Minimum variance unbiased estimators

- In this handout, we consider the estimating problem of $g(\theta)$ based on data X, where

  (*) $X = (X_1, \ldots, X_n)$ is a sample, X has PDF (or PMF) f_θ, where $\theta \in \Theta$.

- Suppose that (*) holds. Suppose that $T(X)$ is a UE (unbiased estimator) of $g(\theta)$, then $T(X)$ is a MVUE (minimum variance unbiased estimator) of $g(\theta)$ if
$$Var(T(X)) \leq Var(U(X)) \text{ for every } \theta \in \Theta$$
for every $U(X)$ that is an UE of $g(\theta)$.

- A result due to Rao and Blackwell.

  Fact 1. *Suppose that S is a sufficient statistic for θ and T is an UE of $g(\theta)$. Then, $E(T|S)$ is also an UE of $g(\theta)$ and*
$$Var(T) \geq Var[E(T|S)]$$
for every $\theta \in \Theta$.

  Note. The inequality in the above fact follows from the fact that
$$E[X_0 - h(Y)]^2 = E[X_0 - E(X_0|Y)]^2 + E[E(X_0|Y) - h(Y)]^2 \qquad (1)$$
for X_0 and $h(Y)$ such that $E(X_0)$ is finite. Apply (1) with $h(Y) = E(X_0)$ and we have
$$Var(X_0) = E[X_0 - E(X_0|Y)]^2 + Var[E(X_0|Y)], \qquad (2)$$
so $Var(X_0) \geq Var[E(X_0|Y)]$. If $Var(X_0) < \infty$, then the equality $Var(X_0) = Var[E(X_0|Y)]$ holds if and only if $P(X_0 = E(X_0|Y)) = 1$.

- Lehmann and Scheffé Theorem. Suppose that $X = (X_1, \ldots, X_n)$ is a sample, X has PDF (or PMF) f_θ, where $\theta \in \Theta$. Suppose that

  (i) T is a UE (unbiased estimator) of $g(\theta)$ and

  (ii) S is a sufficient and complete statistic for θ,

  then $E(T|S)$ is a MVUE of $g(\theta)$. Moreover, if $Var[E(T|S)]$ is finite for every $\theta \in \Theta$ and T_* is a MVUE of $g(\theta)$, then
$$P(E(T|S) = T_*) = 1 \text{ for every } \theta \in \Theta.$$

- Definition of a complete statistic. Suppose that $X = (X_1, \ldots, X_n)$ is a sample, X has PDF (or PMF) f_θ, where $\theta \in \Theta$. For a statistic T, if
$$E[g(T)] = 0 \text{ for every } \theta \in \Theta \Rightarrow P(g(T) = 0) = 1 \text{ for every } \theta \in \Theta, \qquad (3)$$
then we say that T is a complete statistic (or the distribution of T belongs to a complete family).

- Notation of a binomial distribution. For a positive integer n and for $\theta \in [0, 1]$, we will use $Bin(n, \theta)$ to denote the binomial distribution with PMF p_θ, where

  $$p_\theta(x) = C_k^n \theta^k (1-\theta)^{n-k} I_{\{0,1,\ldots,n\}}(x)$$

  for $x \in (-\infty, \infty)$.

- Example 1. Suppose that $(X_1, \ldots, X_n)$ is a random sample from $Bin(1, \theta)$, where $\theta \in (0, 1)$. Let $\bar{X} = \sum_{i=1}^n X_i/n$. Show that $\bar{X}$ is a complete statistic for θ.

  Sol. Suppose that $E[g(\bar{X})] = 0$ for every $\theta \in (0, 1)$. Since $n\bar{X} \sim Bin(n, \theta)$, we have

  $$\sum_{k=0}^{n} g\left(\frac{k}{n}\right) C_k^n \theta^k (1-\theta)^{n-k} = 0 \text{ for every } \theta \in (0, 1). \tag{4}$$

  Let $a_k = g(k/n)C_k^n$ for $k \in \{0, \ldots, n\}$, then (4) implies that

  $$\sum_{k=0}^{n} a_k \lambda^k = 0 \text{ for every } \lambda \in \left\{\frac{\theta}{1-\theta} : \theta \in (0, 1)\right\} = (0, \infty). \tag{5}$$

  Note that from (5), we have a polynomial of λ that is equal to 0 on $(0, \infty)$, so $a_k = 0$ for $k \in \{0, \ldots, n\}$. Since $g(k/n) = a_k/C_k^n$, we have $g(k/n) = 0$ for each $k \in \{0, \ldots, n\}$. Note that $g(\bar{X})$ takes values in

  $$\{g(k/n) : k \in \{0, \ldots, n\}\} = \{0\},$$

  so $P(g(\bar{X}) = 0) = 1$ for every $\theta \in (0, 1)$. We have verified that $\bar{X}$ is a complete statistic.

- Exponential family (class) of PDFs (or PMFs). Suppose $\{f_\theta : \theta \in \Theta\}$ is a family of PDFs (or PMFs) on R^n. If there exist functions h, c_0, c_1, ..., c_k, T_1, ..., T_k such that

  $$f_\theta(x) = \exp\left(c_0(\theta) + \sum_{j=1}^{k} c_j(\theta) T_j(x)\right) h(x) \text{ for every } x \in R^n \tag{6}$$

  for every $\theta \in \Theta$, then $\{f_\theta : \theta \in \Theta\}$ is called an exponential family of PDFs (or PMFs). If the set

  $$\{(c_1(\theta), \ldots, c_k(\theta)) : \theta \in \Theta\}$$

  contains a nonempty open set in R^k, then the exponential family $\{f_\theta : \theta \in \Theta\}$ is said to be of full rank.

Fact 2. *Suppose that $X = (X_1, \ldots, X_n)$ is a sample and X has PDF (or PMF) f_θ, where $\theta \in \Theta$. Suppose that (6) holds for every $\theta \in \Theta$. Suppose that the exponential family $\{f_\theta : \theta \in \Theta\}$ is of full rank, then* $(T_1(X), \ldots, T_k(X))$ *is a k-dimensional sufficient and complete statistic for θ.*

- Some remarks on the exponential family $\{f_\theta : \theta \in \Theta\}$ with f_θ given in (6).
  - Usually, we consider $\Theta \subset R^m$ with $m \leq k$ to avoid the problem of non-identifiability.
  - To make sure that the exponential family is of full rank, we usually need to ensure that the following conditions hold:
    (i) $\Theta \subset R^k$,
    (ii) Θ contains a nonempty open set in R^k, and
    (iii) there is no $c_i(\theta)$ that can be expressed as a function of $c_j(\theta)$s, where $j \in \{1, \ldots k\}$ and $j \neq i$.

    Here (iii) generally implies that it is impossible to find constants a_0, a_1, ..., a_k such that a_0, a_1, ..., a_k are not all zeros and

    $$P\left(a_0 + \sum_{j=1}^{k} a_j T_j(X) = 0\right) = 1 \text{ for all } \theta \in \Theta.$$

  - Another way to check whether the exponential family is of full rank is to apply the following result and verify the conditions numerically.

    Fact 3. *Consider the exponential family $\{f_\theta : \theta \in \Theta\}$ with f_θ given in (6). Suppose $\Theta \subset R^k$ and let $g(\theta) = (c_1(\theta), \ldots, c_k(\theta)) : \theta \in \Theta\}$ and let $D_g(\theta)$ be the $k \times k$ matrix with the (i, j)-th element*

    $$\frac{\partial}{\partial \theta_j} c_i(\theta).$$

    Suppose there exists θ_0 in Θ such that

    $$B(\theta_0, \delta) = \{\eta \in R^k : \|\eta - \theta_0\| < \delta\}$$

    is contained in Θ for some $\delta > 0$ and each element of $D_g(\theta)$ is a continuous function of θ on $B(\theta_0, \delta)$. If the determinant of $D_g(\theta)$ is nonzero on $B(\theta_0, \delta)$, then the exponential family $\{f_\theta : \theta \in \Theta\}$ is of full rank.

- Example 2. Suppose that $(X_1, \ldots, X_n)$ is a random sample from $N(\mu, \sigma^2)$, where $n \geq 2$, $\mu \in (-\infty, \infty)$ and $\sigma > 0$. Find a sufficient and complete statistic for (μ, σ).

  Ans. $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ is a sufficient and complete statistic for (μ, σ). However, the answer is not unique according to Fact 4. For instance, $(\bar{X}, \sum_{i=1}^n (X_i - \bar{X})^2)$ is also a sufficient and complete statistic for (μ, σ).

Fact 4. *Suppose that $(T_1, \ldots, T_k)$ is a k-dimensional sufficient and complete statistic for θ. Suppose that $(U_1, \ldots, U_k)$ is a statistic so that*

$$(U_1, \ldots, U_k) = g(T_1, \ldots, T_k)$$

for some function g and

$$(T_1, \ldots, T_k) = h(U_1, \ldots, U_k)$$

for some function h, then $(U_1, \ldots, U_k)$ is also a sufficient and complete statistic for θ.

- To find a MVUE using Lehman Scheffé Theorem, we need to compute the conditional expectation given a sufficient and complete statistic. The following properties are useful (assuming all conditional expectations involved are finite):
  - $E(E(X|Y)) = E(X)$.
  - Suppose that X and Y are independent, then $E(X|Y) = E(X)$.
  - $E(Xh(Y)|Y) = h(Y)E(X|Y)$.
  - $E(h(Y)|Y) = h(Y)$.
  - Suppose that a is a constant, then $E(aX|Y) = aE(X|Y)$.
  - $E((X_1 + X_2)|Y) = E(X_1|Y) + E(X_2|Y)$.

- Example 3. Suppose that $(X_1, \ldots, X_n)$ is a random sample from $N(\mu, \sigma^2)$, where $\mu \in (-\infty, \infty)$ and $\sigma > 0$. Find a MVUE of μ.

  Sol. Since $E(\bar{X}) = \mu$, $\bar{X}$ is an UE of μ. Let $S = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$, then S is a sufficient and complete statistic for (μ, σ) from Example 2. By Lehmann and Scheffé theorem, a MVUE of μ is $E(\bar{X}|S) = \bar{X}$. Here $E(\bar{X}|S) = \bar{X}$ since $\bar{X} = g(S)$ for some function g.

- Definition. Suppose that (*) holds. Suppose that T is a statistic and the distribution of T does not depend on θ, then T is called an ancillary statistic (輔助統計量).

- Basu's theorem.

  Fact 5. *Suppose that S is an m-dimensional complete sufficient statistic and U is an n-dimensional ancillary statistic. Then S and U are independent.*

  The proof of Basu's theorem is left as a homework problem.

- Example 4. Suppose that $(X_1, \ldots, X_n)$ is a random sample from $U(0, \theta)$, where $\theta > 0$. It can be shown that

  (a) $X_{(n)} = \max(X_1, \ldots, X_n)$ is a complete sufficient statistic for θ and

(b) $2\bar{X}/X_{(n)}$ is an ancillary statistic.

Show that $E(2\bar{X}|X_{(n)}) = c_n X_{(n)}$ for some constant c_n using the results in (a) and (b).

Sol. Note that

$$
\begin{aligned}
E(2\bar{X}|X_{(n)}) &= E\left(\frac{2\bar{X}}{X_{(n)}} \cdot X_{(n)} \middle| X_{(n)} \right) \\
&= X_{(n)} E\left(\frac{2\bar{X}}{X_{(n)}} \middle| X_{(n)} \right) \\
&= X_{(n)} E\left(\frac{2\bar{X}}{X_{(n)}} \right),
\end{aligned}
$$

where the last equality holds since $2\bar{X}/X_{(n)}$ and $X_{(n)}$ are independent, which follows from Basu's theorem. Take $c_n = E(2\bar{X}/X_{(n)})$, then we have $E(2\bar{X}|X_{(n)}) = c_n X_{(n)}$.

Note.

- The constant c_n does not depend on θ since $2\bar{X}/X_{(n)}$ is an ancillary statistic.
- c_n can be solved by $c_n E X_{(n)} = \theta$, which gives $c_n = (n+1)/n$. In the calculation of $EX_{(n)}$, we need a PDF of $X_{(n)}$, which is given by

  $$f_n(t) = \frac{n}{\theta} \left(\frac{t}{\theta} \right)^{n-1} I_{[0,\theta]}(t)$$

  for $t \in (-\infty, \infty)$.
- $c_n X_{(n)} = (n+1)X_{(n)}/n$ is the MVUE of θ by Lehmann and Scheffé theorem.