STAT 611 Homework 2 Solutions

1 (a) The joint pdf of the data is

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \exp\left(-\frac{X_i - \theta}{\theta}\right) \mathbf{1}(x \geq \theta)$$
$$= \left(\frac{e}{\theta}\right)^n \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} X_i\right) \mathbf{1}(\min X_i \geq \theta)$$

Let $T(\mathbf{x}) = (\sum_{i=1}^n X_i, \min X_i) = (\sum_{i=1}^n X_i, X_{(1)})$ which by the factorization theorem is a sufficient statistic for θ. If $\mathbf{x}$ and $\mathbf{y}$ are data points, then the ratio

$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} = \frac{\left(\frac{e}{\theta}\right)^n \exp\left(-\frac{1}{\theta}\sum_{i=1}^n X_i\right) \mathbf{1}(\min X_i \geq \theta)}{\left(\frac{e}{\theta}\right)^n \exp\left(-\frac{1}{\theta}\sum_{i=1}^n Y_i\right) \mathbf{1}(\min Y_i \geq \theta)}$$
$$= \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n}(X_i - Y_i)\right) \frac{\mathbf{1}(\min X_i \geq \theta)}{\mathbf{1}(\min Y_i \geq \theta)}$$

being constant as a function of θ implies that $T(\mathbf{x}) = T(\mathbf{y})$.

(b) Note that

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = n\mathbb{E}[X_1]$$

Now,

$$\mathbb{E}[X - \theta] = \int_{\theta}^{\infty} (t - \theta)\frac{1}{\theta} e^{-\frac{t-\theta}{\theta}} dt = \theta$$

and so since $\mathbb{E}[X] = 2\theta$, we have

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = 2n\theta$$

To calculate the expectation of the first order statistic, we need to find its distribution. First note that

$$P(X \geq x) = \int_{x}^{\infty} \frac{1}{\theta} e^{-\frac{t-\theta}{\theta}} dt = e^{-\frac{x-\theta}{\theta}}$$

So

$$P(X_{(1)} \geq x) = 1 - P(X_{(1)} \geq x) = 1 - \prod_{i=1}^{n} P(X_i \geq x) = 1 - e^{-\frac{n(x-\theta)}{\theta}}$$

Differentiating, we have

$$f_{X_{(1)}}(x) = \frac{n}{\theta} e^{-\frac{n(x-\theta)}{\theta}}$$

Thus,

$$\mathbb{E}[X_{(1)} - \theta] = \int_{\theta}^{\infty} (t-\theta) \frac{n}{\theta} e^{-\frac{n(t-\theta)}{\theta}} dt = \frac{\theta}{n}$$

and

$$\mathbb{E}[X_{(1)}] = \theta \left(1 + \frac{1}{n}\right)$$

Hence,

$$\mathbb{E}\left[\frac{1}{2n} \sum_{i=1}^{n} X_i - \frac{X_{(1)}}{1 + \frac{1}{n}}\right] = \theta - \theta = 0$$

But since $\frac{1}{2n} \sum_{i=1}^{n} X_i - \frac{X_{(1)}}{1+\frac{1}{n}}$ is not a constant, T is not complete.

2 (a) The joint pdf is

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_{i<j} p_{ij}^{x_{ij}} (1 - p_{ij})^{x_{ij}} \\ &= \exp\left(\sum_{i<j} x_{ij} \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) + \sum_{i>j} \log(1 - p_{ij})\right) \end{aligned}$$

The natural parameter is

$$\eta(\theta) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right)$$

and sufficient statistic is

$$T(\mathbf{x}) = (x_{ij})_{i<j}$$

Since neither T nor η satisfy a non-trivial linear constraint, T is a minimal sufficient statistic for θ.

(b) From the hint, we have

$$\frac{p_{ij}}{1-p_{ij}} = \exp(\beta_i + \beta_j)$$

Notice that the first sum in the pdf from (a) can be written as

$$\begin{aligned}
\sum_{i<j} x_{ij} \log\left(\frac{p_{ij}}{1-p_{ij}}\right) &= \sum_{j=1}^{n}\sum_{i=1}^{j-1} x_{ij}(\beta_i + \beta_j) \\
&= \sum_{j=1}^{n}\sum_{i=1}^{j-1} x_{ij}\beta_i + \sum_{j=1}^{n}\sum_{i=1}^{j-1} x_{ij}\beta_j \\
&= \sum_{i=1}^{n}\sum_{j=i+1}^{n} x_{ji}\beta_i + \sum_{i=1}^{n}\sum_{i=1}^{j-1} x_{ji}\beta_j \\
&= \sum_{i=1}^{n}\beta_i \sum_{j=i+1}^{n} x_{ji} + \sum_{i=1}^{n}\beta_i \sum_{j=1}^{i-1} x_{ji} \\
&= \sum_{i=1}^{n}\beta_i \left[\sum_{j=i+1}^{n} x_{ji} + \sum_{j=1}^{i-1} x_{ji}\right]
\end{aligned}$$

So the joint pdf under the parameterization $\theta = (\beta_1, \ldots, \beta_n)$ is

$$f(\mathbf{x}|\theta) = \exp\left(\sum_{i=1}^{n}\beta_i\left(\sum_{j=i+1}^{n} x_{ji} + \sum_{j=1}^{i-1} x_{ji}\right) - \sum_{i\neq j}\log(1+\exp(\beta_i+\beta_j))\right)$$

Or equivalently,

$$T(\mathbf{x}) = \left(\sum_{k=1}^{n} x_{ik}\right)_{i=1}^{n}$$

as $x_{ij} = x_{ji}$ and $x_{ii} = 0$ for all i, j.

(c) The natural parameter in (b) is in $\mathbb{R}^k$ which clearly has non-empty interior. Since the above representation is minimal, this implies full rank and so T is complete.