CMSC25025 HW1

Zihao Wang

April 10, 2018

1. (a)

$$\Pr(X < x) = \Pr(F^{-1}(U) < x) = \Pr(U < F(x)) = F(x)$$

(b) i.

$$\begin{aligned}
&\Pr(Z < z) \\
&= \Pr(X - Y < z) \\
&= \int_{x-y<z} f_{X,Y}(x,y)dxdy \\
&= \int_y e^{-\mu y} \int_{x \le y+z} e^{-\lambda x} dxdy \\
&= \int_y e^{-\mu y} + e^{-(\mu+y)} * e^{-\lambda z} dy \\
&= \frac{1}{\mu} - \frac{e^{-\lambda z}}{\mu + \lambda}
\end{aligned}$$

ii.

$$\begin{aligned}
&\Pr(Z < z) \\
&= 1 - \Pr(Z \ge z) \\
&= 1 - \Pr(X \ge z, Y \ge z) \\
&= 1 - \Pr(X \ge z) \Pr(Y \ge z) \\
&= 1 - e^{-(\mu+\lambda)}
\end{aligned}$$

(c)

$$\mathrm{E}(Y) = E(e^X) = \int_R e^x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$
$$= e^{1/2} \int_R e^x \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-1)^2}{2}} dx = e^{1/2}$$

For Var(y), first work out E(Y^2)

$$
\begin{aligned}
&\mathrm{E}(Y^2) = \\
&\int_R e^{2x}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \\
&= e^2\int_R e^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-2)^2}{2}}dx = e^2
\end{aligned}
$$

Thus By $\mathrm{Var}(Y) = E(Y^2) - E(Y)^2$, we have:

$$\mathrm{Var}(Y) = e^2 - (e^{1/2})^2 = e^2 - e$$

(d)

$$
\begin{aligned}
&\mathrm{Var}(Y) = \mathrm{E}(Y^2) - \mathrm{E}(Y)^2 \\
&= \mathrm{E}(\mathrm{E}(Y^2 \mid X)) - (\mathrm{E}(\mathrm{E}(Y \mid X)))^2 \\
&= \mathrm{E}(\mathrm{E}(Y^2 \mid X)) - \mathrm{E}((\mathrm{E}(Y \mid X))^2) + \mathrm{E}((\mathrm{E}(Y \mid X))^2) - (\mathrm{E}(\mathrm{E}(Y \mid X)))^2 \\
&= \mathrm{Var}(Y \mid X) + \mathrm{E}(\mathrm{Var}(Y \mid X))
\end{aligned}
$$

2. From the nonsingularity of X^TX we can see meaning X is a full rank matrix, thus $det(X) > 0$ and $rank(X) = d$

   (a) The problem of finding the least square estimates can be stated in an optimization problem:
   Goal: Minimize $g(\hat{y}) = (\hat{y} - y)^T(\hat{y} - y))$
   s.t: $\hat{y} \in L = \{\hat{y} = X\hat{\beta}; \beta \in R^n\}$
   Then the goal can be changed into minimize $h(\hat{\beta})$ where $h(\hat{\beta}) = g(X\hat{\beta}), \hat{\beta} \in R^n$

   $$\nabla h = 2X^TX - 2X^Ty$$

   Thus $\nabla h = 0$ when $\hat{\beta} = (X^TX)^{-1}X^Ty$ Since we have:

   $$\nabla^2 h(\hat{\beta}) = 2det(X^TX) > 0$$

   We can say the obtained is indeed the least square estimates.

   (b)
   $$HX = X(X^TX)^{-1}X^TX = X(X^TX)^{-1}(X^TX) = X$$

   (c)
   $$H^T = (X(X^TX)^{-1}X^T)^T = X(X^TX)^{-1}X^T = H$$

   (d)
   $$H^2 = (X(X^TX)^{-1}X^T)^2 = X(X^TX)^{-1}X^T = H$$

(e)

$$\hat{y} = Hy = X\hat{\beta} =$$

$$\begin{bmatrix} x_1 & x_2 & ... & x_d \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ ... \\ b_d \end{bmatrix}$$

This is a linear combination of the columnns in X, thus a projection of y into L.

(f) First, I show 1 is the only eigenvalue of H:
For any nonzero vector $\gamma \in L$, we can find $\beta \in R^n s.t \gamma = X\beta$; then from

$$H\gamma = \lambda\gamma$$

we have

$$X\beta = \lambda X\beta$$

Then it is easy to show that $X\beta \neq 0$ since otherwise:

$$(X\beta)^T(X\beta) = 0 ==> \beta^T X^T X\beta = 0$$

From the positive definite property of $X^T X$, we can get $\beta = 0$
Then it follows that $\lambda = 1$
Without 0 eigenvalue we can see H is a full rank matrix, thus $rank(H) = d$, which also shows H has d eigenvalues. Then by $tr(H) = \sum_{i=1}^{d} \lambda_i = d$. ($rank(X) = d$ has been at the front of the problem)

3. For convenience of notation, we denote the diagnol values of H to be σ_i with $\sigma_i = 0, i > r$

(a)

$$XX^T = U\sum(\sum)^T U^T$$

$$XX^T \begin{bmatrix} u_1 & u_2 & ...u_m \end{bmatrix} = U\sum(\sum)^T = \begin{bmatrix} u_1 & u_2 & ...u_m \end{bmatrix} \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_m^2 \end{bmatrix}$$

Thus the columns of U are eigenvectors of XX^T, and the eigenvalues for u_i is σ_i^2
In the same way we can show the columns of U are eigenvectors of $X^T X$, and the eigenvalues for v_i is σ_i^2

(b)

$$XV = X(v_1, v_2, ...v_n)$$

$$LHS = U\sum = \begin{bmatrix} u_1 & u_2 & ...u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_m \end{bmatrix} = \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & ...\sigma_m u_m \end{bmatrix}$$

Thus we have $Xv_i = \sigma_i u_i$
In the same way we can show $X^T u_i = \sigma_i v_i$

(c)

$$LHS = \sqrt{tr(X^T X)} = \sqrt{\sum_{i=1}^{d} \sigma_i^2}$$

(d)

$$|X|^2 = |X^T|\,|X| = |X^T X| = \prod_{i=1}^{r} \sigma^2$$

Thus

$$|det(X)| = \sqrt{\prod_{i=1}^{r} \sigma^2}$$

(e)

$$H = X(X^T X)^{-1} X^T = (U\sum V^T)(U\sum(\sum)^T V^T)^{-1}(V\sum U^T) = U\Sigma(\Sigma^T\Sigma)^{-1}\Sigma^T U^T$$

(f) In this case $(\Sigma^{(k)})^T\Sigma^{(k)}$ may not be invertible. Therefore we use the psudoinverse of it. Then the hat function becomes:

$$H^{(k)} = U\Sigma^{(k)}((\Sigma^{(k)})^T\Sigma^{(k)})^{+}(\Sigma^{(k)})^T U^T$$

And the least square estimate becomes $\hat{y} = H^{(k)} y$