate more data throughout time. For example, in estimating the mean return and risk of a stock, one can keep updating the estimator daily as one gather more data. On the other hand, conjugate prior is restrictive, if we have firm belief that μ and σ^2 do not have anything to do with eachother, and we have a lot of computation power at hand, we can use the non-conjugate product prior. □

Problem 3. Constant Fisher information

Proof. Let g be the inverse of h or $\theta = g(\eta)$

(a) Binomial distribution $Bin(n, \theta)$.

The Fisher information for $\eta = h(\theta)$ is:

$$\begin{aligned}
\tilde{I}(\eta) &= I(\theta)\left[g'(\eta)\right]^2 \\
&= -\left[g'(\eta)\right]^2 \mathbb{E}_\theta \frac{\partial^2 \log p_\theta(X)}{\partial \theta^2} \\
&= -\left[g'(\eta)\right]^2 \mathbb{E}_\theta \frac{\partial^2 \log\left(\frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}\right)}{\partial \theta^2} \\
&= -\left[g'(\eta)\right]^2 \mathbb{E}_\theta \frac{\partial^2 \left(\log n! - \log k! - \log(n-k)! + k \log\theta + (n-k)\log(1-\theta)\right)}{\partial \theta^2} \\
&= -\left[g'(\eta)\right]^2 \mathbb{E}_\theta \left[\frac{\partial}{\partial \theta}\left(\frac{k}{\theta} - \frac{n-k}{1-\theta}\right)\right] \\
&= -\left[g'(\eta)\right]^2 \mathbb{E}_\theta \left[-\frac{k}{\theta^2} - \frac{n-k}{(1-\theta)^2}\right] \\
&= \left[g'(\eta)\right]^2 \left(\frac{n\theta}{\theta^2} + \frac{n - n\theta}{(1-\theta)^2}\right) \\
&= \left[g'(\eta)\right]^2 n \left(\frac{1}{\theta} + \frac{1}{1-\theta}\right) \\
&= \left[g'(\eta)\right]^2 n \frac{1}{\theta(1-\theta)}
\end{aligned}$$

We want $\tilde{I}(\eta) = a$. Thus:

$$\begin{aligned}
g'(\eta)^2 =& \frac{a}{n}\theta(1-\theta) \\
\Leftrightarrow g'(\eta)^2 =& \frac{a}{n} g(\eta)(1 - g(\eta)) \\
\Rightarrow g(\eta) =& \frac{1}{2}(\cos\left[\sqrt{a/n}(\eta + c)\right] + \frac{1}{2}
\end{aligned}$$

So we can pick a function for example $g(\eta) = \frac{1}{2}\cos\eta + \frac{1}{2} \Rightarrow 2\theta = \cos\eta + 1 \Rightarrow \eta = \arccos(2\theta - 1)$. Thus $h(\theta) = \arccos(2\theta - 1), \theta \in (0, 1)$

(b) The Fisher information for for $\eta = h(\theta)$ is:

$$
\begin{aligned}
\tilde{I}(\eta) &= -\left[g'(\eta)\right]^2 \mathbb{E}_\theta \frac{\partial^2 \log p_\theta(X)}{\partial \theta^2} \\
&= -\left[g'(\eta)\right]^2 \mathbb{E}_\theta \frac{\partial^2}{\partial \theta^2} \left\{ -\log \Gamma(a) - a \log \theta + (a-1) \log x - \frac{x}{\theta} \right\} \\
&= -\left[g'(\eta)\right]^2 \mathbb{E}_\theta \frac{\partial}{\partial \theta} \left\{ -\frac{a}{\theta} + \frac{x}{\theta^2} \right\} \\
&= -\left[g'(\eta)\right]^2 \mathbb{E}_\theta \left\{ \frac{a}{\theta^2} - \frac{2x}{\theta^3} \right\} \\
&= \left[g'(\eta)\right]^2 \frac{a}{\theta^2} = \frac{\left[g'(\eta)\right]^2}{\left[g(\eta)\right]^2} a
\end{aligned}
$$

We want $\tilde{I}(\eta) = c$ constant, we can choose $g(\eta) = \exp \frac{1}{\sqrt{a}} \eta \Rightarrow h(\theta) = \sqrt{a} \log \theta, \theta \in (0, \infty)$.

(c) The Fisher information for θ is:

$$
\begin{aligned}
I(\theta) &= -\mathbb{E}_\theta \frac{\partial^2 \left(\frac{3}{2} \log \theta + 2 \log x - \frac{\theta x^2}{2} \right)}{\partial \theta^2} \\
&= -\mathbb{E}_\theta \frac{\partial \left(\frac{3}{2\theta} - \frac{x^2}{2} \right)}{\partial \theta} \\
&= -\mathbb{E}_\theta - \frac{3}{2\theta^2} = \frac{3}{2} \theta^2
\end{aligned}
$$

Now $\tilde{I}(\eta) = \left[g'(\eta)\right]^2 I(\theta)$, and we want $\tilde{I}(\eta)$ to be constant, which mean $(g'(\eta))^2 g^2(\eta)$ constant. So we can pick $g(\eta) = \sqrt{\eta} \Rightarrow h(\theta) = \theta^2$. □

Problem 4. Linear Regression Model

Proof. 1. The posterior distribution:

$$
\begin{aligned}
p_{\beta|y} \prec & p_\beta p_{y|\beta} \\
& \prec \exp \left\{ -\frac{1}{2} \beta^T g X^T X \beta \right\} \exp \left\{ -\frac{1}{2} (y - X\beta)^T (y - X\beta) \right\} \\
& \prec \exp \left\{ -\frac{1}{2} \left[-2y^T X \beta + (1+g) \beta^T X^T X \beta \right] \right\}
\end{aligned}
$$

We want the expression inside exp to have the form $(\beta - A)^T (1+g) X^T X (\beta - A)$. This means we need: $(1+g) A^T X^T X = y^T X \Leftarrow (1+g) A^T = y^T X (X^T X)^{-1} \Leftrightarrow A^T = \frac{1}{1+g} y^T X (X^T X)^{-1}$. $\Rightarrow A = \frac{1}{1+g} (X^T X)^{-1} X^T y$. With this choice of A, we see that $\beta \mid y$ is normal with mean A, covariance matrix $\frac{1}{1+g} (X^T X)^{-1}$.

2. $\mathbb{E}(\beta \mid y) = \frac{1}{1+g} (X^T X)^{-1} X^T y = \frac{1}{1+g} \hat{\beta}$. For β is the usual MLE of β.

3. $\mathbb{E}\left[\mu \mid y\right] = \mathbb{E}\left[X\beta \mid y\right] = \frac{1}{1+g} X (X^T X)^{-1} X^T y$, which is the usual least square $\hat{y}$ multiplied with $\frac{1}{1+g}$.

4. $\mathrm{Var}\left[\mu \mid y\right] = \mathrm{Var}\left[X\beta \mid y\right] = X \left(\mathrm{Var} \beta \mid y\right) X^T = \frac{1}{1+g} X (X^T X)^{-1} X^T$.

5. First, since $\beta \mid y$ is normal, $\mu = X\beta \mid y$ is normal. So μ_i and μ_k are independent iff the covariance matrix is diagonal. But since $X^T X = I_p \Rightarrow \mathrm{Var}\left[\mu \mid y\right] = \frac{1}{1+g} X X^T$ which is not guarranteed to be diagonal. For example let $X^T = u_i^T$ for u_i is a vector norm 1 in $\mathbb{R}^n$, then $X^T X = 1$, but $X X^T$ is in general not I_n. So the answer is no. □

Problem 5. Bernoulli

Proof. 1. From question 5.1, we have the conditional distribution of $X_1, X_2, ..., X_n$ on θ is:

$$p_{X_1,...,X_n}(x_1, ..., x_n) = \int \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\mu(d\theta)$$

Since this distribution only depends on $\sum x_i$, and reordering X_i does not change the sum of them. We have that a permutation of $X_i's$ having the same distribution as $X_i's$.

2. We have:

$$\begin{aligned}
\mathrm{Cov}\left(X_i, X_j\right) =&\mathbb{E}\left[X_i X_j\right] - \mathbb{E}\left[X_i\right]\mathbb{E}\left[X_j\right] \\
=&\mathbb{E}\mathbb{E}\left[X_i X_j \mid \theta\right] - \mathbb{E}^2\left[X_i\right] \\
=&\mathbb{E}\left[\mathbb{E}\left[X_i \mid \theta\right]\mathbb{E}\left[X_j \mid \theta\right]\right] - \mathbb{E}^2\mathbb{E}\left[X_i \mid \theta\right] \\
=&\mathbb{E}\theta^2 - \mathbb{E}^2\theta = \mathrm{Var}\theta \geq 0.
\end{aligned}$$

This covariance is zero iff $\mathrm{Var}\theta = 0$. $\square$