# Problem Set 10

Hoang Nguyen, Huy Nguyen

February 7, 2020

**Problem 1.**

1. We have this result from problem 2 in Problem Set 3:

$$l(\theta) = \sum_{i=1}^{n} \Big( -\frac{1}{2}log(\theta) - \frac{1}{2\theta}X_i^2 \Big)$$

Taking deravite ans set to equal 0, we get:

$$\frac{\partial l(\theta)}{\partial \theta} = \sum_{i=1}^{n}(-\frac{1}{2\theta} + \frac{1}{2\theta^2}X_i^2) = 0 \Longleftrightarrow \theta^{MLE} = \frac{\sum_{i=1}^{n} X_i^2}{n} \tag{1}$$

2. From (1), we have the second deraviate of log-likelihood:

$$\frac{\partial l(\theta)}{\partial \theta^2} = \sum_{i=1}^{n} \Big( \frac{1}{2\theta^2} - \frac{1}{2\theta^3}X_i^2 \Big) \Leftrightarrow I(\theta) = -\sum_{i=1}^{n} \mathbb{E}[\frac{1}{2\theta^2} - \frac{1}{2\theta^3}X_i^2] = \frac{n}{2\theta^2}$$

Hence,

$$(\theta^{MLE} - \theta) \xrightarrow[n\to\infty]{\text{(d)}} N(0, \frac{2\theta^2}{n})$$

3.
a) From previous question, we know: $I(\theta) = \frac{n}{2\theta^2}$. Hence, $\pi(\theta) = c\sqrt{detI(\theta)} = c\sqrt{\frac{n}{2\theta^2}}$. This is improper prior
b) We have

$$\pi(\theta|X_1,..,X_n) = \frac{\pi(\theta)p_n(X_1,..,X_n|\theta))}{\int_0^{\infty} p_n(X_1,..,X_n|t)d\pi(t)} = \frac{c\sqrt{\frac{n}{2\theta^2}}\frac{1}{(2\pi\theta)^{\frac{n}{2}}}e^{-\frac{1}{2\theta}\sum_i^n X_i^2}}{\int_0^{\infty} p_n(X_1,..,X_n|t)d\pi(t)} = \frac{c\sqrt{n}}{(2\pi\theta)^{\frac{n}{2}}\sqrt{2\theta^2}}e^{-\frac{1}{2\theta}\sum_i^n X_i^2}$$

Hence, we can rewrite the posterio:

$$\frac{c\sqrt{n}}{(2\pi\theta)^{\frac{n}{2}}\sqrt{2\theta^2}}e^{-\frac{1}{2\theta}\sum_i^n X_i^2} = \frac{c\sqrt{n}}{(2\pi)^{\frac{n}{2}}\sqrt{2}}\frac{e^{-\frac{\frac{1}{2}\sum_{i=1}^n}{\theta}}}{\theta^{\frac{n}{2}+1}} = \frac{c\sqrt{n}}{(2\pi)^{\frac{n}{2}}\sqrt{2}}\frac{e^{-\frac{\beta}{\theta}}}{\theta^{\alpha+1}}$$

Where $\alpha = \frac{n}{2}$, $\beta = \frac{\sum_{i=1}^n X_i^2}{2}$ and $c$ is the constant satisfied $\frac{c\sqrt{n}}{(2\pi)^{\frac{n}{2}}\sqrt{2}} = \frac{\beta^{\alpha}}{\tau(\alpha)}$. This posterior is Gamma distribution with parameters $\alpha$ and $\beta$. c)
We know $\theta^{\hat{(\pi)}} = \int_0^{\infty} \theta d\pi(\theta|X_1,..,X_n)$ is the expectation of Gamma distribution with parameters $\alpha$ and $\beta$. Hence, $\theta^{\hat{(\pi)}} = \frac{\beta}{\alpha-1} = \frac{\sum_{i=1}^n X_i^2}{n-2}$.

**Problem 2.**

1. Conditionally on $\beta$, we have $Y - X\beta \sim N_p(0, \sigma^2 I_n)$. Hence, $Y \sim N_p(X\beta, \sigma^2 I_n)$.

2.
a) We know:

$$\pi(\beta|X_1,..,X_n) \propto \pi(\beta)p_n(X_1,..,X_n|\beta)) \propto e^{\frac{1}{\sigma^2}||Y-X\beta||_2^2} e^{\frac{1}{\tau^2}||\beta||_2^2} = e^{\frac{1}{\sigma^2}\Big(||Y-X\beta||_2^2 + \frac{\sigma^2}{\tau^2}||\beta||_2^2\Big)}$$

b) Let $\pi(\beta|X_1,..,X_n) = Ke^{\frac{1}{\sigma^2}\Big(||Y-X\beta||_2^2 + \frac{\sigma^2}{\tau^2}||\beta||_2^2\Big)}$
We know:

$$\frac{1}{\sigma^2}\Big(||Y-X\beta||_2^2 + \frac{\sigma^2}{\tau^2}||\beta||_2^2\Big) = \frac{1}{\sigma^2}\Big((Y-X\beta)^T(Y-X\beta) + \frac{\sigma^2}{\tau^2}\beta^T\beta\Big)$$

$$= \frac{1}{\sigma^2}\Big(Y^TY - 2\beta^TX^TY + \beta^TX^TX\beta + \frac{\sigma^2}{\tau^2}\beta^T\beta\Big) = \frac{1}{\sigma^2}\Big(Y^TY - 2\beta^TX^TY\Big) + \beta^T(\frac{1}{\sigma^2}X^TX + \frac{1}{\tau^2}I)\beta$$

Hence,

$$\pi(\beta|X_1,..,X_n) = Ke^{\frac{1}{\sigma^2}\Big(Y^TY-2\beta^TX^TY\Big)+\beta^T(\frac{1}{\sigma^2}X^TX+\frac{1}{\tau^2}I)\beta} = K_1e^{\frac{-2}{\sigma^2}\beta^TX^TY+\beta^T(\frac{1}{\sigma^2}X^TX+\frac{1}{\tau^2}I)\beta}$$

$$= K_1e^{\beta^T(\frac{1}{\sigma^2}X^TX+\frac{1}{\tau^2}I)\beta-\frac{2}{\sigma^2}\beta^TX^TY}$$

Let $\sum^{-1} = \frac{1}{\sigma^2}X^TX + \frac{1}{\tau^2}I$ and $\mu = (X^TX + \frac{\sigma^2}{\tau^2}I)^{-1}X^Ty$. We can see that $\frac{1}{\sigma^2}X^Ty = \sum^{-1}\mu$. Thereforce, we can rewrite

$$\pi(\beta|X_1,..,X_n) = K_1e^{\beta^T\sum^{-1}\beta-2\beta\sum^{-1}\mu} = K_1e^{\beta^T\sum^{-1}\beta-2\beta\sum^{-1}\mu} = K_1e^{\beta^T\sum^{-1}\beta-2\beta\sum^{-1}\mu+\mu^T\sum^{-1}\mu-\mu^T\sum^{-1}\mu}$$

We see that $\mu^T\sum^{-1}\mu$ depends on $X$ and $Y$ which are constants, hence we can get rid this constant by rewrite:

$$\pi(\beta|X_1,..,X_n) = K_2e^{\beta^T\sum^{-1}\beta-2\beta\sum^{-1}\mu+\mu^T\sum^{-1}\mu} = K_2e^{(\beta-\mu)^T\sum^{-1}(\beta-\mu)}$$

c) The posterior mean of $\beta$ is the expectation of $g(\beta)$ which is $\mu = (X^TX + \frac{\sigma^2}{\tau^2}I)^{-1}X^Ty$

3.
a) Consider function $f(t) = ||Y-Xt||^2 + \lambda||t||^2$. We take derivate and set it to be zero:

$$\frac{\partial f(t)}{\partial t} = 2X^TXt - 2X^TY + 2\lambda = 0 \Longleftrightarrow \hat{\beta} = t = (X^TX + \lambda I)^{-1}X^TY$$

b) We know the posterio mean is $\mu = (X^TX + \frac{\sigma^2}{\tau^2}I)^{-1}X^Ty$. Hence, there exists a $\tau^2$ such that $\frac{\sigma^2}{\tau^2} = \lambda \Longleftrightarrow \tau^2 = \frac{\sigma^2}{\lambda}$.
c) We know $Y \sim N_p(X\beta, \sigma^2I)$. Hence

$$\hat{\beta} = (X^TX+\lambda I)^{-1}X^TY \sim N_pBig(0, \sigma^2\Big((X^TX+\lambda I)^{-1}X^T\Big)^T\Big((X^TX+\lambda I)^{-1}X^T\Big)\Big) = N_p(0, \sigma^2(X^TX+\lambda I)^{-1})$$

d) We have $\mathbb{E}[||\hat{\beta}-\beta||_2^2] = \sum_{i=1}^D \mathbb{E}[(\hat{\beta}_i - \beta_i)^2] = \sum_{i=1}^D Var(\hat{\beta}_i) = \sigma^2 tr((X^TX + \lambda I)^{-1})$.

**Problem 3.**

1 Covariance matrix is a square matrix giving the covariance between each pair of elements of a given random vector. In the matrix diagonal there are variances, i.e., the covariance of each element with itself. The dimension is $n$.

2 The sample covariance matrix is the matrix whose elements are pairwise covariances of the vectors.

3 We have covariance matrix $C = \mathbb{E}[(X-\bar{X})(X-\bar{X})^T]$. Hence

$$u^TCu = u^T\mathbb{E}[(X-\bar{X})(X-\bar{X})^T]u = \mathbb{E}[u^T(X-\bar{X})(X-\bar{X})^Tu] = \mathbb{E}[\sigma^2] = \sigma^2 \geqslant 0$$

5

a) Let $B = AX$ and $b_i$ be the element ith of B. We have the element on i-row, j-column of covarinace maxtrix $B'$ of B is calculated:

$$B'_{ij} = Cov(B_i, B_j) = Cov(A_iX, A_jX) = \mathbb{E}[(A_iX)(A_jX)^T]) - \mathbb{E}[A_iX]\mathbb{E}[A_jX]^T = A_i\mathbb{E}[XX^T]A_j^T - A_i\mathbb{E}[X]\mathbb{E}[X]^T A_j^T$$

$$= A_i(\mathbb{E}[XX^T] - \mathbb{E}[X]\mathbb{E}[X]^T)A_j^T = A_i \sum A_j^T$$

where $A_i$ denotes whole j row of matrix A. Hence, the covariance of $AX$ is $A \sum A^T$. b)