Mathematical Statistics I
Recitation Session 3

Theorem. *Let $X_1, \ldots, X_n$ be iid observations from a pdf or pmf $f(x; \boldsymbol{\theta})$ that belongs to an exponential family given by*

$$f(x; \boldsymbol{\theta}) = h(x)c(\theta) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x)\right),$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d), d \leq k$. Then

$$T(\mathbf{X}) = \left(\sum_{j=1}^{n} t_1(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j)\right)$$

is a sufficient statistic for $\boldsymbol{\theta}$.

Definition. A sufficient statistic $T(\mathbf{X})$ is called a minimal sufficient statistic if, for any other sufficient statistic $T'(\mathbf{X}), T(\mathbf{x})$ is a function of $T'(\mathbf{x})$.

Theorem. *Let $f(\mathbf{x}; \theta)$ be the pmf or pdf of a sample $\mathbf{X}$. Suppose there exists a function $T(\mathbf{x})$ such that, for every two sample points $\mathbf{x}$ and $\mathbf{y}$, the ratio $f(\mathbf{x}; \theta)/f(\mathbf{y}; \theta)$ is constant as a function of θ if and only if $T(\mathbf{x}) = T(\mathbf{y})$. Then $T(\mathbf{X})$ is a minimal sufficient statistic for θ.*

Definition. Let $f(t; \theta)$ be a family of pdfs or pmfs for a statistic $T(\mathbf{X})$. The family of probability distributions is called complete if $\mathrm{E}_\theta(g(T)) = 0$ for all θ implies $P_\theta(g(T) = 0) = 1$ for all θ. Equivalently, $T(\mathbf{X})$ is called a complete statistic.

Theorem. *Let $X_1, \ldots, X_n$ be iid observations from an exponential family with pdf or pmf of the form*

$$f(x; \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{j=1}^{k} w(\theta_j) t_j(x)\right),$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$. Then the statistic

$$T(\mathbf{X}) = \left(\sum_{i=1}^{n} t_1(X_i), \sum_{i=1}^{n} t_2(X_i), \ldots, \sum_{i=1}^{n} t_k(X_i)\right)$$

is complete as long as the parameter space Θ contains an open set in $\Re^k$.

Definition. A statistic $S(X)$ whose distribution does not depend on the parameter θ is called an ancillary statistic.

Exercise. Let $X_1, \ldots, X_n$ be independent random variables with densities

$$f_{X_i}(x;\theta) = \begin{cases} e^{i\theta - x} & x \geq i\theta \\ 0 & x < i\theta. \end{cases}$$

Prove that $T = \min_i (X_i/i)$ is a sufficient statistic for θ.

Exercise. Let $X_1, \ldots, X_n$ be independent random variables with pdfs

$$f(x_i;\theta) = \begin{cases} \frac{1}{2i\theta} & -i(\theta-1) < x_i < i(\theta+1) \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$. Find a two-dimensional sufficient statistic for θ.

Exercise. Suppose $X_1, \ldots, X_n$ are iid uniform observations on the interval $(0,\theta), 0 < \theta < \infty$. Show that $T(\mathbf{X}) = X_{(n)}$ is a complete statistic.

Exercise. Let $X_1, \ldots, X_n$ be i.i.d. from the $N\left(\theta, \theta^2\right)$ distribution, where $\theta > 0$ is a parameter. Find a minimal sufficient statistic for θ and show whether it is complete.

Exercise. Suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d. random 2 -vectors having the normal distribution with $E(X_1) = E(Y_1) = 0, \operatorname{Var}(X_1) = \operatorname{Var}(Y_1) = 1$, and $\operatorname{Cov}(X_1, Y_1) = \theta \in (-1,1)$.

(a) Find a minimal sufficient statistic for θ.

(b) Show whether the minimal sufficient statistic in (a) is complete or not.

Exercise. For each of the following pdfs let $X_1, \ldots, X_n$ be iid sample. Find a complete sufficient statistic, or show that one does not exist.

(a) $f(x;\theta) = \frac{\theta}{(1+x)^{1+\theta}}, \quad 0 < x < \infty, \quad \theta > 0$

(b) $f(x;\theta) = \frac{(\log\theta)\theta^x}{\theta-1}, \quad 0 < x < 1, \quad \theta > 1$

Exercise. Let $X_1, \ldots, X_n$ be a random sample from the inverse Gaussian distribution with pdf

$$f(x;\mu,\lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} e^{\frac{-\lambda(x-\mu)^2}{2\mu^2 x}}, \quad 0 < x < \infty.$$

Show that the statistics

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad T = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i} - \frac{1}{\bar{X}}}$$

are sufficient and complete.