MA 3140: Statistical Inference

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

SUMMARY

Goal: To obtain UMVUE.

Action Plan:

- Find an estimator (MME or MLE)
- Check if it has certain desirable properties (Unbiasedness, Consistency)
- Evaluate the estimator (MSE, Variance, CRLB, Efficiency)

Principles of Data Reduction: Sufficiency

Sufficient Statistic

Let $X_1, \ldots, X_n$ be a random sample from a population P_θ, $\theta \in \Theta$.

A statistic that captures all the information about the parameter θ contained in the sample is said to be sufficient.

Formally, a statistic $T = T(\boldsymbol{X})$ is said to be sufficient for $P = \{P_\theta : \theta \in \Theta\}$ if the conditional distribution of the sample $\boldsymbol{X}$ given the value of $T(\boldsymbol{X})$ does not depend on θ.

Example 1: Binomial Distribution

Let $X_1, \ldots, X_n$ be a random sample from Bernoulli distribution with parameter p, $0 < p < 1$. Check whether $T = \sum_{i=1}^{n} X_i$ is a sufficient statistic or not.

Solution: Consider the conditional distribution of $X_1, \ldots, X_n$ given $T = t$

$$P(X_1 = x_1, \ldots, X_n = x_n | T = t) = \frac{P(X_1 = x_1, \ldots, X_n = x_n, T = t)}{P(T = t)}$$

$$= \begin{cases} \frac{P\left(X_1 = x_1, \ldots, X_n = t - \sum_{i=1}^{n-1} x_i\right)}{P(T=t)}, & \text{if } t = \sum_{i=1}^{n} x_i \\ 0, & \text{if } t \neq \sum_{i=1}^{n} x_i \end{cases}$$

Example 1: Binomial Distribution cont'd

For $t = \sum_{i=1}^{n} x_i$,

$$\frac{P(X_1 = x_1)\dots P(X_{n-1} = x_{n-1})P\left(X_n = t - \sum_{i=1}^{n-1} x_i\right)}{P(T = t)}$$

$$= \frac{p^{x_1}(1-p)^{1-x_1} \;\dots\; p^{x_{n-1}}(1-p)^{1-x_{n-1}} \; p^{t-\sum_{i=1}^{n-1} x_i}(1-p)^{1-t+\sum_{i=1}^{n-1} x_i}}{\binom{n}{t} p^t (1-p)^{n-t}}$$

$$= \frac{p^t(1-p)^{n-t}}{\binom{n}{t} p^t (1-p)^{n-t}}$$

$$= \frac{1}{\binom{n}{t}}$$

Example 1: Binomial Distribution cont'd

So,

$$P(X_1 = x_1, \ldots, X_n = x_n | T = t) = \begin{cases} \frac{1}{\binom{n}{t}}, & \text{if } t = \sum x_i \\ 0, & \text{if } t \neq \sum x_i \end{cases}$$

This is independent of p.

Thus, $T = \sum X_i$ is sufficient for $\{Ber(p) : 0 < p < 1\}$.

Interpretation: The total number of 1s in this Bernoulli sample contains all the information about p.

Example 2A: Poisson Distribution

Let X_1, X_2 be iid $P(\lambda)$ random variables. Check whether $T = X_1 + X_2$ is a sufficient statistic or not.

Solution: Consider the conditional distribution of X_1, X_2 given $T = t$

$$
\begin{aligned}
&P(X_1 = x_1, X_2 = x_2 | X_1 + X_2 = t) \\
&\quad = \frac{P(X_1 = x_1, X_2 = x_2, X_1 + X_2 = t)}{P(X_1 + X_2 = t)} \\
&\quad = \begin{cases} \frac{P(X_1 = x_1, X_2 = t - x_1)}{P(X_1 + X_2 = t)}, & \text{if } t = x_1 + x_2, x_i = 0, 1, 2 \ldots \\ 0, & \text{o/w} \end{cases}
\end{aligned}
$$

Example 2A: Poisson Distribution cont'd

For $x_i = 0, 1, 2, \ldots$, $x_1 + x_2 = t$, we have

$$\frac{P(X_1 = x_1, X_2 = t - x_1)}{P(X_1 + X_2 = t)} = \frac{\frac{e^{-\lambda}\ \lambda^{x_1}}{x_1!}\ \frac{e^{-\lambda}\ \lambda^{t-x_1}}{(t-x_1)!}}{\frac{e^{-2\lambda}(2\lambda)^t}{t!}}$$
$$= \binom{t}{x_1}\left(\frac{1}{2}\right)^t$$

which is independent of λ.

Thus, $T = X_1 + X_2$ is sufficient for $\{P(\lambda) : \lambda > 0\}$.

Example 2B: Poisson Distribution

Let $X_1, \ldots, X_n$ be a random sample from Poisson distribution with parameter λ, $\lambda > 0$. Check whether $T = \sum_{i=1}^{n} X_i$ is a sufficient statistic or not.

Solution: Consider the conditional distribution of $X_1, \ldots, X_n$ given $T = t$

$$P(X_1 = x_1, \ldots, X_n = x_n | T = t) = \frac{P(X_1 = x_1, \ldots, X_n = x_n, T = t)}{P(T = t)}$$

$$= \begin{cases} \frac{P\left(X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = t - \sum_{i=1}^{n-1} x_i\right)}{P(T=t)}, & \text{if } t = \sum_{i=1}^{n} x_i \\ 0, & \text{if } t \neq \sum_{i=1}^{n} x_i \end{cases}$$

Example 2B: Poisson Distribution cont'd

For $t = \sum_{i=1}^{n} x_i$,

$$\frac{P\left(X_1 = x_1\right) \ \ldots \ P\left(X_{n-1} = x_{n-1}\right) \ P\left(X_n = t - \sum_{i=1}^{n-1} x_i\right)}{P(T = t)}$$

$$= \frac{\dfrac{e^{-\lambda} \ \lambda^{x_1}}{x_1!} \ \ldots \ \dfrac{e^{-\lambda} \ \lambda^{x_{n-1}}}{x_{n-1}!} \ \dfrac{e^{-\lambda} \ \lambda^{\left(t - \sum_{i=1}^{n-1} x_i\right)}}{\left(t - \sum_{i=1}^{n-1} x_i\right)!}}{\dfrac{e^{-n\lambda}(n\lambda)^t}{t!}}$$

$$= \frac{\left[\prod_{i=1}^{n-1} \dfrac{e^{-\lambda} \ \lambda^{x_i}}{x_i!}\right] \dfrac{e^{-\lambda} \ \lambda^{\left(t - \sum_{i=1}^{n-1} x_i\right)}}{\left(t - \sum_{i=1}^{n-1} x_i\right)!}}{\dfrac{e^{-n\lambda} \ (n\lambda)^t}{t!}} = \frac{t!}{x_1! \ \ldots x_{n-1}! \left(t - \sum_{i=1}^{n-1} x_i\right)!} \left(\frac{1}{n}\right)^t$$

Example 2B: Poisson Distribution cont'd

So,

$$P(X_1 = x_1, \dots, X_n = x_n | T = t)$$
$$= \begin{cases} \frac{t!}{x_1! \ \dots x_{n-1}! \left(t - \sum_{i=1}^{n-1} x_i\right)!} \left(\frac{1}{n}\right)^t, & \text{if } t = \sum x_i \\ 0, & \text{if } t \neq \sum x_i \end{cases}$$

This is independent of λ.

Thus, $T = \sum X_i$ is sufficient for $\{P(\lambda) : \lambda > 0\}$.

Remark 1: Not every statistic is sufficient

Example: Let X_1, X_2 be iid $P(\lambda)$ random variables. Let $T = X_1 + 2X_2$.

Solution: Consider the conditional distribution of X_1, X_2 given $T = t$

$$\begin{aligned}
P(X_1 = 0, X_2 = 1 | X_1 + 2X_2 = 2) \\
&= \frac{P(X_1 = 0, X_2 = 1, X_1 + 2X_2 = 2)}{P(X_1 + 2X_2 = 2)} \\
&= \frac{P(X_1 = 0, X_2 = 1)}{P(X_1 + 2X_2 = 2)} \\
&= \frac{e^{-\lambda}(\lambda e^{-\lambda})}{P(X_1 = 0, X_2 = 1) + P(X_1 = 2, X_2 = 0)} \\
&= \frac{\lambda e^{-2\lambda}}{\lambda e^{-2\lambda} + (\frac{\lambda^2}{2})e^{-2\lambda}} = \frac{1}{1 + \frac{\lambda}{2}}
\end{aligned}$$

which is not independent of λ.

Remark 2

- Let T be sufficient for $P = \{P_\theta : \theta \in \Omega\}$, and let T be a function of U. Then U is also sufficient for P.

- Entire sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ is always sufficient, and is called trivial sufficient statistics.

$$P(X_1 = x_1,\ \ldots X_n = x_n | X_1 = t_1,\ \ldots X_n = t_n) = \begin{cases} 1, & \text{if } \boldsymbol{t} = \boldsymbol{x} \\ 0, & \text{if } \boldsymbol{t} \neq \boldsymbol{x} \end{cases}$$

which is always free from the parameter.

Neyman-Fisher Factorization Theorem

Let $X_1, \ldots, X_n$ be discrete (or continuous) random variables with pmf (or pdf) $f(\boldsymbol{x}, \theta)$, $\theta \in \Theta$. Then $T(\boldsymbol{X})$ is sufficient if, and only if,

$$f(\boldsymbol{x}, \theta) = g(T(\boldsymbol{x}), \theta)\ h(\boldsymbol{x}), \forall\ \theta \in \Theta.$$

Here, h is a non-negative function of $\boldsymbol{x}$ and does not depend on θ, and g is a non-negative function of θ and $T(\boldsymbol{x})$ only.

Proof: Let $f(\boldsymbol{x}, \theta) = g(T(\boldsymbol{x}), \theta)\ h(\boldsymbol{x})$. Consider

$$\begin{aligned} P_\theta(T(\boldsymbol{X}) = t) = \sum_{\boldsymbol{x}: T(\boldsymbol{x}) = t} f(\boldsymbol{x}, \theta) &= \sum_{\boldsymbol{x}: T(\boldsymbol{x}) = t} g(T(\boldsymbol{x}), \theta)\ h(\boldsymbol{x}) \\ &= g(t, \theta) \sum_{\boldsymbol{x}: T(\boldsymbol{x}) = t} h(\boldsymbol{x}) \end{aligned}$$

Neyman-Fisher Factorization Theorem cont'd

Now,

$$P_\theta(\boldsymbol{X} = \boldsymbol{x} | T = t) = \frac{P_\theta(\boldsymbol{X} = \boldsymbol{x}, T(\boldsymbol{x}) = t)}{P_\theta(T(\boldsymbol{x}) = t)}$$

$$= \begin{cases} \frac{P_\theta(\boldsymbol{X}=\boldsymbol{x})}{P_\theta(T(\boldsymbol{x})=t)}, & \text{if } T(\boldsymbol{x}) = t \\ 0, & \text{if } T(\boldsymbol{x}) \neq t \end{cases}$$

Thus, if $T(\boldsymbol{x}) = t$, then

$$\frac{P_\theta(\boldsymbol{X} = \boldsymbol{x})}{P_\theta(T(\boldsymbol{x}) = t)} = \frac{g(t,\theta)h(\boldsymbol{x})}{g(t,\theta) \sum_{T(\boldsymbol{x})=t} h(\boldsymbol{x})}$$

which is free of θ.

So the conditional distribution of X given T is independent of the parameter, and hence T is a sufficient statistic.

Neyman-Fisher Factorization Theorem cont'd

Conversely, let T is sufficient for θ.

$$
\begin{aligned}
\implies \quad P_\theta(\boldsymbol{X} = \boldsymbol{x} | T(\boldsymbol{X}) = t) &= c(\boldsymbol{x}, t) \qquad \text{(indp. of } \theta) \\
\implies \quad \frac{P_\theta(\boldsymbol{X} = \boldsymbol{x}, T(\boldsymbol{X}) = t)}{P_\theta(T(\boldsymbol{X}) = t))} &= c(\boldsymbol{x}, t) \qquad \text{(if} \quad T(\boldsymbol{x}) = t) \\
\implies \quad P_\theta(\boldsymbol{X} = \boldsymbol{x}) &= c(\boldsymbol{x}, t)\ P_\theta(T(\boldsymbol{X}) = t) \\
&= c(\boldsymbol{x}, t) g(t, \theta) \\
&= h(\boldsymbol{x}) g(t, \theta)
\end{aligned}
$$

Hence proved.

Remark

(i) The theorem holds if $\boldsymbol{\theta}$ and $\boldsymbol{T}$ are vectors. In fact, their dimensions need not be same.

(ii) If T is sufficient and T is a function of U, say $\alpha(U)$, then,

$$\begin{aligned} f(x,\theta) &= g(T(x),\theta)\ h(x) \\ &= g(\alpha(U(x)),\theta)\ h(x) \\ &= g(\alpha(U),\theta)\ h(x) \end{aligned}$$

So, U is also sufficient by factorization theorem.

Remark cont'd

(iii) However, if V is a function of T then V need not be sufficient. If V is a one-to-one function of T, then V is sufficient.

Let $V = \beta(T)$. Then $T = \beta^{-1}(V)$.

Thus,

$$g(T, \theta) = g(\beta^{-1}(V), \theta) = g^*(V, \theta),$$

and V is also sufficient.

Thanks for your patience!