MA 3140: Statistical Inference

Dr. Sameen Naqvi
Department of Mathematics, IIT Hyderabad
Email id: sameen@math.iith.ac.in

Some natural questions

- Is the dimension of the sufficient statistic and the parameter always same?
- Is it always possible to reduce the data dimension?
- Are the sufficient statistics related to MLE and/or Fisher's Information?
- Is the sufficient statistic unique? If not, how to decide which one is a better sufficient statistic?

Example 1

Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.
Case 1: σ^2 is known (say 1). Find the sufficient statistic for μ.
Solution:

$$\begin{aligned} f(\boldsymbol{x}, \mu) &= \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i - \mu)^2} \right\} \\ &= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\sum (x_i - \mu)^2} \\ &= \underbrace{\frac{1}{(\sqrt{2\pi})^n} e^{-\frac{\sum x_i^2}{2}}}_{h(\boldsymbol{x})} \underbrace{e^{-\frac{n\mu^2}{2} + n\mu\overline{x}}}_{g(\overline{x}, \mu)} \end{aligned}$$

Thus, Factorization Theorem gives $\overline{X}$ as a sufficient statistic for $\{N(\mu, 1) : \mu \in \mathbb{R}\}$.

Example 1 cont'd

Case 2: μ is known (say $\mu = \mu_0$). Find the sufficient statistic for σ^2.

Solution:

$$
\begin{aligned}
f(\boldsymbol{x}, \sigma^2) &= \prod_{i=1}^{n} \left\{ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu_0)^2} \right\} \\
&= \underbrace{\frac{1}{(\sqrt{2\pi})^n}}_{h(\boldsymbol{x})} \underbrace{\frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu_0)^2}}_{g\left(\sum(x_i-\mu_0)^2, \sigma^2\right)}
\end{aligned}
$$

Thus, Factorization Theorem gives $\sum(X_i - \mu_0)^2$ as a sufficient statistic for $\{N(\mu_0, \sigma^2) : \sigma^2 > 0\}$.

Example 1 cont'd

Remark:

$$f(\boldsymbol{x}, \sigma^2) = \frac{1}{(\sqrt{2\pi})^n} \; \frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2}\sum(x_i - \mu_0)^2}$$

$$= \frac{1}{(\sqrt{2\pi})^n} \; \underbrace{\frac{1}{\sigma^n} e^{-\frac{\sum x_i^2}{2\sigma^2} + \frac{\mu_0 n\bar{x}}{\sigma^2} - \frac{n\mu_0^2}{2\sigma^2}}}_{g^*\left(\bar{x},\ \sum x_i^2, \sigma^2\right)}$$

Thus, $(\overline{X}, \sum X_i^2)$ is also sufficient for σ^2.

- **Question: Which sufficient statistic should be preferred for σ^2?**

  Since $\sum(X_i - \mu_0)^2$ is one-dimensional and gives larger data reduction as compared to 2 dimensional sufficient statistic $(\overline{X}, \sum X_i^2)$, we prefer $\sum(X_i - \mu_0)^2$.

Case 3: Both μ and σ^2 are unknown. Find the sufficient statistics for μ and σ^2.

$$f(\boldsymbol{x}, \mu, \sigma^2) = \prod_{i=1}^{n} \left\{ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2} \right\}$$

$$= \underbrace{\frac{1}{(\sqrt{2\pi})^n}}_{h(\boldsymbol{x})} \; \underbrace{\frac{1}{\sigma^n} e^{-\frac{\sum x_i^2}{2\sigma^2} + \frac{\mu\sum x_i}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}}}_{g^*\left(\sum x_i,\ \sum x_i^2, \mu, \sigma^2\right)}$$

Thus, $(\sum X_i, \sum X_i^2)$ is sufficient.

Example 1 cont'd

Remark:

$$f(\boldsymbol{x}, \mu, \sigma^2) = \prod_{i=1}^{n} \left\{ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2} \right\}$$
$$= \frac{1}{(\sqrt{2\pi})^n} \frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2}\sum(x_i-\overline{x})^2 - \frac{n}{2\sigma^2}(\overline{x}-\mu)^2}$$

Thus, $(\overline{X}, S^2)$ is sufficient for $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$.

- Note that $(\sum X_i, \sum X_i^2)$ are one-to-one function of $(\overline{X}, S^2)$.

Hence, when both parameters are unknown, sample mean $\overline{X}$ and sample variance S^2 are sufficient.

Example 1 cont'd

Case 4: $\sigma = \mu$.

Then, $X_1, \ldots, X_n \sim N(\mu, \mu^2)$.

$$f(\boldsymbol{x}, \mu) = \underbrace{\frac{1}{(\sqrt{2\pi})^n}}_{h(\boldsymbol{x})} \underbrace{\frac{1}{\mu^n} e^{-\frac{n}{2} - \frac{\sum x_i^2}{2\mu^2} + \frac{\mu \sum x_i}{\mu^2}}}_{g^*\left(\sum x_i,\ \sum x_i^2,\ \mu\right)}$$

So, although parameter is one-dimensional, sufficient statistic $(\sum X_i, \sum X_i^2)$ is two-dimensional.

Thus, $(\sum X_i, \sum X_i^2)$ is sufficient for $\{N(\mu, \mu^2) : \mu > 0\}$..

Example 2

Let $X_1, \ldots, X_n$ be a random sample from distribution with pdf

$$f(x,\theta) = \begin{cases} e^{-(x-\theta)}, & x > \theta \\ 0, & \text{o/w.} \end{cases}$$

Find the sufficient statistic for θ.

Solution: The joint density of $X_1, \ldots, X_n$ is

$$\begin{aligned} f(\boldsymbol{x},\theta) &= \begin{cases} e^{-\sum x_i + n\theta}, & x_i > \theta, i = 1, \ldots, n \\ 0, & \text{o/w.} \end{cases} \\ &= \begin{cases} e^{-\sum x_i + n\theta}, & \infty > x_{(n)} > \ldots > x_{(1)} > \theta \\ 0, & \text{o/w.} \end{cases} \\ &= e^{-\sum x_i} \; e^{n\theta} \; I_{(\theta,\infty)} x_{(1)} \; \prod_{i=2}^{n} I_{(x_{(1)},\infty)} x_{(i)} \\ &= g(x_{(1)},\theta) h(\boldsymbol{x}), \end{aligned}$$

Example 2 contd.

where,

$$h(\boldsymbol{x}) = e^{-\sum x_i} \prod_{i=2}^{n} I_{(x_{(1)},\infty)} x_{(i)} = e^{-\sum x_i},$$

and

$$g(x_{(1)}, \theta) = e^{n\theta} \; I_{(\theta,\infty)} x_{(1)}.$$

Thus, $X_{(1)}$ is sufficient for θ.

Example 3

Let $X_1, \ldots, X_n$ be a random sample from a two parameter exponential distribution with pdf

$$f(x, \theta, \sigma) = \begin{cases} \frac{1}{\sigma}\, e^{-\frac{x-\theta}{\sigma}}, & x > \theta \\ 0, & \text{o/w.} \end{cases}$$

Find the sufficient statistic for θ and σ.

Solution: The joint density of $X_1, \ldots, X_n$ is

$$f(\boldsymbol{x}, \theta, \sigma) = \begin{cases} \frac{1}{\sigma^n} e^{\frac{n\theta}{\sigma}} e^{-\frac{\sum x_i}{\sigma}}, & x_i > \theta, i = 1, \ldots, n \\ 0, & \text{o/w.} \end{cases}$$

$$= \underbrace{\frac{1}{\sigma^n} e^{\frac{n\theta}{\sigma}} e^{-\frac{\sum x_i}{\sigma}} I_{(\theta,\infty)} x_{(1)}}_{g\left(x_{(1)}, \sum x_i, \theta, \sigma\right)} \underbrace{\prod_{i=2}^{n} I_{(x_{(1)},\infty)} x_{(i)}}_{h(\boldsymbol{x})}.$$

So, $(X_{(1)}, \sum X_i)$ is sufficient or $(X_{(1)}, \overline{X})$ is sufficient.

Example 4

Let $X_1, \ldots, X_n$ be a random sample from Double Exponential distribution wih pdf

$$f(x,\theta) = \frac{1}{2} e^{-|x-\theta|}, \quad x \in \mathbb{R}, \theta \in \mathbb{R}.$$

Find the sufficient statistic for θ.

Solution: The joint density of $X_1, \ldots, X_n$ is

$$f(\boldsymbol{x},\theta) = \frac{1}{2^n} e^{-\sum |x_i - \theta|} = \underbrace{\frac{1}{2^n}}_{h(\boldsymbol{x})} \underbrace{e^{-\sum |x_{(i)} - \theta|}}_{g(x_{(1)},\ldots,x_{(n)},\theta)}$$

Thus, $(X_{(1)}, \ldots, X_{(n)})$ is sufficient.

Example 5

Let $X_1, \ldots, X_n$ be a random sample from $U(0, \theta)$, $\theta > 0$. Find the sufficient statistic for θ.

Solution: The joint density of $X_1, \ldots, X_n$ is

$$
\begin{aligned}
f(\boldsymbol{x}, \theta) &= \begin{cases} \frac{1}{\theta^n}, & 0 < x_i < \theta, \ i = 1, \ldots, n \\ 0, & \text{o/w.} \end{cases} \\
&= \begin{cases} \frac{1}{\theta^n}, & 0 < x_{(1)} < \ldots < x_{(n)} < \theta \\ 0, & \text{o/w.} \end{cases} \\
&= \underbrace{\frac{1}{\theta^n} \; I_{(0,\theta)} x_{(n)}}_{g(x_{(n)}, \theta)} \underbrace{\prod_{i=1}^{n-1} I_{(0, x_{(n)})} x_{(i)}}_{h(\boldsymbol{x})}
\end{aligned}
$$

Thus, $X_{(n)}$ is sufficient.

Example 6

Let $X_1, \ldots, X_n \sim U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, $\theta > 0$. Find the sufficient statistic for θ.

Solution: The joint density of $X_1, \ldots, X_n$ is

$$f(\boldsymbol{x}, \theta) = \begin{cases} 1, & \theta - \frac{1}{2} < x_{(1)} < \ldots < x_{(n)} < \theta + \frac{1}{2} \\ 0, & \text{o/w.} \end{cases}$$

$$= \underbrace{I_{\left(\theta-\frac{1}{2},\theta+\frac{1}{2}\right)}^{x_{(1)}} I_{\left(\theta-\frac{1}{2},\theta+\frac{1}{2}\right)}^{x_{(n)}}}_{g(x_{(1)},x_{(n)},\theta)} \underbrace{\prod_{i=2}^{n-1} I_{\left(x_{(1)},x_{(n)}\right)}^{x_{(i)}}}_{h(\boldsymbol{x})}$$

Thus, $(X_{(1)}, X_{(n)})$ is sufficient.

Example 7: Exponential Family

For a multi-parameter exponential family with pdf

$f(x, \boldsymbol{\theta}) = c(\boldsymbol{\theta}) h(x) e^{\sum\limits_{i=1}^{k} Q_i(\boldsymbol{\theta}) T_i(x)}$, $\boldsymbol{\theta} \in \mathbb{R}^k$, find the sufficient statistics for θ.

Solution: The joint density of $X_1, \ldots, X_n$ is

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = c^n(\boldsymbol{\theta}) \prod_{j=1}^{n} h(x_j) \; e^{\sum\limits_{j=1}^{n} \sum\limits_{i=1}^{k} Q_i(\boldsymbol{\theta}) T_i(x_j)}$$

$$= \underbrace{c^n(\boldsymbol{\theta}) \; e^{\sum\limits_{i=1}^{k} Q_i(\boldsymbol{\theta}) \sum\limits_{j=1}^{n} T_i(x_j)}}_{g\left(\sum\limits_{j=1}^{n} T_1(x_j), \sum\limits_{j=1}^{n} T_2(x_j), \ldots, \sum\limits_{j=1}^{n} T_k(x_j), \boldsymbol{\theta}\right)} \quad \underbrace{\prod_{j=1}^{n} h(x_j)}_{h(\boldsymbol{x})}$$

Thus, $\left(\sum\limits_{j=1}^{n} T_1(x_j), \sum\limits_{j=1}^{n} T_2(x_j), \ldots, \sum\limits_{j=1}^{n} T_k(x_j) \right)$ is sufficient.

Fisher's Information and Sufficiency

Recall that under regularity conditions, Fisher's information contained in random variable X about θ is defined as

$$I_X(\theta) = E\left[\frac{\partial}{\partial\theta}\log f(X,\theta)\right]^2.$$

If T is any statistic with density $\phi(t,\theta)$, then Fisher's information contained in T is defined as

$$I_T(\theta) = E\left[\frac{\partial}{\partial\theta}\log \phi(T,\theta)\right]^2.$$

Theorem: $I_{\mathbf{X}}(\theta) \geq I_T(\theta)$ with equality holding if, and only if, T is sufficient.

Example 1

Let $X_1, \dots, X_n \sim P(\lambda)$, $\lambda > 0$. Compare the information contained in each of the following statistics

$$T_1 = X_1, \qquad T_2 = X_1 + X_2, \qquad T_3 = X_1 + \cdots X_n = \sum X_i.$$

Solution: Here, $T_1 \sim P(\lambda)$, $T_2 \sim P(2\lambda)$ and $T_3 \sim P(n\lambda)$. Consider

$$\begin{aligned} f_\lambda(x) &= \frac{e^{-\lambda}\lambda^x}{x!} \\ \log f_\lambda(x) &= -\lambda + x \log \lambda - \log x! \\ \frac{\partial}{\partial \lambda} \log f_\lambda(x) &= -1 + \frac{x}{\lambda} = \frac{x-\lambda}{\lambda} \\ E\left[\frac{\partial}{\partial \lambda} \log f_\lambda(X)\right]^2 &= \frac{1}{\lambda^2} E(X-\lambda)^2 = \frac{1}{\lambda^2}\lambda = \frac{1}{\lambda} \end{aligned}$$

Example 1 contd.

Thus,

$$I_{T_1}(\lambda) = \frac{1}{\lambda}$$
$$I_{T_2}(\lambda) = \frac{2}{\lambda}$$
$$I_{T_3}(\lambda) = \frac{n}{\lambda}$$
$$I_{\boldsymbol{X}}(\lambda) = \frac{n}{\lambda}$$

Thus, $I_{\boldsymbol{X}}(\lambda) = I_{T_3}(\lambda)$ as T_3 is sufficient.

Example 2

Let $X_1, \ldots, X_n \sim N(\mu, 1)$. Compare the information contained in

$$T_1 = X_1 - X_2, \qquad T_2 = X_1 + \cdots X_n = \sum X_i.$$

Solution: Here, $T_1 \sim N(0, 2)$ and $T_2 \sim N(n\mu, n)$.

Consider T_1.

$$f_{T_1}(t_1) = \frac{1}{\sqrt{2\pi}\sqrt{2}} \exp^{-\frac{t_1^2}{4}}$$

$$\log f = \text{indp of } \mu$$

$$\frac{\partial}{\partial \mu} \log f = 0$$

Thus, $I_{T_1}(\mu) = 0$.

Example 2 contd.

Consider T_2.

$$f_{T_2}(t_2) = \frac{1}{\sqrt{2\pi}\sqrt{n}} \exp^{-\frac{1}{2n}(t_2 - n\mu)^2}$$

$$\log f_{T_2}(t_2) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log n - \frac{1}{2n}(t_2 - n\mu)^2$$

$$\frac{\partial}{\partial\mu}\log f_{T_2}(t_2) = t_2 - n\mu$$

$$E\left[\frac{\partial}{\partial\mu}\log f_{T_2}(t_2)\right]^2 = E(T_2 - n\mu)^2 = n.$$

Thus, $I_{T_2}(\mu) = n$ which is same as $I_{\boldsymbol{X}}(\mu) = n$.

Thanks for your patience!