2.2.3 Minimum Variance Unbiased Estimators

If an unbiased estimator has the variance equal to the CRLB, it must have the minimum variance amongst all unbiased estimators. We call it the **minimum variance unbiased estimator** (MVUE) of ϕ.

Sufficiency is a powerful property in finding unbiased, minimum variance estimators. If $T(\boldsymbol{Y})$ is an unbiased estimator of ϑ and S is a statistic sufficient for ϑ, then there is a function of S that is also an unbiased estimator of ϑ and has no larger variance than the variance of $T(\boldsymbol{Y})$. The following theorem formalizes this statement.

Theorem 2.5. *Rao-Blackwell theorem.*
*Let $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^{\mathrm{T}}$ be a random sample, $\boldsymbol{S} = (S_1, \ldots, S_p)^{\mathrm{T}}$ be jointly sufficient statistics for $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_p)^{\mathrm{T}}$ and $T(\boldsymbol{Y})$ (which is **not** a function of $\boldsymbol{S}$) be an unbiased estimator of $\phi = g(\boldsymbol{\vartheta})$. Then, $U = E(T|\boldsymbol{S})$ is a statistic such that*

(a) $\mathrm{E}(U) = \phi$*, so that U is an unbiased estimator of ϕ, and*

(b) $\mathrm{var}(U) < \mathrm{var}(T)$.

Proof. First, we note that U is a statistic. Indeed, since $\boldsymbol{S}$ are jointly sufficient for $\boldsymbol{\vartheta}$, the conditional distribution $\boldsymbol{Y}|\boldsymbol{S}$ does not depend on the parameters and so the conditional distribution of a function $T(\boldsymbol{Y})$ given $\boldsymbol{S}$, $T|\boldsymbol{S}$, does not depend on $\boldsymbol{\vartheta}$ either. Thus, $U = \mathrm{E}(T|\boldsymbol{S})$ is a function of the random sample only, not a function of $\boldsymbol{\vartheta}$, therefore it is a statistic.

Next, we will use the known facts about the conditional expectation and variance given Exercise 1.15. Since T is an unbiased estimator of ϕ, we have

$$\mathrm{E}(U) = \mathrm{E}[\mathrm{E}(T|\boldsymbol{S})] = \mathrm{E}(T) = \phi.$$

So U is also an unbiased estimator of ϕ, which proves (a). Finally, we get

$$\begin{aligned} \mathrm{var}(T) &= \mathrm{var}[\mathrm{E}(T|\boldsymbol{S})] + \mathrm{E}[\mathrm{var}(T|\boldsymbol{S})] \\ &= \mathrm{var}(U) + \mathrm{E}[\mathrm{var}(T|\boldsymbol{S})]. \end{aligned}$$

However, since T is not a function of $\boldsymbol{S}$ we have $\mathrm{var}(T|\boldsymbol{S}) > 0$, thus, it follows that $\mathrm{E}[\mathrm{var}(T|\boldsymbol{S})] > 0$, and hence (b) is proved. □

It means that, if we have an unbiased estimator, T, of ϕ, which is not a function of the sufficient statistics, we can always find an unbiased estimator which has smaller variance, namely $U = E(T|S_1, \ldots, S_p)$ which is a function of $\boldsymbol{S}$. We thus have the following result.

Corollary 2.1. *MVUEs must be functions of sufficient statistics.*

Example 2.11. Suppose that $Y_1, \ldots, Y_n$ are independent $\text{Poisson}(\lambda)$ random variables. Then $T = Y_i$, for any $i = 1, \ldots, n$, is an unbiased estimator of λ. Also, $S = \sum_{i=1}^n Y_i$ is a sufficient statistic for λ and T is not a function of S.

Hence, a better unbiased estimator is given by any of $\text{E}(Y_1|\sum_{i=1}^n Y_i), \ldots, \text{E}(Y_n|\sum_{i=1}^n Y_i)$.

Now, since $\text{E}(Y_1|\sum_{i=1}^n Y_i) = \ldots = \text{E}(Y_n|\sum_{i=1}^n Y_i)$ and

$$\text{E}(Y_1|\sum_{i=1}^n Y_i) + \text{E}(Y_2|\sum_{i=1}^n Y_i) + \ldots + \text{E}(Y_n|\sum_{i=1}^n Y_i) = \text{E}\left[\sum_{i=1}^n Y_i|\sum_{i=1}^n Y_i\right] = \sum_{i=1}^n Y_i,$$

we have

$$n\,\text{E}(Y_1|\sum_{i=1}^n Y_i) = \sum_{i=1}^n Y_i.$$

Hence $U = \text{E}(Y_1|\sum_{i=1}^n Y_i) = \overline{Y}$ is a better unbiased estimator of λ than Y_1 or than any of Y_i. In fact, $\overline{Y}$ is a MVUE of λ as its variance is equal to the Cramer-Rao Lower Bound for λ. (See Example 2.12)

□

2.2.4 Complete sufficient statistics

$T(\boldsymbol{Y})$ is a function of random sample $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$ and so it is a random variable as well. Hence, we may ask about the distribution of $T(\boldsymbol{Y})$. For example, assume that σ^2 is known equal to σ_o^2. Then, to make inference about μ we may "reduce" the random sample to its mean. We know that $T(\boldsymbol{Y}) = \overline{Y} \sim \mathcal{N}\left(\mu, \frac{\sigma_o^2}{n}\right)$ if $Y_i \underset{iid}{\sim} \mathcal{N}(\mu, \sigma_o^2)$ and we may write

$$f_T(t; \mu|\sigma_o^2) = \frac{1}{\sqrt{2\pi}\sigma_o} e^{n(t-\mu)^2/2\sigma_o^2}.$$

Due to this "data reduction" we can make inference about μ based on the distribution of $\overline{Y}$ only rather than on the multivariate distribution of the whole random sample $\boldsymbol{Y}$.

A minimal sufficient statistic reduces data maximally while retaining all the information about the parameter $\boldsymbol{\vartheta}$. We would also like such a statistic to be independent of any so called *ancillary* functions of the random sample whose distributions do not depend on the parameter of interest. Such an independent statistic is called *complete*.

First, we introduce a notion of a complete family of distributions.

Definition 2.8. *A family of distributions $\mathcal{P} = \{P_{\boldsymbol{\vartheta}} : \boldsymbol{\vartheta} \in \Theta\}$ defined on a common space $\mathcal{Y}$ is called **complete** if for any real measurable function $h(Y)$*

$$\mathrm{E}[h(Y)] = 0 \quad implies\ that\ P_{\boldsymbol{\vartheta}}\big(h(Y) = 0\big) = 1 \quad for\ all\ \boldsymbol{\vartheta} \in \Theta.$$

□

Note: $P_{\boldsymbol{\vartheta}}\big(h(Y) = 0\big) = 1$ can also be written as $P_{\boldsymbol{\vartheta}}\big(h(Y) \neq 0\big) = 0$, which means that function $h(Y)$ may have non-zero values only on a set $\mathcal{B} \subset \mathcal{Y}$ such that $P(Y \in \mathcal{B}) = 0$. Then we say that $h(Y) = 0$ *almost surely in* $\mathcal{Y}$.

Definition 2.9. *A statistic $T(\boldsymbol{Y})$ is called **complete** for the family $\mathcal{P} = \{P_{\boldsymbol{\vartheta}} : \boldsymbol{\vartheta} \in \Theta\}$ on $\mathcal{Y}$ if the family of probability distributions $\mathcal{P}_T = \{P_{\boldsymbol{\vartheta},T} : \boldsymbol{\vartheta} \in \Theta\}$ is complete for all $\boldsymbol{\vartheta}$, that is,*

$$\mathrm{E}[h(T)] = 0 \text{ implies that } P\{h(T) = 0\} = 1.$$

□

Example 2.12. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ be a random sample from a family of $\mathrm{Bernoulli}(p)$ distributions for $0 < p < 1$. We will show that $T(\boldsymbol{Y}) = \sum_{i=1}^n Y_i$ is a complete sufficient statistic for p.

Sufficiency The pmf of each Y_i is $P(Y_i = y_i) = p^{y_i}(1-p)^{1-y_i}$ and the joint pmf for $\boldsymbol{Y}$ can be factorized as follows:

$$\begin{aligned} P(\boldsymbol{Y} = \boldsymbol{y}) &= \prod_{i=1}^n p^{y_i}(1-p)^{1-y_i} \\ &= p^{\sum_{i=1}^n y_i}(1-p)^{n-\sum_{i=1}^n y_i} \times 1 \end{aligned}$$

Hence, $T(\boldsymbol{Y}) = \sum_{i=1}^n Y_i$ is a sufficient statistic for p.

Completeness Now, we know that a sum of independent Bernoulli rvs has a Binomial distribution, i.e,

$$T \sim \text{Bin}(n, p) \text{ for } 0 < p < 1, \ t = 0, 1, \ldots, n.$$

Let $h(T)$ be such that $\text{E}[h(T)] = 0$. Then

$$\begin{aligned} 0 &= \text{E}[h(T)] \\ &= \sum_{t=0}^{n} h(t)\, {}^{n}C_t p^t (1-p)^{n-t} \\ &= (1-p)^n \sum_{t=0}^{n} h(t)\, {}^{n}C_t \left(\frac{p}{1-p}\right)^t. \end{aligned}$$

The factor $(1-p)^n \neq 0$ for any $p \in (0, 1)$. Thus, it must be that

$$\begin{aligned} 0 &= \sum_{t=0}^{n} h(t)\, {}^{n}C_t \left(\frac{p}{1-p}\right)^t \\ &= \sum_{t=0}^{n} h(t)\, {}^{n}C_t r^t \end{aligned}$$

for all $r = \frac{p}{1-p} > 0$. The last expression is a polynomial of degree n in r. For the polynomial to be zero for all r the coefficients $h(t)\, {}^{n}C_t$ must all be zero. It means that $h(t) = 0$ for all $t \in \{0, 1, \ldots, n\}$. Since $P(T = t, t \in \{0, 1, \ldots, n\}) = 1$ it means that $P\{h(T) = 0\} = 1$ for all p. Hence, T is a complete statistic for p.

□

The following theorem gives a connection between complete and minimal sufficient statistics:

Theorem 2.6. *If $T(\boldsymbol{Y})$ is a complete sufficient statistic for a family of distributions with parameter $\boldsymbol{\vartheta}$, then $T(\boldsymbol{Y})$ is a minimal sufficient statistic for the family.* □

Exercise 2.7. Suppose that $Y_1, Y_2, \ldots, Y_n$ is a random sample from a $\text{Poisson}(\lambda)$ distribution. Show that $T(\boldsymbol{Y}) = \sum_{i=1}^{n} Y_i$ is a complete sufficient statistic for λ.

The following Theorem establishes the minimum variance property of complete sufficient statistics.

Theorem 2.7. *Lehmann-Scheffé Theorem.*
Let $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^{\mathrm{T}}$ *be a random sample. If* $\boldsymbol{S}(\boldsymbol{Y})$ *is a jointly complete sufficient statistic and* $T(\boldsymbol{Y})$ *is an unbiased estimator for* $\phi = g(\boldsymbol{\vartheta})$ *then*

$$U = \mathrm{E}[T|\boldsymbol{S}]$$

is, with probability 1, a unique MVUE of ϕ*.*

Proof. First, to prove that U is a $MVUE$ of $g(\boldsymbol{\vartheta})$, we show that whatever unbiased estimator $T(\boldsymbol{Y})$ we take we obtain the same $\mathrm{E}[T|\boldsymbol{S}]$, i.e., the same U. Then, by Rao-Blackwell Theorem, condition (b), U must be MVUE of $g(\boldsymbol{\vartheta})$.

Suppose that $T(\boldsymbol{Y})$ and $T'(\boldsymbol{Y})$ are any two unbiased estimators of $g(\boldsymbol{\vartheta})$. Let

$$\begin{aligned} U &= \mathrm{E}[T|\boldsymbol{S}] \\ U' &= \mathrm{E}[T'|\boldsymbol{S}]. \end{aligned}$$

Then we have

$$\begin{aligned} &\mathrm{E}\{U - U'\} \\ &= \mathrm{E}\{\mathrm{E}[T|\boldsymbol{S}] - \mathrm{E}[T'|\boldsymbol{S}]\} \\ &= \mathrm{E}(T) - \mathrm{E}(T') \\ &= g(\boldsymbol{\vartheta}) - g(\boldsymbol{\vartheta}) \\ &= 0. \end{aligned}$$

Hence, by completeness of $\boldsymbol{S}(\boldsymbol{Y})$ we get

$$P\left[U\big(\boldsymbol{S}(\boldsymbol{Y})\big) = U'\big(\boldsymbol{S}(\boldsymbol{Y})\big)\right] = 1$$

for all $\boldsymbol{\vartheta}$. This proves the first part of the theorem. Now, we will show uniqueness.

Suppose that U and $T^\star$ are two MVUE of $g(\boldsymbol{\vartheta})$. Then if $T^\star$ is a function of the sufficient statistics $\boldsymbol{S}(\boldsymbol{Y})$ then, as shown above, it must be equal to U. If $T^\star$ is not a function of $\boldsymbol{S}(\boldsymbol{Y})$ then $\mathrm{var}(U) < \mathrm{var}(T^\star)$, hence $T^\star$ cannot be a MVUE. Hence, U is a unique MVUE of $g(\boldsymbol{\vartheta})$. □

Note: Lehmann-Scheffé Theorem may be used to construct MVUE of $g(\boldsymbol{\vartheta})$ by two methods. Both are based on complete sufficient statistics $\boldsymbol{S}(\boldsymbol{Y})$.

- Method 1: If we can find a function of $\boldsymbol{S} = \boldsymbol{S}(\boldsymbol{Y})$, say $U(\boldsymbol{S})$ such that $\mathrm{E}[U(\boldsymbol{S})] = g(\boldsymbol{\vartheta})$ then $U(\boldsymbol{S})$ is a unique MVUE of $g(\boldsymbol{\vartheta})$.

- Method 2: If we can find any unbiased estimator $T = T(\boldsymbol{Y})$ of $g(\boldsymbol{\vartheta})$, then $U(\boldsymbol{S}) = \mathrm{E}[T|\boldsymbol{S}]$ is a unique MVUE of $g(\boldsymbol{\vartheta})$.

Example 2.13. Method 1. Let $Y_i \underset{iid}{\sim} \text{Bernoulli}(p)$, $i = 1, \ldots, n$. Earlier we showed that $\sum_{i=1}^n Y_i$ is a complete sufficient statistic for p. Denote it by $S(\boldsymbol{Y})$.

$\overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i = \frac{1}{n}S(\boldsymbol{Y})$ is an unbiased estimator of p, hence, as a function of a complete sufficient statistic, it is the unique MVUE of p.

Now, let $g(p) = \text{var}(Y) = p(1-p)$. The sample variance

$$\frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y})^2$$

is an unbiased estimator of $g(p)$. It is in fact a function of the complete sufficient statistic $S(\boldsymbol{Y}) = \sum_{i=1}^n Y_i$. Hence, it is the unique MVUE of $g(p) = p(1-p)$. □

Exercise 2.8. Suppose that $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ is a random sample from a $\text{Poisson}(\lambda)$ distribution. Find a MVUE of $\phi = \lambda^2$.

2.2.5 Exponential families

There is a class of distributions, including the normal, Poisson, binomial, gamma, chi-squared, exponential and others for which complete sufficient statistics always exist.

Definition 2.10. *A family of probability distributions $\mathcal{P} = \{P_{\boldsymbol{\vartheta}} : \boldsymbol{\vartheta} \in \Theta\}$ is called* ***exponential*** *if for every distribution belonging to the family, its pdf (pmf) can be written in the form*

$$f(y; \boldsymbol{\vartheta}) = h(y)\exp\left\{\sum_{j=1}^p a_j(\boldsymbol{\vartheta})b_j(y) + c(\boldsymbol{\vartheta})\right\}.$$

□

Note: For the one-parameter exponential family, this reduces to

$$f(y; \vartheta) = h(y)\exp\{a(\vartheta)b(y) + c(\vartheta)\}.$$

Example 2.14. Suppose that $Y \sim \text{Bin}(m, p)$, where m is known. Then we may write

$$\begin{aligned} P(Y=y;p) &= \binom{m}{y} p^y (1-p)^{m-y} \\ &= \binom{m}{y} \exp\{y \log p + (m-y)\log(1-p)\} \\ &= \binom{m}{y} \exp\{y \log p - y\log(1-p) + m\log(1-p)\} \\ &= \binom{m}{y} \exp\left\{y \log\left(\frac{p}{1-p}\right) + m\log(1-p)\right\}. \end{aligned}$$

Thus, we have $a(p) = \log\{p/(1-p)\}$, $b(y) = y$, $c(p) = m\log(1-p)$ and $h(y) = \binom{m}{y}$. Hence, $\mathcal{P} = \{P(Y=y;p) : p \in (0,1)\}$ is an exponential family of distributions. □

Exercise 2.9. Show that $\mathcal{P} = \{P(Y=y;\lambda) = \frac{\lambda^y e^{-\lambda}}{y!} I_{\{0,1,2,\ldots\}} : \lambda > 0\}$ is an exponential family of distributions.

Example 2.15. Suppose that $Y \sim \mathcal{N}(\mu, \sigma^2)$. Then we may write

$$\begin{aligned} f(y;\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \\ &= \exp\left\{-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(y-\mu)^2}{2\sigma^2}\right\} \\ &= \exp\left\{-\frac{y^2}{2\sigma^2} + \frac{\mu y}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right\} \\ &= \exp\left\{\frac{\mu}{\sigma^2}y - \frac{1}{2\sigma^2}y^2 - \frac{\mu^2}{2\sigma^2} - \log\sigma - \frac{1}{2}\log(2\pi)\right\}. \\ &= \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{\mu}{\sigma^2}y - \frac{1}{2\sigma^2}y^2 - \frac{\mu^2}{2\sigma^2} - \log\sigma\right\}. \end{aligned}$$

Thus, we have $a_1(\mu,\sigma^2) = \mu/\sigma^2$, $b_1(y) = y$, $a_2(\mu,\sigma^2) = -1/(2\sigma^2)$, $b_2(y) = y^2$, $c(\mu,\sigma^2) = -\mu^2/(2\sigma^2) - \log\sigma$ and $h(y) = \frac{1}{\sqrt{2\pi}}$. Hence, the family of normal distributions, $\mathcal{P} = \{f(y;\mu,\sigma^2) : \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+\}$, is a family of exponential distributions. □

Lemma 2.3. *Let $\boldsymbol{Y} = (Y_1, \dots, Y_n)^{\mathrm{T}}$ be a random sample from a distribution belonging to an exponential family of distributions. Then, there exists a non-trivial jointly sufficient statistic $\boldsymbol{S} = (S_1, \dots, S_p)^{\mathrm{T}}$ for $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_p)^{\mathrm{T}}$ such that*

$$S_j = \sum_{i=1}^{n} b_j(Y_i), \quad j = 1, \dots, p. \tag{2.3}$$

Proof. Note that for members of the exponential family the joint pdf (pmf) of $\boldsymbol{Y}$ can be written as

$$\begin{aligned}
&\prod_{i=1}^{n} \left[h(y_i) \exp\left\{ \sum_{j=1}^{p} a_j(\boldsymbol{\vartheta}) b_j(y_i) + c(\boldsymbol{\vartheta}) \right\} \right] \\
&= \left[\prod_{i=1}^{n} h(y_i) \right] \exp\left\{ \sum_{i=1}^{n} \sum_{j=1}^{p} a_j(\boldsymbol{\vartheta}) b_j(y_i) + nc(\boldsymbol{\vartheta}) \right\} \\
&= \left[\prod_{i=1}^{n} h(y_i) \right] \exp\left\{ \sum_{j=1}^{p} a_j(\boldsymbol{\vartheta}) \big[\sum_{i=1}^{n} b_j(y_i) \big] + nc(\boldsymbol{\vartheta}) \right\}
\end{aligned}$$

Hence, by the Neyman's Factorization Theorem $\boldsymbol{S} = \big(\sum_{i=1}^{n} b_1(Y_i), \dots, \sum_{i=1}^{n} b_p(Y_i) \big)^{\mathrm{T}}$ is a jointly sufficient statistic for $\boldsymbol{\vartheta}$. □

A stronger statement, given here without proof, is following:

Theorem 2.8. *Lehmann's Theorem*
If $\boldsymbol{Y} = (Y_1, \dots, Y_n)^{\mathrm{T}}$ is a random sample from a distribution belonging to an exponential family, then $S_j = \sum_{i=1}^{n} b_j(Y_i)$ for $j = 1, 2, \dots, p$, are the joint complete sufficient statistics for $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_p)^{\mathrm{T}}$. □

From this and from Theorem 2.7 we have the following:

Corollary 2.2. *If a distribution belongs to an exponential family than any function of the jointly complete sufficient statistics $\boldsymbol{S} = (S_1, \dots, S_p)^{\mathrm{T}}$, which is an unbiased estimator of $g(\boldsymbol{\vartheta})$, is the unique MVUE of $\phi = g(\boldsymbol{\vartheta})$.* □

Example 2.16. Suppose that $Y_i \underset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$. Normal distributions belong to the family of exponential distributions, hence from Example 2.15 and Theorem 2.8, it follows that $S_1 = \sum_{i=1}^{n} Y_i$ and $S_2 = \sum_{i=1}^{n} Y_i^2$ are the joint complete sufficient statistics for μ and σ^2. Then, $\overline{Y}$ and $S^2 = \sum_{i=1}^{n}(Y_i - \overline{Y})^2/(n-1)$ are MVUEs of μ and σ^2. □

Exercise 2.10. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ be a random sample from a Gamma distribution, $\mathrm{Gamma}(\lambda, \alpha)$, with the following pdf

$$f(y; \lambda, \alpha) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y}, \quad \text{for } y > 0.$$

(a) Show that the distribution belongs to an exponential family.

(b) Identify the joint complete sufficient statistics for $(\lambda, \alpha)^{\mathrm{T}}$.