Statistical Inference Assignment

i. Location Family

- **Definition**: Let $f(.;\theta, \theta \in R)$ be a family of densities indexed by a parameter θ. The parameter θ is defined to be a **location parameter** if and only $f(x;\theta)$ can be written as a function of $x - \theta$. $f(x;\theta) = h(x - \theta)$ for some function $h(.)$ Equivalently, θ is a location parameter for the density $f(x;\theta)$ of a random variable X if and only if the distribution of $X - \theta$ does not depend on θ

- **Importance**: To determine the location where the distrbution is centered (population mean) along horizontal axis. If θ is increased, the graph of the probability function shifts to the right on the horizontal axis and vice versa.

- **Example**: Let $X_1, X_2, \ldots, X_n$ are iid (identical and independent distributed) random sample from logistic distribution with location parameter $\theta \in R$ and scale parameter $b = 1$. The probability density function (pdf): $f(x;\theta)$ is given by:

$$f(x;a) = \frac{e^{(x-\theta)}}{1 + e^{x-\theta}}, x \in R$$

  - By letting $y = x - \theta$, we have the following expression in which the distribution of $y = x - \theta$ does not depend on θ.

$$f(y;a) = \frac{e^{y}}{1 + e^{y}}, x \in R$$

ii. Scale Family

- **Definition**: Let $f(.;\theta, \theta > 0)$ be a family of densities indexed by a parameter θ. The parameter θ is defined to be a **location parameter** if and only $f(x;\theta)$ can be written as a function of $\frac{1}{\theta}h(\frac{x}{\theta})$. $f(x;\theta) = h(x - \theta)$ for some function $h(.)$. Equivalently, θ is a location parameter for the density $f(x;\theta)$ of a random variable X if and only if the distribution of $\frac{X}{\theta}$ does not depend on θ.

- **Importance**: To scale the shape of a distribution, which usually steches, squeeze or even change the general shape of distribution along horizontal axis.

- **Example**: Let $X_1, X_2, \ldots, X_n$ are iid random sample from exponential distribution with mean of θ. The probability density function (pdf): $f(x;\theta)$ is given by:

$$f(x;\theta) = \frac{1}{\theta}e^{(-\frac{x}{\theta})},\ for\ x \geq 0$$

  - By letting $y = \frac{x}{\theta}$, we have the following expression in which the distribution of $y = \frac{x}{\theta}$ does not depend on θ.

$$f(y;\theta) = \frac{1}{\theta} \cdot e^y,\ for\ x \geq 0$$

.

iii. Sufficient statistic

- **Definition**: Let $X_1, X_2, \ldots, X_n$ be random sample from density $f(\cdot, \theta)$ where θ may be a vector. A statistic S = $s(X_1, X_2, \ldots, X_n)$ is defined to be sufficient statistics iff the conditional distribution of $X_1, X_2, \ldots, X_n$ given S =s does not depend on θ.

- **Importance**: To act as a form of data reduction;To condense all the inportant information of sample in a statistics in such as way that no information of parameter θ is lost.

- **Factorization Theorem (single sufficient statistic)**: Let $X_1, X_2, \ldots, X_n$ be random sample from density $f(\cdot, \theta)$ where θ may be a vector. A statistics S = $s(X_1, X_2, \ldots, X_n)$ is sufficient iff the joint density of $X_1, X_2, \ldots, X_n$ can be expressed as:

$$\prod_{i=1}^{n} f(x_i;\theta) = g(s(x_1, \ldots, x_n);\theta)h(x_1, \ldots, x_n) = g(s;\theta)h(x_1, \ldots, x_n)$$

- **Example**: Let $X_1, X_2, \ldots, X_n$ are iid random sample from a Poisson distribution with mean of $\theta > 0$. The probability mass function (pmf): $f(x;\theta)$ is given by:

$$f(x;\theta) = \frac{\theta^x e^{-\theta}}{x!}$$

  - Using a statistic of $s = \sum_{i=1}^{n} x_i$, its likelihood is expressed by:

$$\begin{aligned} L(\lambda; x_i) &= \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!} && (1) \\ &= \frac{\theta^{\sum_{i=1}^{n} x_i} e^{-n\theta}}{\prod_{i=1}^{n}(x_i)!} && (2) \\ &= \underbrace{\theta^s e^{-n\theta}}_{g(s;\theta)} \cdot \underbrace{\frac{1}{\prod_{i=1}^{n}(x_i!)}}_{h(x_1,\ldots,x_n)} && (3) \end{aligned}$$

  - Hence by factorization theorem, we can say that a statistic of $s = \sum_{i=1}^{n} x_i$ is **sufficient** to condense all important information of sample drawn from a Poisson distribution, without losing the information of θ.

iv. Minimal sufficient statistic

- **Definition**: A set of jointly sufficient statistics is defined to be **minimal sufficient** iff it is a **function** of every other set of sufficient statistics. Given two sample points of x and y, $S(X)$ is a minimal sufficient iff the ratio of likelihood, $\frac{L(\theta;x)}{L(\theta;y)}$ is a constant as a function of θ iff $S(X) = S(Y)$.

- **Importance**: A statistic captures all possible information of θ most efficiently.

- **Example**: Let $X_1, X_2, \ldots, X_n$ are iid random sample from a **Bernoulli distribution**, to show a test statistic $s = \sum X_i$ is minimal sufficient statistic:

  - The ratio of likelihood of 2 samples, X and Y is expressed as below:

$$\frac{L(\theta;x)}{L(\theta;y)} = \frac{\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}}{\prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}} \tag{4}$$

$$= \frac{p^{\sum_{i=1}^{n} x_i}(1-p)^{\sum_{i=1}^{n}(1-x_i)}}{p^{\sum_{i=1}^{n} y_i}(1-p)^{\sum_{i=1}^{n}(1-y_i)}} \tag{5}$$

  - By algebra, we know that $\sum_{i=1}^{n} x_i = n\bar{X}$ and $\sum_{i=1}^{n} y_i = n\bar{Y}$

$$\frac{L(\theta;x)}{L(\theta;y)} = \frac{p^{n\bar{X}}(1-p)^{n(1-\bar{X})}}{p^{n\bar{Y}}(1-p)^{n(1-\bar{Y})}} \tag{6}$$

$$= p^{n(\bar{X}-\bar{Y})}(1-p)^{n(\bar{Y}-\bar{X})} \tag{7}$$

  - The ratio of likelihood is a constant iff $\bar{X} = \bar{Y}$, hence, it implies that the sample mean $\bar{X}$ is minimum sufficient statistic. Any one-to-one function of $\bar{X}$ such as $\sum_{i=1}^{n} X_i$ is also a sufficient statistic.

v. Complete sufficient statistic

- **Definition**: Let $X_1, X_2, \ldots, X_n$ be random sample from density $f(\cdot, \theta)$ with parameter space Ω, and let T = $t(X_1, X_2, \ldots, X_n)$ be a statistic. The family of densitiies of T is defined to be complete iff $E_\theta[(z(T)] \equiv 0$ for all $\theta \in \Omega$ implied that $P_\theta[(z(T) = 0] \equiv 1$ for all $\theta \in \Omega$, where $z(T)$ is a setatistic. The statistic T is said to be complete iff its family of densities is complete.

- **Importance**: It ensures that the distributions corresponding to different values of the parameters are distinct. According to **Lehmann–Scheffé theorem**, it states that leting T = $t(X_1, X_2, \ldots, X_n)$ be a sufficient statistic for a parameter θ if there is a function of T that is an unbiased estimator of θ, then this function of T is the unique **minimum variance unbiased estimator (MVUE)** of θ.

- **Example**: Let $X_1, X_2, \ldots, X_n$ are iid random sample from a **Exponential distribution** with mean of $\frac{1}{\theta}$ and $\theta > 0$, to show a test statistic $T = \sum X_i$ is a complete sufficient statistic:

  - The pdf of Exponential distribution is:

$$f(x;\lambda) = \theta e^{-\theta x_i} \tag{8}$$

  - We need to show that $E_\theta[(z(T)] \equiv 0$

$$E_\theta[(z(T)] = \int z(t)P_\theta(t)\ dt \tag{9}$$

$$= \int_0^\theta z(t)\theta e^{-\theta t}\ dt \tag{10}$$

$$= \underbrace{\theta}_{>0} \int_0^\theta z(t)e^{-\theta t}\ dt \tag{11}$$

- Looking at the fundamental theorem of calculus:

$$\frac{d}{d\theta}\int_0^{k(\theta)} f(u)\ du = f(k(\theta)) \cdot \frac{dk(\theta)}{d\theta} \tag{12}$$

- Let $k(\theta) = \theta$ which implies $\frac{dk(\theta)}{d\theta} = 1$, and $f(u) = z(t)e^{-\theta t}$:

$$\frac{d}{d\theta}\int_0^{\theta} z(t)e^{-\theta t}\ dt = f(k(\theta)) \cdot 1 = 0 \tag{13}$$

- This also implies that

$$z(t) \underbrace{e^{-\theta t}}_{e^x > 0\ for x \in R} = 0 \tag{14}$$

$$z(t) = 0 \tag{15}$$

- This also implies that $z(t) = 0$ for all $t \in R$, hence, $T = \sum X_i$ is a complete sufficient statistic.

vi. Ancillary statistic

- **Definition**: A statistic S = $s(X_1, X_2, \ldots, X_n)$ whose distribution **does not depend** on the parameters.
- **Importance**:The value of statistic does not change as the parameter of the distribution changes.
- **Example**: The following statistics do not contain any information of parameters, but only the descriptive statistic from samples:
  - Sample mean, $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$,
  - Sample variance $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$

2. Exponential Family of Distributions

i. Definition

- A one-parameter family (θ is undimensional of densities $f(\cdot;\theta)$ that can be expressed as the following:

$$f(x;\theta) = a(\theta)b(x)\exp[c(\theta)d(x)]$$

is defined to belong to the **exponential family** for $-\infty < x < \infty$, for all $\theta \in R$.

- A k-parameter family (θ is undimensional of densities $f(\cdot;\theta_1,\ldots,\theta_k)$ that can be expressed as the following:

$$f(x;\theta_1,\ldots,\theta_k) = a(\theta_1,\ldots,\theta_k)b(x)\exp\sum_{j=1}^{k}[c_j(\theta_1,\ldots,\theta_k)d_j(x)]$$

is defined to belong to the **exponential family** for $-\infty < x < \infty$, for all $\theta \in R$.

ii. Advantages

a. Mathematical convenience.
b. Demonstration of **complete** minimal sufficient statistics

a. Mathematical convenience.

- The form is designed for **mathematical convenience** to express useful algebraic properties.
- It is much clearer in determining likelihood expression; direct applying multiplication and summation at $a(\theta)b(x)$ and $\exp[c(\theta)d(x)]$, respectively.
- For example, given $X_1, X_2, \ldots, X_n$ are iid. (identical and independent distributed) observation from an exponential distribution with mean of $\frac{1}{\lambda}$, the pdf is in the form of:

$$f(x;\lambda) = \lambda e^{(-\lambda x)} for x \geq 0$$

- The pdf can also be expressed in the form of exponential family as such:

$$f(x;\lambda) = \lambda \cdot 1 \cdot exp(-\lambda \cdot x)$$

- Hence, for likelihood, it can be easily expressed in the form of:

$$L(\lambda; x_i) = \prod_{i=1}^{n} \lambda \cdot 1 \cdot exp(-\lambda \cdot x) \tag{16}$$
$$= \lambda^n exp(-\lambda \cdot \sum x_i) \tag{17}$$

b. Demonstration of complete minimal sufficient statistics

- **Theorem**: given $X_1, X_2, \ldots, X_n$ are iid random sample from an exponential family distribution, then $\sum d(X_i)$ is a complete minimal sufficient statistics.
- For example, given $X_1, X_2, \ldots, X_n$ are iid from an exponential distribution with mean of $\frac{1}{\lambda}$, its pdf expression in the form of exponential family is:

$$f(x;\lambda) = \lambda \cdot 1 \cdot exp(-\lambda \cdot x)$$

- d(x) is x, hence by convenience of theorem without explicitly proving, $\sum d(X_i) = \sum(X_i)$ is a complete minimal sufficient statistics.

3 Basu's Theorem

- **Theorem**: If T(X) is a complete and minimal sufficient statistics, then T(X) is independent of every ancillary statistics.

- **Example**: Given $X_1, X_2, \ldots, X_n$ are iid from a normal distribution with known mean and variance of μ and σ^2, respectively. We want to show that the sample mean as a complete and minimal sufficient statistic: $T(X) = \frac{\sum_{i=1}^{n} X_i}{n}$ is independent of an ancillary statistic such as sample variance.

  - Probability density function(pdf) of $N(\mu, \sigma)$ is defined as :

$$\begin{aligned} f(x; \mu, \sigma) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} && (18) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(\mu)^2}{2\sigma^2}} \cdot 1 \cdot exp[\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x] && (19) \end{aligned}$$

- The pdf is in the form of k-parameter exponential family in the form of $a(\theta)b(x)\exp\sum_{j=1}^{k}[c_j(\theta_1, \ldots, \theta_k)d_j(x)]$, with k = 2, by having:
  - $a(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(\mu)^2}{2\sigma^2}}$
  - $b(x) = 1$
  - $c_1(\mu, \sigma) = -\frac{1}{2\sigma^2}$
  - $c_2(\mu, \sigma) = -\frac{\mu}{\sigma^2}$
  - $d_1(x) = x^2$
  - $d_2(x) = x$

- By theorem, $\sum d_1(X_i) = \sum_i^n x_i^2$ and $\sum d_2(X_i) = \sum_i^n x_i$ are complete minimal statistics. Only referring to $\sum d_2(X_i) = \sum_i^n x_i$, this implies that the sample mean is also a complete minimal statistic because n (number of observation is known).

- Sample variance $S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$, is one of the ancilliary statistics as it does not depend on the parameters of μ and σ.

- According to Basu's Theorem, $T(X) = \bar{X} = \sum_i^n x_i$ is independent of sample variance S^2.

Citation

1. Casella, G., & Berger, R. L. (2002). Statistical inference. Belmont, CA: Duxbury.