Problem Set 10

Hoang Nguyen, Huy Nguyen

February 22, 2020

Problem 1.

1. We have PDF: $p(x) = p^x(1-p)^{1-x} = e^{xlog(p)+(1-x)log(1-p)}$. Hence we have:

$$\begin{cases} \eta_1 = log(p) \\ \eta_2 = log(1-p) \\ T_1(x) = x \\ T_2(x) = 1-x \\ h(x) = 1 \\ B(p) = 0 \end{cases}$$

2. We have PDF $p(x) = e^{\mu x - \frac{\mu^2}{2}} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}$.Hence

$$\begin{cases} \eta = \mu \\ T(x) = x \\ h(x) = \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} \\ B(p) = \frac{\mu^2}{2} \end{cases}$$

3. We have PDF $p(x) = e^{\frac{\mu}{\sigma^2} - \frac{1}{2\sigma^2}x^2 - \frac{\mu^2}{2\sigma^2}} \frac{1}{\sigma\sqrt{2\pi}}$. Hence

$$\begin{cases} \eta_1 = \frac{\mu}{\sigma^2} \\ \eta_2 = -\frac{1}{2\sigma^2} \\ T_1(x) = x \\ T_2(x) = x^2 \\ h(x) = 1 \\ B(p) = \frac{\mu^2}{2\sigma^2} + log(\sigma\sqrt{2\pi}) \end{cases}$$

4. We have $p(x) = \lambda e^{-\lambda x}$. Hence

$$\begin{cases} \eta = -\lambda \\ T(x) = x \\ h(x) = \lambda \\ B(p) = 0 \end{cases}$$

5. We have $p(x) = \frac{1}{\upsilon} = e^{-log(\upsilon)}$. Hence We have $p(x) = \lambda e^{-\lambda x}$. Hence

$$\begin{cases} \eta = -log(\upsilon) \\ T(x) = 1 \\ h(x) = 1 \\ B(p) = 0 \end{cases}$$

6. We have $p(x) = \frac{\beta^\alpha}{\mathbb{I}(\alpha)} x^{\alpha-1} e^{-\beta x} = e^{-\beta x + (\alpha-1)log(x)} \frac{\beta^\alpha}{\mathbb{I}(\alpha)}$. Hence

$$\begin{cases} \eta_1 = -\beta \\ \eta_2 = \alpha - 1 \\ T_1(x) = x \\ T_2(x) = log(x) \\ h(x) = 1 \\ B(p) = -log(\frac{\beta^\alpha}{\mathbb{I}(\alpha)}) \end{cases}$$

7. We have $p(x) = \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda + log(\lambda)x} \frac{1}{x!}$. Hence

$$\begin{cases} \eta_1 = log(\lambda) \\ T(x) = x \\ h(x) = \frac{1}{x!} \\ B(p) = -\lambda \end{cases}$$

Problem 2.

1. Conditionally on β, 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:

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

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

We know:

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

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

Hence,

$$\pi(\beta|X_1, .., X_n) = Ke^{\frac{1}{\sigma^2}\left(Y^TY - 2\beta^TX^TY\right) + \beta^T(\frac{1}{\sigma^2}X^TX + \frac{1}{\tau^2}I)\beta} = K_1 e^{\frac{-2}{\sigma^2}\beta^TX^TY + \beta^T(\frac{1}{\sigma^2}X^TX + \frac{1}{\tau^2}I)\beta}$$

$$= K_1 e^{\beta^T(\frac{1}{\sigma^2}X^TX + \frac{1}{\tau^2}I)\beta - \frac{2}{\sigma^2}\beta^TX^TY}$$

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

$$\pi(\beta|X_1, .., X_n) = K_1 e^{\beta^T\sum^{-1}\beta - 2\beta\sum^{-1}\mu} = K_1 e^{\beta^T\sum^{-1}\beta - 2\beta\sum^{-1}\mu} = K_1 e^{\beta^T\sum^{-1}\beta - 2\beta\sum^{-1}\mu + \mu^T\sum^{-1}\mu - \mu^T\sum^{-1}\mu}$$

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:

$$\pi(\beta|X_1, .., X_n) = K_2 e^{\beta^T\sum^{-1}\beta - 2\beta\sum^{-1}\mu + \mu^T\sum^{-1}\mu} = K_2 e^{(\beta-\mu)^T\sum^{-1}(\beta-\mu)}$$

c) The posterior mean of β 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 τ^2 such that $\frac{\sigma^2}{\tau^2} = \lambda \Longleftrightarrow \tau^2 = \frac{\sigma^2}{\lambda}$.
c) We know $Y \sim N_p(X\beta, \sigma^2 I)$. Hence

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

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 p by p.

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

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

4.
We know $S = \frac{1}{n}\sum_{i=1}^n X_iX_i^T - \bar{X}\bar{X}^T$. Hence:

$$u^TSu = \frac{1}{n}\sum_{i=1}^n u^TX_i(X_i^Tu) - u\bar{X}(\bar{X}^Tu) = \frac{1}{n}\sum_{i=1}^n (u^TX_i)^2 - (u^T\bar{X})^2$$

On the other hand, $u^T\bar{X} = \frac{1}{n}\sum_{i=1}^n u^TX_i = u^{\bar{T}}X$. Hence $u^TSu = \frac{1}{n}\sum_{i=1}^n u^TX_i(X_i^Tu) - u^{\bar{T}}X^2 \Rightarrow u^TSu$ is sample variance of u^X, hecne this has to be greater or equal to 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:

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

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

where A_i denotes whole j row of matrix A. Hence, the covariance of AX is $A\sum A^T$.
b) For any non-zero vector $u \in \mathbb{R}^q$, we have:

$$u^TA\sum A^Tu^T = (A^Tu)\sum(A^Tu)^T$$

If A^T has full column rank which means for any non-zero u, $A^Tu = 0$ is impossible, hence $u^TA\sum A^Tu^T = (A^Tu)\sum(A^Tu)^T > 0 \Rightarrow$ all eigenvalues of $A\sum A^T$ are positive, hence covariance of AX is invertible.
If A^T does not have full column rank, there exsists non-zero vector u such that $A^Tu = 0$. That u make $u^TA\sum A^Tu^T = (A^Tu)\sum(A^Tu)^T = 0$ which means there is at least an eigenvalue of $A\sum A^T$ that is equal to 0, hence covariance of AX is not invertible.
c) We have $\sum = \mathbb{E}[(u^TX)(X^Tu)] - \mathbb{E}[u^TX]\mathbb{E}[u^TX]^T = u^T\mathbb{E}[XX^T]u - u^T\mathbb{E}[X]\mathbb{E}[X]^Tu = u^T\sum_X u$.

6.
a) Using similar methhof from previous questions, sample covariance matrix of $BX_1, BX_2, .., BX_n$ are $B\hat{\sum}_1 B^T, B\hat{\sum}_2 B^T, ..., B\hat{\sum}_n B^T$.
b) Using similar methhof from previous questions, sample covariance matrix of $u^TX_1, u^TX_2, .., u^TX_n$ are $u^T\hat{\sum}_1 u, u^T\hat{\sum}_2 u, ..., u^T\hat{\sum}_n u$.

7.
We know $\mathbb{E}[X^T A^T AX] = \sum_{i=1}^k \mathbb{E}[(A_i X)^T (A_i X)]$
On the other hand,

$$||A\mu||_2^2 + Tr(A \sum A^T) = \sum_{i=1}^k A_i^T \mu\mu^T A_i + Cov(A_i^T X, A_i^T X) = \sum_{i=1}^k A_i^T \mu\mu^T A_i + \mathbb{E}[A_i^T XX^T A_i] - A_i^T \mu\mu^T A_i$$

$$= \sum_{i=1}^k \mathbb{E}[A_i^T XX^T A_i] = \sum_{i=1}^k \mathbb{E}[(A_i X)^T (A_i X)]$$