# Chapter 3

# Unbiased Estimation

Unbiased or asymptotically unbiased estimation plays an important role in point estimation theory. Unbiasedness of point estimators is defined in §2.3. In this chapter we discuss in detail how to derive unbiased estimators and, more importantly, how to find the best unbiased estimators in various situations. Although an unbiased estimator (even the best unbiased estimator if it exists) is not necessarily better than a slightly biased estimator in terms of their mse's (see Exercise 52 in §2.6), unbiased estimators can be used as "building blocks" for the construction of better estimators. Furthermore, one may give up the exact unbiasedness, but cannot give up asymptotic unbiasedness since it is necessary for consistency (see §2.5). Properties and the construction of asymptotically unbiased estimators are studied in the last part of this chapter.

## 3.1 The UMVUE

Let $X$ be a sample from an unknown population $P \in \mathcal{P}$ and $\vartheta$ be a real-valued parameter related to $P$. Recall that an estimator $T(X)$ of $\vartheta$ is unbiased if $E[T(X)] = \vartheta$ for any $P \in \mathcal{P}$. If there exists an unbiased estimator of $\vartheta$, then $\vartheta$ is called an *estimable* parameter.

**Definition 3.1.** An unbiased estimator $T(X)$ of $\vartheta$ is called the *uniformly minimum variance unbiased estimator* (UMVUE) if and only if $\mathrm{Var}(T(X)) \leq \mathrm{Var}(U(X))$ for any $P \in \mathcal{P}$ and any other unbiased estimator $U(X)$ of $\vartheta$. ∎

Since the mse of any unbiased estimator is its variance, a UMVUE is $\Im$-optimal in mse with $\Im$ being the class of all unbiased estimators. One can similarly define the uniformly minimum risk unbiased estimator in statistical decision theory when we use an arbitrary loss instead of the squared error loss that corresponds to the mse.

### 3.1.1 Sufficient and complete statistics

The derivation of a UMVUE is relatively simple if there exists a sufficient and complete statistic for $P \in \mathcal{P}$.

**Theorem 3.1.** Suppose that there exists a sufficient and complete statistic $T(X)$ for $P \in \mathcal{P}$. If $\vartheta$ is estimable, then there is a unique unbiased estimator of $\vartheta$ that is of the form $h(T)$ with a Borel function $h$. (Two estimators that are equal a.s. $\mathcal{P}$ are treated as one estimator.) Furthermore, $h(T)$ is the unique UMVUE of $\vartheta$. ∎

This theorem is a consequence of Theorem 2.5(ii) (Rao-Blackwell's theorem). One can easily extend this theorem to the case of uniformly minimum risk unbiased estimator under any loss function $L(P, a)$ which is strictly convex in $a$. The uniqueness of the UMVUE follows from the completeness of $T(X)$.

There are two typical ways to derive a UMVUE when a sufficient and complete statistic $T$ is available. The first one is solving for $h$ when the distribution of $T$ is available. The following are two typical examples.

**Example 3.1.** Let $X_1, ..., X_n$ be i.i.d. from the uniform distribution on $(0, \theta)$, $\theta > 0$. Let $\vartheta = g(\theta)$, where $g$ is a differentiable function on $(0, \infty)$. Since the sufficient and complete statistic $X_{(n)}$ has the Lebesgue p.d.f. $n\theta^{-n}x^{n-1}I_{(0,\theta)}(x)$, an unbiased estimator $h(X_{(n)})$ of $\vartheta$ must satisfy

$$\theta^n g(\theta) = n\int_0^\theta h(x)x^{n-1}dx \qquad \text{for all } \theta > 0.$$

Assuming that $h$ is continuous and differentiating both sizes of the previous equation lead to

$$n\theta^{n-1}g(\theta) + \theta^n g'(\theta) = nh(\theta)\theta^{n-1}.$$

Hence, the UMVUE of $\vartheta$ is $h(X_{(n)}) = g(X_{(n)}) + n^{-1}X_{(n)}g'(X_{(n)})$. In particular, if $\vartheta = \theta$, then the UMVUE of $\theta$ is $(1 + n^{-1})X_{(n)}$. ∎

**Example 3.2.** Let $X_1, ..., X_n$ be i.i.d. from the Poisson distribution $P(\theta)$ with an unknown $\theta > 0$. Then $T(X) = \sum_{i=1}^n X_i$ is sufficient and complete for $\theta > 0$ and has the Poisson distribution $P(n\theta)$. Suppose that $\vartheta = g(\theta)$, where $g$ is a smooth function such that $g(x) = \sum_{j=0}^\infty a_j x^j$, $x > 0$. An unbiased estimator $h(T)$ of $\vartheta$ must satisfy

$$\begin{aligned}\sum_{t=0}^{\infty} \frac{h(t)n^t}{t!}\theta^t &= e^{n\theta}g(\theta)\\ &= \sum_{k=0}^{\infty}\frac{n^k}{k!}\theta^k \sum_{j=0}^{\infty} a_j\theta^j\\ &= \sum_{t=0}^{\infty}\left(\sum_{j,k:j+k=t}\frac{n^k a_j}{k!}\right)\theta^t\end{aligned}$$

for any $\theta > 0$. Thus, a comparison of coefficients in front of $\theta^t$ leads to

$$h(t) = \frac{t!}{n^t}\sum_{j,k:j+k=t}\frac{n^k a_j}{k!},$$

i.e., $h(T)$ is the UMVUE of $\vartheta$. In particular, if $\vartheta = \theta^r$ for some fixed integer $r \geq 1$, then $a_r = 1$ and $a_k = 0$ if $k \neq r$ and

$$h(t) = \begin{cases} 0 & t < r \\ \frac{t!}{n^r(t-r)!} & t \geq r. \end{cases} \qquad \blacksquare$$

The second method of deriving a UMVUE when there is a sufficient and complete statistic $T(X)$ is conditioning on $T$, i.e., if $U(X)$ is any unbiased estimator of $\vartheta$, then $E[U(X)|T]$ is the UMVUE of $\vartheta$. To apply this method, we do not need the distribution of $T$, but need to work out the conditional expectation $E[U(X)|T]$. From the uniqueness of the UMVUE, it does not matter which $U(X)$ is used and, thus, we should choose $U(X)$ so as to make the calculation of $E[U(X)|T]$ as easy as possible.

**Example 3.3.** Consider the estimation problem in Example 2.26, where $\vartheta = 1 - F_\theta(t)$ and $F_\theta(x) = (1 - e^{-x/\theta})I_{(0,\infty)}(x)$. Since $\bar{X}$ is sufficient and complete for $\theta > 0$ and $U(X) = 1 - F_n(t)$ is unbiased for $\vartheta$, $T(X) = E[U(X)|\bar{X}] = E[1 - F_n(t)|\bar{X}]$ is the UMVUE of $\vartheta$. Since $X_i$'s are i.i.d.,

$$E[1 - F_n(t)|\bar{X}] = \frac{1}{n}\sum_{i=1}^{n} P(X_i > t|\bar{X}) = P(X_1 > t|\bar{X}).$$

If the conditional distribution of $X_1$ given $\bar{X}$ is available, then we can calculate $P(X_1 > t|\bar{X})$ directly. But the following technique can be applied to avoid the derivation of conditional distributions. By Proposition 1.12(vii),

$$P(X_1 > t|\bar{X} = \bar{x}) = P(X_1/\bar{X} > t/\bar{X}|\bar{X} = \bar{x}) = P(X_1/\bar{X} > t/\bar{x}|\bar{X} = \bar{x}).$$

By Basu's theorem (Theorem 2.4), $X_1/\bar{X}$ and $\bar{X}$ are independent. Hence

$$P(X_1 > t|\bar{X} = \bar{x}) = P(X_1/\bar{X} > t/\bar{x}).$$

To compute this probability, we need the distribution of

$$X_1 \bigg/ \sum_{i=1}^{n} X_i = X_1 \bigg/ \left( X_1 + \sum_{i=2}^{n} X_i \right).$$

Using the transformation technique discussed in §1.3.1 and the fact that $\sum_{i=2}^{n} X_i$ is independent of $X_1$ and has a gamma distribution, we obtain that $X_1/\sum_{i=1}^{n} X_i$ has the Lebesgue p.d.f. $(n-1)(1-x)^{n-2}I_{(0,1)}(x)$. Hence

$$P(X_1 > t|\bar{X} = \bar{x}) = (n-1)\int_{t/(n\bar{x})}^{1} (1-x)^{n-2}dx = \left(1 - \frac{t}{n\bar{x}}\right)^{n-1}$$

and the UMVUE of $\vartheta$ is $T(X) = E[1 - F_n(t)|\bar{X}] = \left(1 - \frac{t}{n\bar{X}}\right)^{n-1}$. ∎

We now show more examples of applying these two methods to find UMVUE's.

**Example 3.4.** Let $X_1, ..., X_n$ be i.i.d. from $N(\mu, \sigma^2)$ with unknown $\mu \in \mathcal{R}$ and $\sigma^2 > 0$. From Example 2.18, $T = (\bar{X}, S^2)$ is sufficient and complete for $\theta = (\mu, \sigma^2)$ and $\bar{X}$ and $(n-1)S^2/\sigma^2$ are independent and have the $N(\mu, \sigma^2/n)$ and chi-square distribution $\chi^2_{n-1}$, respectively. Using the method of solving for $h$ directly, we find that the UMVUE for $\mu$ is $\bar{X}$; the UMVUE of $\mu^2$ is $\bar{X}^2 - S^2/n$; the UMVUE for $\sigma^r$ with $r > 1-n$ is $k_{n-1,r}S^r$, where

$$k_{n,r} = \frac{n^{r/2}\Gamma(n/2)}{2^{r/2}\Gamma\left(\frac{n+r}{2}\right)}$$

(exercise); and the UMVUE of $\mu/\sigma$ is $k_{n-1,-1}\bar{X}/S$, if $n > 2$.

Suppose that $\vartheta$ satisfies $P(X_1 \le \vartheta) = p$ with a fixed $p \in (0,1)$. Let $\Phi$ be the c.d.f. of the standard normal distribution. Then $\vartheta = \mu + \sigma\Phi^{-1}(p)$ and its UMVUE is $\bar{X} + k_{n-1,1}S\Phi^{-1}(p)$.

Let $c$ be a fixed constant and $\vartheta = P(X_1 \le c) = \Phi\left(\frac{c-\mu}{\sigma}\right)$. We can find the UMVUE of $\vartheta$ using the method of conditioning and the technique used in Example 3.3. Since $I_{(-\infty,c)}(X_1)$ is an unbiased estimator of $\vartheta$, the UMVUE of $\vartheta$ is $E[I_{(-\infty,c)}(X_1)|T] = P(X_1 \le c|T)$. By Basu's theorem, the ancillary statistic $Z(X) = (X_1 - \bar{X})/S$ is independent of $T = (\bar{X}, S^2)$. Then

$$\begin{aligned} P\left(X_1 \le c|T = (\bar{x}, s^2)\right) &= P\left( \frac{X_1 - \bar{X}}{S} \le \frac{c - \bar{x}}{s} \middle| T = (\bar{x}, s^2) \right) \\ &= P\left( Z \le \frac{c-\bar{x}}{s} \right). \end{aligned}$$

It can be shown that $Z$ has the Lebesgue p.d.f.

$$f(z) = \frac{\sqrt{n}\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}(n-1)\Gamma\left(\frac{n-2}{2}\right)}\left[1 - \frac{nz^2}{(n-1)^2}\right]^{(n/2)-2} I_{(0,(n-1)/\sqrt{n})}(|z|) \quad (3.1)$$

(exercise). Hence the UMVUE of $\vartheta$ is

$$P(X_1 \leq c|T) = \int_{-(n-1)/\sqrt{n}}^{(c-\bar{X})/S} f(z)dz \quad (3.2)$$

with $f$ given by (3.1).

Suppose that we would like to estimate $\vartheta = \frac{1}{\sigma}\Phi'\left(\frac{c-\mu}{\sigma}\right)$, the Lebesgue p.d.f. of $X_1$ evaluated at a fixed $c$, where $\Phi'$ is the first-order derivative of $\Phi$. By (3.2), the conditional p.d.f. of $X_1$ given $\bar{X} = \bar{x}$ and $S^2 = s^2$ is $s^{-1}f\left(\frac{x-\bar{x}}{s}\right)$. Let $f_T$ be the joint p.d.f. of $T = (\bar{X}, S^2)$. Then

$$\vartheta = \int\int \frac{1}{s} f\left(\frac{c-\bar{x}}{s}\right) f_T(t)dt = E\left[\frac{1}{S} f\left(\frac{c-\bar{X}}{S}\right)\right].$$

Hence the UMVUE of $\vartheta$ is

$$\frac{1}{S} f\left(\frac{c-\bar{X}}{S}\right). \quad \blacksquare$$

**Example 3.5.** Let $X_1, ..., X_n$ be i.i.d. from a power series distribution (see Exercise 13 in §2.6), i.e.,

$$P(X_i = x) = \gamma(x)\theta^x / c(\theta), \qquad x = 0, 1, 2, ...$$

with a known function $\gamma(x) \geq 0$ and an unknown parameter $\theta > 0$. It turns out that the joint distribution of $X = (X_1, ..., X_n)$ is in an exponential family with a sufficient and complete statistic $T(X) = \sum_{i=1}^n X_i$. Furthermore, the distribution of $T$ is also in a power series family, i.e.,

$$P(T = t) = \gamma_n(t)\theta^t / [c(\theta)]^n, \qquad t = 0, 1, 2, ...,$$

where $\gamma_n(t)$ is the coefficient of $\theta^t$ in the power series expansion of $[c(\theta)]^n$ (exercise). This result can help us to find the UMVUE of $\vartheta = g(\theta)$. For example, by comparing both sides of

$$\sum_{t=0}^{\infty} h(t)\gamma_n(t)\theta^t = [c(\theta)]^{n-p}\theta^r,$$

we conclude that the UMVUE of $\theta^r / [c(\theta)]^p$ is

$$h(T) = \begin{cases} 0 & T < r \\ \frac{\gamma_{n-p}(T-r)}{\gamma_n(T)} & T \geq r, \end{cases}$$

where $r$ and $p$ are nonnegative integers. In particular, the case of $p = 1$ produces the UMVUE $\gamma(r)h(T)$ of the probability $P(X_1 = r) = \gamma(r)\theta^r/c(\theta)$ for any nonnegative integer $r$. ∎

**Example 3.6.** Let $X_1, ..., X_n$ be i.i.d. from an unknown population $P$ in a nonparametric family $\mathcal{P}$. We have discussed in §2.2 that in many cases the vector of order statistics, $T = (X_{(1)}, ..., X_{(n)})$, is sufficient and complete for $P \in \mathcal{P}$. Note that an estimator $\varphi(X_1, ..., X_n)$ is a function of $T$ if and only if the function $\varphi$ is symmetric in its $n$ arguments. Hence, if $T$ is sufficient and complete, then a symmetric unbiased estimator of any estimable $\vartheta$ is the UMVUE. For example, $\bar{X}$ is the UMVUE of $\vartheta = EX_1$; $S^2$ is the UMVUE of $\mathrm{Var}(X_1)$; $n^{-1}\sum_{i=1}^n X_i^2 - S^2$ is the UMVUE of $(EX_1)^2$; and $F_n(t)$ is the UMVUE of $P(X_1 \le t)$ for any fixed $t$.

Note that these conclusions are not true if $T$ is *not* sufficient and complete for $P \in \mathcal{P}$. For example, if $\mathcal{P}$ contains all symmetric distributions having Lebesgue p.d.f.'s and finite means, then there is no UMVUE for $\vartheta = EX_1$ (exercise). ∎

More discussions of UMVUE's in nonparametric problems are provided in §3.2.

### 3.1.2 A necessary and sufficient condition

When a complete and sufficient statistic is not available, it is usually very difficult to derive a UMVUE. In some cases, the following result can be applied, if we have enough knowledge about unbiased estimators of 0.

**Theorem 3.2.** Let $\mathcal{U}$ be the set of all unbiased estimators of 0 with finite variances and $T$ be an unbiased estimator of $\vartheta$ with $E(T^2) < \infty$.
(i) A necessary and sufficient condition for $T(X)$ to be a UMVUE of $\vartheta$ is that $E[T(X)U(X)] = 0$ for any $U \in \mathcal{U}$ and any $P \in \mathcal{P}$.
(ii) Suppose that $T = h(\tilde{T})$, where $\tilde{T}$ is a sufficient statistic for $P \in \mathcal{P}$ and $h$ is a Borel function. Let $\mathcal{U}_{\tilde{T}}$ be the subset of $\mathcal{U}$ containing Borel functions of $\tilde{T}$. Then a necessary and sufficient condition for $T$ to be a UMVUE of $\vartheta$ is that $E[T(X)U(X)] = 0$ for any $U \in \mathcal{U}_{\tilde{T}}$ and any $P \in \mathcal{P}$.
**Proof.** (i) Suppose that $T$ is a UMVUE of $\vartheta$. Then $T_c = T + cU$, where $U \in \mathcal{U}$ and $c$ is a fixed constant, is also unbiased for $\vartheta$ and, thus,

$$\mathrm{Var}(T_c) \ge \mathrm{Var}(T) \qquad c \in \mathcal{R},\ P \in \mathcal{P},$$

which is the same as

$$c^2\mathrm{Var}(U) + 2c\mathrm{Cov}(T, U) \ge 0 \qquad c \in \mathcal{R},\ P \in \mathcal{P}.$$

This is impossible unless $\mathrm{Cov}(T, U) = E(TU) = 0$ for any $P \in \mathcal{P}$.

Suppose now $E(TU) = 0$ for any $U \in \mathcal{U}$ and $P \in \mathcal{P}$. Let $T_0$ be another unbiased estimator of $\vartheta$ with $\mathrm{Var}(T_0) < \infty$. Then $T - T_0 \in \mathcal{U}$ and, hence,

$$E[T(T - T_0)] = 0 \qquad P \in \mathcal{P},$$

which with the fact that $ET = ET_0$ implies that

$$\mathrm{Var}(T) = \mathrm{Cov}(T, T_0) \qquad P \in \mathcal{P}.$$

By inequality (1.34), $[\mathrm{Cov}(T, T_0)]^2 \leq \mathrm{Var}(T)\mathrm{Var}(T_0)$. Hence $\mathrm{Var}(T) \leq \mathrm{Var}(T_0)$ for any $P \in \mathcal{P}$.
(ii) It suffices to show that $E(TU) = 0$ for any $U \in \mathcal{U}_{\tilde{T}}$ and $P \in \mathcal{P}$ implies that $E(TU) = 0$ for any $U \in \mathcal{U}$ and $P \in \mathcal{P}$. Let $U \in \mathcal{U}$. Then $E(U|\tilde{T}) \in \mathcal{U}_{\tilde{T}}$ and the result follows from the fact that $T = h(\tilde{T})$ and

$$E(TU) = E[E(TU|\tilde{T})] = E[E(h(\tilde{T})U|\tilde{T})] = E[h(\tilde{T})E(U|\tilde{T})]. \quad \blacksquare$$

Theorem 3.2 can be used to find a UMVUE, to check whether a particular estimator is a UMVUE, and to show the nonexistence of any UMVUE. If there is a sufficient statistic, then by Rao-Blackwell's theorem, we only need to focus on functions of the sufficient statistic and, hence, Theorem 3.2(ii) is more convenient to use.

**Example 3.7.** Let $X_1, ..., X_n$ be i.i.d. from the uniform distribution on the interval $(0, \theta)$. In Example 3.1, $(1 + n^{-1})X_{(n)}$ is shown to be the UMVUE for $\theta$ when the parameter space is $\Theta = (0, \infty)$. Suppose now that $\Theta = [1, \infty)$. Then $X_{(n)}$ is not complete, although it is still sufficient for $\theta$. Thus, Theorem 3.1 does not apply. We now illustrate how to use Theorem 3.2(ii) to find a UMVUE of $\theta$. Let $U(X_{(n)})$ be an unbiased estimator of 0. Since $X_{(n)}$ has the Lebesgue p.d.f. $n\theta^{-n}x^{n-1}I_{(0,\theta)}(x)$,

$$0 = \int_0^1 U(x)x^{n-1}dx + \int_1^\theta U(x)x^{n-1}dx$$

for all $\theta \geq 1$. This implies that $U(x) = 0$ a.e. Lebesgue measure on $[1, \infty)$ and

$$\int_0^1 U(x)x^{n-1}dx = 0.$$

Consider $T = h(X_{(n)})$. To have $E(TU) = 0$, we must have

$$\int_0^1 h(x)U(x)x^{n-1}dx = 0.$$

Thus, we may consider the following function:

$$h(x) = \begin{cases} c & 0 \leq x \leq 1 \\ bx & x > 1, \end{cases}$$

where $c$ and $b$ are some constants. From the previous discussion,

$$E[h(X_{(n)})U(X_{(n)})] = 0, \qquad \theta \geq 1.$$

Since $E[h(X_{(n)})] = \theta$, we obtain that

$$\begin{aligned}\theta &= cP(X_{(n)} \leq 1) + bE[X_{(n)}I_{(1,\infty)}(X_{(n)})]\\ &= c\theta^{-n} + [bn/(n+1)](\theta - \theta^{-n}).\end{aligned}$$

Thus, $c = 1$ and $b = (n+1)/n$. The UMVUE of $\theta$ is then

$$T = \begin{cases} 1 & 0 \leq X_{(n)} \leq 1 \\ (1+n^{-1})X_{(n)} & X_{(n)} > 1. \end{cases}$$

This estimator is better than $(1+n^{-1})X_{(n)}$ which is the UMVUE when $\Theta = (0,\infty)$ and does not make use of the information about $\theta \geq 1$. ∎

**Example 3.8.** Let $X$ be a sample (of size 1) from the uniform distribution $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, $\theta \in \mathcal{R}$. We now apply Theorem 3.2 to show that there is no UMVUE of $\vartheta = g(\theta)$ for any nonconstant continuous $g$. Note that an unbiased estimator $U(X)$ of 0 must satisfy

$$\int_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} U(x)dx = 0 \qquad \theta \in \mathcal{R}$$

and, hence, $U(x) = U(x+1)$ a.e. $m$, where $m$ is the Lebesgue measure on $\mathcal{R}$. If $T$ is a UMVUE, then $T(X)U(X)$ is unbiased for 0 and, hence, $T(x)U(x) = T(x+1)U(x+1)$ a.e. $m$, which implies that $T(x) = T(x+1)$ a.e. $m$. If $T$ is unbiased for $g(\theta)$, then

$$g(\theta) = \int_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} T(x)dx \qquad \theta \in \mathcal{R},$$

which implies that

$$g'(\theta) = T\left(\theta + \tfrac{1}{2}\right) - T\left(\theta - \tfrac{1}{2}\right) = 0 \quad \text{a.e. } m \quad ∎$$

As a consequence of Theorem 3.2, we have the following useful result.

**Corollary 3.1.** (i) Let $T_j$ be a UMVUE of $\vartheta_j$, $j = 1,...,k$, where $k$ is a fixed positive integer. Then $\sum_{j=1}^k c_jT_j$ is a UMVUE of $\vartheta = \sum_{j=1}^k c_j\vartheta_j$ for any constants $c_1,...,c_k$.
(ii) Let $T_1$ and $T_2$ be two UMVUE's of $\vartheta$. Then $T_1 = T_2$ a.s. $P$ for any $P \in \mathcal{P}$. ∎

### 3.1.3 Information inequality

Suppose that we have a lower bound for the variances of all unbiased estimators of $\vartheta$ and that there is an unbiased estimator $T$ of $\vartheta$ whose variance is always the same as the lower bound. Then $T$ is a UMVUE of $\vartheta$. Although this is not an effective way to find UMVUE's (compared with the methods introduced in §3.1.1 and §3.1.2), it provides a way of assessing the performance of UMVUE's. The following result provides such a lower bound in some cases.

**Theorem 3.3** (The Cramér-Rao lower bound). Let $X = (X_1, ..., X_n)$ be a sample from $P \in \mathcal{P} = \{P_\theta : \theta \in \Theta\}$, where $\Theta$ is an open set in $\mathcal{R}^k$. Suppose that $T(X)$ is an estimator with $E[T(X)] = g(\theta)$ being a differentiable function of $\theta$; the joint distribution of $X$ has a p.d.f. $f_\theta$ w.r.t. a measure $\nu$ for all $\theta \in \Theta$; and $f_\theta$ is differentiable as a function of $\theta$ and satisfies

$$\frac{\partial}{\partial \theta} \int h(x) f_\theta(x) d\nu = \int h(x) \frac{\partial}{\partial \theta} f_\theta(x) d\nu, \qquad \theta \in \Theta, \tag{3.3}$$

for $h(x) \equiv 1$ and $h(x) = T(x)$. Then

$$\mathrm{Var}(T(X)) \geq \tfrac{\partial}{\partial \theta} g(\theta) [I(\theta)]^{-1} \left[\tfrac{\partial}{\partial \theta} g(\theta)\right]^\tau, \tag{3.4}$$

where

$$I(\theta) = E\left[\frac{\partial}{\partial \theta} \log f_\theta(X)\right]^\tau \left[\frac{\partial}{\partial \theta} \log f_\theta(X)\right] \tag{3.5}$$

is assumed to be positive definite for any $\theta \in \Theta$.

**Proof.** We prove the univariate case ($k = 1$) only. The proof for the multivariate case ($k > 1$) is left to the reader. When $k = 1$, (3.4) reduces to

$$\mathrm{Var}(T(X)) \geq \frac{[g'(\theta)]^2}{E\left[\frac{\partial}{\partial \theta} \log f_\theta(X)\right]^2}. \tag{3.6}$$

From inequality (1.34), we only need to show that

$$E\left[\frac{\partial}{\partial \theta} \log f_\theta(X)\right]^2 = \mathrm{Var}\left(\frac{\partial}{\partial \theta} \log f_\theta(X)\right)$$

and

$$g'(\theta) = \mathrm{Cov}\left(T(X), \frac{\partial}{\partial \theta} \log f_\theta(X)\right).$$

These two results are consequences of condition (3.3). ∎

The $k \times k$ matrix $I(\theta)$ in (3.5) is called the *Fisher information matrix.* The greater $I(\theta)$ is, the easier it is to distinguish $\theta$ from neighboring values and, therefore, the more accurately $\theta$ can be estimated. In fact, if the equality in (3.6) holds for an unbiased estimator $T(X)$ of $g(\theta)$ (which is then a UMVUE), then the greater $I(\theta)$ is, the smaller $\mathrm{Var}(T(X))$ is. Thus, $I(\theta)$ is the information that $X$ contains about the unknown parameter $\theta$. The inequalities in (3.4) and (3.6) are called *information inequalities*.

The following result is helpful in finding the Fisher information matrix.

**Proposition 3.1.** (i) Let $X$ and $Y$ be independent with the Fisher information matrices $I_X(\theta)$ and $I_Y(\theta)$, respectively. Then the Fisher information about $\theta$ contained in $(X, Y)$ is $I_X(\theta) + I_Y(\theta)$. In particular, if $X_1, ..., X_n$ are i.i.d. and $I(\theta)$ is the Fisher information about $\theta$ contained in a single $X_i$, then the Fisher information about $\theta$ contained in $X_1, ..., X_n$ is $nI(\theta)$.
(ii) Suppose that $X$ has the p.d.f. $f_\theta$ which is twice differentiable in $\theta$ and that (3.3) holds with $h(x) \equiv 1$ and $f_\theta$ replaced by $\partial f_\theta / \partial\theta$. Then

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta\partial\theta^\tau} \log f_\theta(X)\right]. \tag{3.7}$$

**Proof.** Result (i) follows from the independence of $X$ and $Y$ and the definition of the Fisher information. Result (ii) follows from the equality

$$\frac{\partial^2}{\partial\theta\partial\theta^\tau} \log f_\theta(X) = \frac{\frac{\partial^2}{\partial\theta\partial\theta^\tau} f_\theta(X)}{f_\theta(X)} - \left[\frac{\partial}{\partial\theta} \log f_\theta(X)\right]^\tau \left[\frac{\partial}{\partial\theta} \log f_\theta(X)\right]. \quad \blacksquare$$

The following example provides a formula for the Fisher information matrix for many parametric families with a two-dimensional parameter $\theta$.

**Example 3.9.** Let $X_1, ..., X_n$ be i.i.d. with the Lebesgue p.d.f. $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, where $f(x) > 0$ and $f'(x)$ exists for all $x \in \mathcal{R}$, $\mu \in \mathcal{R}$, and $\sigma > 0$ (a location-scale family). Let $\theta = (\mu, \sigma)$. Then the Fisher information about $\theta$ contained in $X_1, ..., X_n$ is (exercise)

$$I(\theta) = \frac{n}{\sigma^2} \begin{pmatrix} \int \frac{[f'(x)]^2}{f(x)} dx & \int x \frac{[f'(x)]^2}{f(x)} dx \\ \int x \frac{[f'(x)]^2}{f(x)} dx & \int \frac{[xf'(x) + f(x)]^2}{f(x)} dx \end{pmatrix}. \quad \blacksquare$$

Note that $I(\theta)$ depends on the particular parameterization. If $\theta = \psi(\eta)$ and $\psi$ is differentiable, then the Fisher information that $X$ contains about $\eta$ is

$$\tfrac{\partial}{\partial\eta}\psi(\eta) I(\psi(\eta)) \left[\tfrac{\partial}{\partial\eta}\psi(\eta)\right]^\tau.$$

However, it is easy to see that the Cramér-Rao lower bound in (3.4) or (3.6) is not affected by any one-to-one reparameterization.

If we use inequality (3.4) or (3.6) to find a UMVUE $T(X)$, then we obtain a formula for $\text{Var}(T(X))$ at the same time. On the other hand, the Cramér-Rao lower bound in (3.4) or (3.6) is typically not sharp. Under some regularity conditions, the Cramér-Rao lower bound is attained if and only if $f_\theta$ is in an exponential family; see Propositions 3.2 and 3.3 and the discussion in Lehmann (1983, p. 123). Some improved information inequalities are available (see, e.g., Lehmann (1983, Sections 2.6 and 2.7)).

**Proposition 3.2.** Suppose that the distribution of $X$ is from an exponential family $\{f_\theta : \theta \in \Theta\}$, i.e., the p.d.f. of $X$ w.r.t. a measure $\nu$ is

$$f_\theta(x) = \exp\big\{T(x)[\eta(\theta)]^\tau - \xi(\theta)\big\}c(x) \tag{3.8}$$

(see §2.1.3), where $\Theta$ is an open subset of $\mathcal{R}^k$.
(i) The regularity condition (3.3) is satisfied for any $h$ with $E|h(X)| < \infty$ and (3.7) holds.
(ii) If $\underline{I}(\eta)$ is the Fisher information matrix for the natural parameter $\eta$, then the variance-covariance matrix $\text{Var}(T) = \underline{I}(\eta)$.
(iii) If $\overline{I}(\vartheta)$ is the Fisher information matrix for the parameter $\vartheta = E[T(X)]$, then $\text{Var}(T) = [\overline{I}(\vartheta)]^{-1}$.
**Proof.** (i) This is a direct consequence of Theorem 2.1.
(ii) From (2.6), the p.d.f. under the natural parameter $\eta$ is

$$\tilde{f}_\eta(x) = \exp\left\{T(x)\eta^\tau - \zeta(\eta)\right\}c(x).$$

From Proposition 1.10 and Theorem 2.1, $E[T(X)] = \frac{\partial}{\partial\eta}\zeta(\eta)$. The result follows from

$$\tfrac{\partial}{\partial\eta}\log\tilde{f}_\eta(x) = T(x) - \tfrac{\partial}{\partial\eta}\zeta(\eta).$$

(iii) Since $\vartheta = E[T(X)] = \frac{\partial}{\partial\eta}\zeta(\eta)$,

$$\underline{I}(\eta) = \tfrac{\partial\vartheta}{\partial\eta}\overline{I}(\vartheta)\tfrac{\partial\vartheta^\tau}{\partial\eta} = \tfrac{\partial^2}{\partial\eta\partial\eta^\tau}\zeta(\eta)\overline{I}(\vartheta)\left[\tfrac{\partial^2}{\partial\eta\partial\eta^\tau}\zeta(\eta)\right]^\tau.$$

By Proposition 1.10, Theorem 2.1, and the result in (ii), $\frac{\partial^2}{\partial\eta\partial\eta^\tau}\zeta(\eta) = \text{Var}(T) = \underline{I}(\eta)$. Hence

$$\overline{I}(\vartheta) = [\underline{I}(\eta)]^{-1}\underline{I}(\eta)[\underline{I}(\eta)]^{-1} = [\underline{I}(\eta)]^{-1} = [\text{Var}(T)]^{-1}. \quad \blacksquare$$

A direct consequence of Proposition 3.2(ii) is that the variance of any linear function of $T$ in (3.8) attains the Cramér-Rao lower bound. The following result gives a necessary condition for $\text{Var}(U(X))$ of an estimator $U(X)$ to attain the Cramér-Rao lower bound.

**Proposition 3.3.** Let $U(X)$ be an estimator of $g(\theta) = E[U(X)]$. Assume that the conditions in Theorem 3.3 hold for $U(x)$ and that $\Theta \subset \mathcal{R}$.
(i) If $\text{Var}(U(X))$ attains the Cramér-Rao lower bound in (3.6), then

$$a(\theta)[U(X) - g(\theta)] = g'(\theta)\frac{\partial}{\partial\theta}\log f_\theta(X) \quad \text{a.s. } f_\theta$$

for some function $a(\theta)$, $\theta \in \Theta$.
(ii) Let $f_\theta$ and $T$ be given by (3.8). If $\text{Var}(U(X))$ attains the Cramér-Rao lower bound, then $U(X)$ is a linear function of $T(X)$ a.s. $f_\theta$, $\theta \in \Theta$. ∎

**Example 3.10.** Let $X_1, ..., X_n$ be i.i.d. from the $N(\mu, \sigma^2)$ distribution with an unknown $\mu \in \mathcal{R}$ and a known $\sigma^2$. Let $f_\mu$ be the joint distribution of $X = (X_1, ..., X_n)$. Then

$$\tfrac{\partial}{\partial\mu}\log f_\mu(X) = \sum_{i=1}^n (X_i - \mu)/\sigma^2.$$

Thus, $I(\mu) = n/\sigma^2$. It is obvious that $\text{Var}(\bar{X})$ attains the Cramér-Rao lower bound in (3.6). Consider now the estimation of $\vartheta = \mu^2$. Since $E\bar{X}^2 = \mu^2 + \sigma^2/n$, the UMVUE of $\vartheta$ is $h(\bar{X}) = \bar{X}^2 - \sigma^2/n$. A straightforward calculation shows that

$$\text{Var}(h(\bar{X})) = \frac{4\mu^2\sigma^2}{n} + \frac{2\sigma^4}{n^2}.$$

On the other hand, the Cramér-Rao lower bound in this case is $4\mu^2\sigma^2/n$. Hence $\text{Var}(h(\bar{X}))$ does not attain the Cramér-Rao lower bound. The difference is $2\sigma^4/n^2$. ∎

Condition (3.3) is a key regularity condition for the results in Theorem 3.3 and Proposition 3.3. If $f_\theta$ is not in an exponential family, then (3.3) has to be checked. Typically, it does not hold if the set $\{x : f_\theta(x) > 0\}$ depends on $\theta$ (Exercise 32). More discussions can be found in Pitman (1979).

### 3.1.4 Asymptotic properties of UMVUE's

UMVUE's are typically consistent (see Exercise 88 in §2.6). If there is an unbiased estimator of $\vartheta$ whose mse is of the order $a_n^{-2}$, where $\{a_n\}$ is a sequence of positive numbers diverging to $\infty$, then the UMVUE of $\vartheta$ (if it exists) has a mse of order $a_n^{-2}$ and is $a_n$-consistent. For instance, in Example 3.3, the mse of $U(X) = 1 - F_n(t)$ is $F_\theta(t)[1 - F_\theta(t)]/n$; hence the UMVUE $T(X)$ is $\sqrt{n}$-consistent and its mse is of the order $n^{-1}$.

UMVUE's are exactly unbiased so that there is no need to discuss their asymptotic biases. Their variances (or mse's) are finite, but amse's can be used to approximate their mse's if the exact forms of these mse's are difficult to obtain. In many cases, although the variance of a UMVUE $T_n$ does not attain the Cramér-Rao lower bound, the limit of the ratio of the amse (or mse) of $T_n$ over the Cramér-Rao lower bound (if it is not 0) is 1. For instance, in Example 3.10,

$$\frac{\mathrm{Var}(\bar{X}^2 - \sigma^2/n)}{\text{the Cramér-Rao lower bound}} = 1 + \frac{\sigma^2}{2\mu^2 n} \to 1$$

if $\mu \neq 0$. In general, under the conditions in Theorem 3.3, if $T_n(X)$ is unbiased for $g(\theta)$ and if for any $\theta \in \Theta$,

$$T_n(X) - g(\theta) = \tfrac{\partial}{\partial\theta} g(\theta)[I(\theta)]^{-1} \left[\tfrac{\partial}{\partial\theta} \log f_\theta(X)\right]^\tau [1 + o_p(1)] \quad \text{a.s. } f_\theta, \tag{3.9}$$

then

$$\mathrm{amse}_{T_n}(\theta) = \text{the Cramér-Rao lower bound} \tag{3.10}$$

whenever the Cramér-Rao lower bound is not 0. Note that the case of zero Cramér-Rao lower bound is not of interest since a zero lower bound does not provide any information on the performance of estimators.

Consider the UMVUE $T_n = \left(1 - \frac{t}{n\bar{X}}\right)^{n-1}$ of $e^{-t/\theta}$ in Example 3.3. Using the fact that

$$\log(1 - x) = -\sum_{j=1}^{\infty} \frac{x^j}{j}, \qquad |x| \leq 1,$$

we obtain that

$$T_n - e^{-t/\bar{X}} = O_p\left(n^{-1}\right).$$

Using Taylor's expansion, we obtain that

$$e^{-t/\bar{X}} - e^{-t/\theta} = g'(\theta)(\bar{X} - \theta)[1 + o_p(1)],$$

where $g(\theta) = e^{-t/\theta}$. On the other hand,

$$[I(\theta)]^{-1} \tfrac{\partial}{\partial\theta} \log f_\theta(X) = \bar{X} - \theta.$$

Hence (3.9) and (3.10) hold. Note that the exact variance of $T_n$ is not easy to obtain. In this example, it can be shown that $\{n[T_n - g(\theta)]^2\}$ is uniformly integrable and, therefore,

$$\begin{aligned} \lim_{n\to\infty} n\mathrm{Var}(T_n) &= \lim_{n\to\infty} n[\underline{\mathrm{amse}}_{T_n}(\theta)] \\ &= \lim_{n\to\infty} n[g'(\theta)]^2[I(\theta)]^{-1} \\ &= \frac{t^2 e^{-2t/\theta}}{\theta^2}. \end{aligned}$$

It is shown in Chapter 4 that if (3.10) holds, then $T_n$ is asymptotically optimal in some sense. Hence UMVUE's satisfying (3.9), which is often true, are asymptotically optimal, although they may be improved in terms of the exact mse's.

## 3.2 U-Statistics

Let $X_1, ..., X_n$ be i.i.d. from an unknown population $P$ in a nonparametric family $\mathcal{P}$. In Example 3.6 we argued that if the vector of order statistic is sufficient and complete for $P \in \mathcal{P}$, then a symmetric unbiased estimator of any estimable $\vartheta$ is the UMVUE of $\vartheta$. In a large class of problems parameters to be estimated are of the form

$$\vartheta = E[h(X_1, ..., X_m)]$$

with a positive integer $m$ and a Borel function $h$ which is symmetric and satisfies $E|h(X_1, ..., X_m)| < \infty$ for any $P \in \mathcal{P}$. It is easy to see that a symmetric unbiased estimator of $\vartheta$ is

$$U_n = \binom{n}{m}^{-1} \sum_c h(X_{i_1}, ..., X_{i_m}), \tag{3.11}$$

where $\sum_c$ denotes the summation over the $\binom{n}{m}$ combinations of $m$ distinct elements $\{i_1, ..., i_m\}$ from $\{1, ..., n\}$.

**Definition 3.2.** The statistic $U_n$ in (3.11) is called a *U-statistic* with kernel $h$ of order $m$. ∎

### 3.2.1 Some examples

The use of U-statistics is an effective way of obtaining unbiased estimators. In nonparametric problems, U-statistics are often UMVUE's, whereas in parametric problems, U-statistics can be used as initial estimators to derive more efficient estimators.

If $m = 1$, $U_n$ in (3.11) is simply a type of sample mean. Examples include the empirical c.d.f. (2.31) evaluated at a particular $t$ and the *sample moments* $n^{-1} \sum_{i=1}^n X_i^k$ for a positive integer $k$. We now consider some examples with $m > 1$.

Consider the estimation of $\vartheta = \mu^m$, where $\mu = EX_1$ and $m$ is a positive integer. Using $h(x_1, ..., x_m) = x_1 \cdots x_m$, we obtain the following U-statistic unbiased for $\vartheta = \mu^m$:

$$U_n = \binom{n}{m}^{-1} \sum_c X_{i_1} \cdots X_{i_m}. \tag{3.12}$$

Consider next the estimation of $\vartheta = \sigma^2 = \mathrm{Var}(X_1)$. Since

$$\sigma^2 = [\mathrm{Var}(X_1) + \mathrm{Var}(X_2)]/2 = E[(X_1 - X_2)^2/2],$$

we obtain the following U-statistic with kernel $h(x_1, x_2) = (x_1 - x_2)^2/2$:

$$U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \frac{(X_i - X_j)^2}{2} = \frac{1}{n-1} \left( \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right) = S^2,$$

which is the sample variance in (2.2).

In some cases we would like to estimate $\vartheta = E|X_1 - X_2|$, a measure of concentration. Using kernel $h(x_1, x_2) = |x_1 - x_2|$, we obtain the following U-statistic unbiased for $\vartheta = E|X_1 - X_2|$:

$$U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} |X_i - X_j|,$$

which is known as *Gini's mean difference.*

Let $\vartheta = P(X_1 + X_2 \le 0)$. Using kernel $h(x_1, x_2) = I_{(-\infty,0)}(x_1 + x_2)$, we obtain the following U-statistic unbiased for $\vartheta$:

$$U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} I_{(-\infty,0)}(X_i + X_j),$$

which is known as the *one-sample Wilcoxon statistic.*

Let $T_n = T_n(X_1, ..., X_n)$ be a given statistic and let $r$ and $d$ be two positive integers such that $r + d = n$. For any $\boldsymbol{s} = \{i_1, ..., i_r\} \subset \{1, ..., n\}$, define

$$T_{r,\boldsymbol{s}} = T_r(X_{i_1}, ..., X_{i_r}),$$

which is the statistic $T_n$ computed after $X_i$, $i \notin \boldsymbol{s}$, are deleted from the original sample. Let

$$U_n = \binom{n}{r}^{-1} \sum_{c} \tfrac{r}{d}(T_{r,\boldsymbol{s}} - T_n)^2. \tag{3.13}$$

Then $U_n$ is a U-statistic with kernel

$$h_n(x_1, ..., x_r) = \tfrac{r}{d}[T_r(x_1, ..., x_r) - T_n(x_1, ..., x_n)].$$

Unlike the kernels in the previous examples, the kernel in this example depends on $n$. The order of the kernel, $r$, may also depend on $n$. The statistic $U_n$ in (3.13) is known as the *delete-d jackknife variance estimator* for $T_n$ (see, e.g., Shao and Tu (1995)), since it is often true that

$$E[h_n(X_1, ..., X_r)] \approx \mathrm{Var}(T_n).$$

It can be shown that if $T_n = \bar{X}$, then $nU_n$ in (3.13) is exactly the same as the sample variance $S^2$ (exercise).

### 3.2.2 Variances of U-statistics

If $E[h(X_1,...,X_m)]^2 < \infty$, then the variance of $U_n$ in (3.11) with kernel $h$ has an explicit form. To derive $\mathrm{Var}(U_n)$, we need some notation. For $k = 1,...,m$, let

$$\begin{aligned} h_k(x_1,...,x_k) &= E[h(X_1,...,X_m)|X_1 = x_1,...,X_k = x_k] \\ &= E[h(x_1,...,x_k,X_{k+1},...,X_m)]. \end{aligned}$$

It can be shown that

$$h_k(x_1,...,x_k) = E[h_{k+1}(x_1,...,x_k,X_{k+1})]. \tag{3.14}$$

Define

$$\tilde{h}_k = h_k - E[h(X_1,...,X_m)]. \tag{3.15}$$

Then, for any $U_n$ defined by (3.11),

$$U_n - E(U_n) = \binom{n}{m}^{-1} \sum_c \tilde{h}_m(X_{i_1},...,X_{i_m}). \tag{3.16}$$

**Theorem 3.4** (Hoeffding's theorem). For a U-statistic $U_n$ given by (3.11) with $E[h(X_1,...,X_m)]^2 < \infty$,

$$\mathrm{Var}(U_n) = \binom{n}{m}^{-1} \sum_{k=1}^{m} \binom{m}{k}\binom{n-m}{m-k}\zeta_k,$$

where

$$\zeta_k = \mathrm{Var}(h_k(X_1,...,X_k)).$$

**Proof.** Consider two sets $\{i_1,...,i_m\}$ and $\{j_1,...,j_m\}$ of $m$ distinct integers from $\{1,...,n\}$ with exactly $k$ integers in common. The number of distinct choices of two such sets is $\binom{n}{m}\binom{m}{k}\binom{n-m}{m-k}$. By the symmetry of $\tilde{h}_m$ and independence of $X_1,...,X_n$,

$$E[\tilde{h}_m(X_{i_1},...,X_{i_m})\tilde{h}_m(X_{j_1},...,X_{j_m})] = \zeta_k \tag{3.17}$$

for $k = 1,...,m$ (exercise). Then, by (3.16),

$$\begin{aligned} \mathrm{Var}(U_n) &= \binom{n}{m}^{-2} \sum_c \sum_c E[\tilde{h}_m(X_{i_1},...,X_{i_m})\tilde{h}_m(X_{j_1},...,X_{j_m})] \\ &= \binom{n}{m}^{-2} \sum_{k=1}^{m} \binom{n}{m}\binom{m}{k}\binom{n-m}{m-k}\zeta_k. \end{aligned}$$

This proves the result. ∎

**Corollary 3.2.** Under the condition of Theorem 3.4,
(i) $\frac{m^2}{n}\zeta_1 \le \mathrm{Var}(U_n) \le \frac{m}{n}\zeta_m$;
(ii) $(n+1)\mathrm{Var}(U_{n+1}) \le n\mathrm{Var}(U_n)$ for any $n > m$;
(iii) For any fixed $m$ and $k = 1, ..., m$, if $\zeta_j = 0$ for $j < k$ and $\zeta_k > 0$, then

$$\mathrm{Var}(U_n) = \frac{k!\binom{m}{k}^2\zeta_k}{n^k} + O\left(\frac{1}{n^{k+1}}\right). \quad \blacksquare$$

It follows from Corollary 3.2 that a U-statistic $U_n$ as an estimator of its mean is consistent in mse (under the finite second moment assumption on $h$). In fact, for any fixed $m$, if $\zeta_j = 0$ for $j < k$ and $\zeta_k > 0$, then the mse of $U_n$ is of the order $n^{-k}$ and, therefore, $U_n$ is $n^{k/2}$-consistent.

**Example 3.11.** Consider first $h(x_1, x_2) = x_1x_2$ which leads to a U-statistic unbiased for $\mu^2$, $\mu = EX_1$. Note that $h_1(x_1) = \mu x_1$, $\tilde{h}_1(x_1) = \mu(x_1 - \mu)$, $\zeta_1 = E[\tilde{h}_1(X_1)]^2 = \mu^2\mathrm{Var}(X_1) = \mu^2\sigma^2$, $\tilde{h}_2(x_1, x_2) = x_1x_2 - \mu^2$, and $\zeta_2 = \mathrm{Var}(X_1X_2) = E(X_1X_2)^2 - \mu^4 = (\mu^2 + \sigma^2)^2 - \mu^4$. By Theorem 3.4, for $U_n = \binom{n}{2}^{-1}\sum_{1\le i<j\le n} X_iX_j$,

$$\begin{aligned}\mathrm{Var}(U_n) &= \binom{n}{2}^{-1}\left[\binom{2}{1}\binom{n-2}{1}\zeta_1 + \binom{2}{2}\binom{n-2}{0}\zeta_2\right] \\ &= \frac{2}{n(n-1)}\left[2(n-2)\mu^2\sigma^2 + (\mu^2+\sigma^2)^2 - \mu^4\right] \\ &= \frac{4\mu^2\sigma^2}{n} + \frac{2\sigma^4}{n(n-1)}.\end{aligned}$$

Comparing $U_n$ with $\bar{X}^2 - \sigma^2/n$ in Example 3.10, which is the UMVUE under the normality and known $\sigma^2$ assumption, we find that

$$\mathrm{Var}(U_n) - \mathrm{Var}(\bar{X}^2 - \sigma^2/n) = \frac{2\sigma^4}{n^2(n-1)}.$$

Next, consider $h(x_1, x_2) = I_{(-\infty,0)}(x_1 + x_2)$ which leads to the one-sample Wilcoxon statistic. Note that $h_1(x_1) = P(x_1 + X_2 \le 0) = F(-x_1)$, where $F$ is the c.d.f. of $P$. Then $\zeta_1 = \mathrm{Var}(F(-X_1))$. Let $\vartheta = E[h(X_1, X_2)]$. Then $\zeta_2 = \mathrm{Var}(h(X_1, X_2)) = \vartheta(1-\vartheta)$. Hence, for $U_n$ being the one-sample Wilcoxon statistic,

$$\mathrm{Var}(U_n) = \frac{2}{n(n-1)}\left[2(n-2)\zeta_1 + \vartheta(1-\vartheta)\right].$$

If $F$ is continuous and symmetric about 0, then $\zeta_1$ can be simplified as

$$\zeta_1 = \mathrm{Var}(F(-X_1)) = \mathrm{Var}(1 - F(X_1)) = \mathrm{Var}(F(X_1)) = \tfrac{1}{12},$$

since $F(X_1)$ has the uniform distribution on $[0,1]$.

Finally, consider $h(x_1,x_2) = |x_1 - x_2|$, which leads to Gini's mean difference. Note that

$$h_1(x_1) = E|x_1 - X_2| = \int |x_1 - y| dP(y),$$

and

$$\zeta_1 = \mathrm{Var}(h_1(X_1)) = \int \left[\int |x - y| dP(y)\right]^2 dP(x) - \vartheta^2,$$

where $\vartheta = E|X_1 - X_2|$. ∎

### 3.2.3 The projection method

Since $\mathcal{P}$ is nonparametric, the exact distribution of any U-statistic is hard to derive. In this section we study asymptotic distributions of U-statistics, using the method of *projection*.

**Definition 3.3.** Let $T_n$ be a given statistic based on $X_1, ..., X_n$. The projection of $T_n$ on $k_n$ random elements $Y_1, ..., Y_{k_n}$ is defined to be

$$\check{T}_n = E(T_n) + \sum_{i=1}^{k_n} [E(T_n|Y_i) - E(T_n)]. \quad ∎$$

Let $\psi_n(X_i) = E(T_n|X_i)$. If $T_n$ is symmetric (as a function of $X_1, ..., X_n$), then $\psi_n(X_1), ..., \psi_n(X_n)$ are i.i.d. with mean $E[\psi_n(X_i)] = E[E(T_n|X_i)] = E(T_n)$. If $E(T_n^2) < \infty$ and $\mathrm{Var}(\psi_n(X_i)) > 0$, then

$$\frac{1}{\sqrt{n\mathrm{Var}(\psi_n(X_1))}} \sum_{i=1}^{n} [\psi_n(X_i) - E(T_n)] \to_d N(0,1) \tag{3.18}$$

by the CLT. Let $\check{T}_n$ be the projection of $T_n$ on $X_1, ..., X_n$. Then

$$T_n - \check{T}_n = T_n - E(T_n) - \sum_{i=1}^{n} [\psi_n(X_i) - E(T_n)]. \tag{3.19}$$

If we can show that $T_n - \check{T}_n$ has a negligible order of magnitude, then we can derive the asymptotic distribution of $T_n$ by using (3.18)-(3.19) and Slutsky's theorem. The order of magnitude of $T_n - \check{T}_n$ can be obtained with the help of the following lemma.

**Lemma 3.1.** Let $T_n$ be a symmetric statistic with $\mathrm{Var}(T_n) < \infty$ for every $n$ and $\check{T}_n$ be the projection of $T_n$ on $X_1, ..., X_n$. Then $E(T_n) = E(\check{T}_n)$ and

$$E(T_n - \check{T}_n)^2 = \mathrm{Var}(T_n) - \mathrm{Var}(\check{T}_n).$$

**Proof.** Since $E(T_n) = E(\check{T}_n)$,

$$E(T_n - \check{T}_n)^2 = \mathrm{Var}(T_n) + \mathrm{Var}(\check{T}_n) - 2\mathrm{Cov}(T_n, \check{T}_n).$$

From Definition 3.3 with $Y_i = X_i$,

$$\mathrm{Var}(\check{T}_n) = n\mathrm{Var}(E(T_n|X_i)).$$

The result follows from

$$\begin{aligned}
\mathrm{Cov}(T_n, \check{T}_n) &= E(T_n\check{T}_n) - [E(T_n)]^2 \\
&= nE[T_nE(T_n|X_i)] - n[E(T_n)]^2 \\
&= nE\{E[T_nE(T_n|X_i)|X_i]\} - n[E(T_n)]^2 \\
&= nE\{[E(T_n|X_i)]^2\} - n[E(T_n)]^2 \\
&= n\mathrm{Var}(E(T_n|X_i)) \\
&= \mathrm{Var}(\check{T}_n). \quad \blacksquare
\end{aligned}$$

This method of deriving the asymptotic distribution of $T_n$ is known as the method of projection and is particularly effective for U-statistics. For a U-statistic $U_n$ given by (3.11), one can show (exercise) that

$$\check{U}_n = E(U_n) + \frac{m}{n}\sum_{i=1}^{n} \tilde{h}_1(X_i), \tag{3.20}$$

where $\check{U}_n$ is the projection of $U_n$ on $X_1, ..., X_n$ and $\tilde{h}_1$ is defined by (3.15). Hence

$$\mathrm{Var}(\check{U}_n) = m^2\zeta_1/n$$

and, by Corollary 3.2 and Lemma 3.1,

$$E(U_n - \check{U}_n)^2 = O(n^{-2}).$$

If $\zeta_1 > 0$, then (3.18) holds with $\psi_n(X_i) = mh_1(X_i)$, which leads to the result in Theorem 3.5(i) stated later.

If $\zeta_1 = 0$, then $\tilde{h}_1 \equiv 0$ and we have to use another projection of $U_n$. Suppose that $\zeta_1 = \cdots = \zeta_{k-1} = 0$ and $\zeta_k > 0$ for an integer $k > 1$. Consider the projection $\check{U}_{kn}$ of $U_n$ on $\{X_{i_1}, ..., X_{i_k}\}$, $1 \le i_1 < \cdots < i_k \le n$. We can establish a result similar to that in Lemma 3.1 (exercise) and show that

$$E(U_n - \check{U}_n)^2 = O(n^{-(k+1)}).$$

Also, see Serfling (1980, §5.3.4).

With these results, we obtain the following theorem.

**Theorem 3.5.** Let $U_n$ be given by (3.11) with $E[h(X_1, ..., X_m)]^2 < \infty$.
(i) If $\zeta_1 > 0$, then

$$\sqrt{n}[U_n - E(U_n)] \to_d N(0, m^2\zeta_1).$$

(ii) If $\zeta_1 = 0$ but $\zeta_2 > 0$, then

$$n[U_n - E(U_n)] \to_d \frac{m(m-1)}{2} \sum_{j=1}^{\infty} \lambda_j(\chi_{1j}^2 - 1), \tag{3.21}$$

where $\chi_{1j}^2$'s are i.i.d. random variables having the chi-square distribution $\chi_1^2$ and $\lambda_j$'s are some constants (which may depend on $P$) satisfying $\sum_{j=1}^{\infty} \lambda_j^2 = \zeta_2$. ∎

We have actually proved Theorem 3.5(i). A proof for Theorem 3.5(ii) is given in Serfling (1980, §5.5.2). One may derive results for the cases where $\zeta_2 = 0$, but the case of either $\zeta_1 > 0$ or $\zeta_2 > 0$ is the most interesting case in applications.

If $\zeta_1 > 0$, it follows from Theorem 3.5(i) and Corollary 3.2(iii) that $\underline{\text{amse}}_{U_n}(P) = \text{Var}(U_n) = m^2\zeta_1/n$. By Theorem 1.8(vii), $\{n[U_n - E(U_n)]^2\}$ is uniformly integrable.

If $\zeta_1 = 0$ but $\zeta_2 > 0$, it follows from Theorem 3.5(ii) that $\underline{\text{amse}}_{U_n}(P) = EY^2/n^2$, where $Y$ denotes the random variable on the right-hand side of (3.21). The following result provides the value of $EY^2$.

**Lemma 3.2.** Let $Y$ be the random variable on the right-hand side of (3.21). Then $EY^2 = \frac{m^2(m-1)^2}{2}\zeta_2$.
**Proof.** Define

$$Y_k = \frac{m(m-1)}{2} \sum_{j=1}^{k} \lambda_j(\chi_{1j}^2 - 1), \quad k = 1, 2, ...$$

It can be shown (exercise) that $\{Y_k^2\}$ is uniformly integrable. Since $Y_k \to_d Y$ as $k \to \infty$, $\lim_{k\to\infty} EY_k^2 = EY^2$ (Theorem 1.8(vii)). Since $\chi_{1j}^2$'s are independent chi-square random variables with $E\chi_{1j}^2 = 1$ and $\text{Var}(\chi_{1j}^2) = 2$, $EY_k = 0$ for any $k$ and

$$\begin{aligned} EY_k^2 &= \frac{m^2(m-1)^2}{4} \sum_{j=1}^{k} \lambda_j^2 \text{Var}(\chi_{1j}^2) \\ &= \frac{m^2(m-1)^2}{4} \left(2\sum_{j=1}^{k} \lambda_j^2\right) \\ &\to \frac{m^2(m-1)^2}{2}\zeta_2. \end{aligned}$$ ∎

It follows from Corollary 3.2(iii) and Lemma 3.2 that $\underline{\text{amse}}_{U_n}(P) = \text{Var}(U_n) = \frac{m^2(m-1)^2}{2}\zeta_2/n^2$ if $\zeta_1 = 0$. Again, by Theorem 1.8(vii), the sequence $\{n^2[U_n - E(U_n)]^2\}$ is uniformly integrable.

We now apply Theorem 3.5 to the U-statistics in Example 3.11. For $U_n = \frac{2}{n(n-1)}\sum_{1\le i<j\le n} X_iX_j$, $\zeta_1 = \mu^2\sigma^2$. Thus, if $\mu \neq 0$, the result in Theorem 3.5(i) holds with $\zeta_1 = \mu^2\sigma^2$. If $\mu = 0$, then $\zeta_1 = 0$, $\zeta_2 = \sigma^4 > 0$, and Theorem 3.5(ii) applies. However, it is not convenient to use Theorem 3.5(ii) to find the limiting distribution of $U_n$. We may derive this limiting distribution using the following technique which is further discussed in §3.5. By the CLT and Theorem 1.10,

$$n\bar{X}^2/\sigma^2 \to_d \chi_1^2$$

when $\mu = 0$, where $\chi_1^2$ is a random variable having the chi-square distribution $\chi_1^2$. Note that

$$\frac{n\bar{X}^2}{\sigma^2} = \frac{1}{\sigma^2 n}\sum_{i=1}^n X_i^2 + \frac{(n-1)U_n}{\sigma^2}.$$

By the SLLN, $\frac{1}{\sigma^2 n}\sum_{i=1}^n X_i^2 \to_{a.s.} 1$. An application of Slutsky's theorem leads to

$$nU_n/\sigma^2 \to_d \chi_1^2 - 1.$$

Since $\mu = 0$, this implies that the right-hand side of (3.21) is $\sigma^2(\chi_1^2 - 1)$, i.e., $\lambda_1 = \sigma^2$ and $\lambda_j = 0$ when $j > 1$.

For the one-sample Wilcoxon statistic, $\zeta_1 = \text{Var}(F(-X_1)) > 0$ unless $F$ is degenerate. Similarly, for Gini's mean difference, $\zeta_1 > 0$ unless $F$ is degenerate. Hence Theorem 3.5(i) applies to these two cases.

Theorem 3.5 does not apply to $U_n$ defined by (3.13), if $r$, the order of the kernel, depends on $n$ and diverges to $\infty$ as $n \to \infty$. We consider the simple case where

$$T_n = \frac{1}{n}\sum_{i=1}^n \psi(X_i) + R_n \tag{3.22}$$

for some $R_n$ satisfying $E(R_n^2) = o(n^{-1})$. Note that (3.22) is satisfied for $T_n$ being a U-statistic (exercise). Assume that $r/d$ is bounded. Let $S_\psi^2 = (n-1)^{-1}\sum_{i=1}^n[\psi(X_i) - n^{-1}\sum_{i=1}^n \psi(X_i)]^2$. Then

$$nU_n = S_\psi^2 + o_p(1) \tag{3.23}$$

(exercise). Under (3.22), if $0 < E[\psi(X_i)]^2 < \infty$, then $\underline{\text{amse}}_{T_n}(P) = E[\psi(X_i)]^2/n$. Hence, the jackknife estimator $U_n$ in (3.13) provides a consistent estimator of $\underline{\text{amse}}_{T_n}(P)$, i.e., $U_n/\underline{\text{amse}}_{T_n}(P) \to_p 1$.

## 3.3 The LSE in Linear Models

One of the most useful statistical models for non-i.i.d. data in applications is the following general linear model

$$X_i = Z_i\beta^\tau + \varepsilon_i, \qquad i = 1, ..., n, \tag{3.24}$$

where $X_i$ is the $i$th observation and is often called the $i$th response; $\beta$ is a $p$-vector of unknown parameters, $p < n$; $Z_i$ is the $i$th value of a $p$-vector of explanatory variables (or covariates); and $\varepsilon_1, ..., \varepsilon_n$ are random errors. Our data in this case are $(X_1, Z_1), ..., (X_n, Z_n)$ ($\varepsilon_i$'s are not observed). Throughout this book $Z_i$'s are considered to be nonrandom or given values of a random $p$-vector, in which case our analysis is conditioned on $Z_1, ..., Z_n$. Each $\varepsilon_i$ can be viewed as a random measurement error in measuring the unknown mean of $X_i$ when the covariate vector is equal to $Z_i$. The main parameter of interest is $\beta$. More specific examples of model (3.24) are provided in this section. Other examples and examples of data from model (3.24) can be found in many standard books for linear models, for example, Draper and Smith (1981) and Searle (1971).

### 3.3.1 The LSE and estimability

Let $X = (X_1, ..., X_n)$, $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)$, and $Z$ be the $n \times p$ matrix whose $i$th row is $Z_i$, $i = 1, ..., n$. Then a matrix form of model (3.24) is

$$X = \beta Z^\tau + \varepsilon. \tag{3.25}$$

**Definition 3.4.** Suppose that the range of $\beta$ in model (3.25) is $B \subset \mathcal{R}^p$. A *least squares estimator* (LSE) of $\beta$ is defined to be any $\hat{\beta} \in B$ such that

$$\|X - \hat{\beta}Z^\tau\|^2 = \min_{b \in B} \|X - bZ^\tau\|^2. \tag{3.26}$$

For any $l \in \mathcal{R}^p$, $\hat{\beta}l^\tau$ is called an LSE of $\beta l^\tau$. ∎

Throughout this book we consider $B = \mathcal{R}^p$, unless otherwise stated. Differentiating $\|X - bZ^\tau\|^2$ w.r.t. $b$, we obtain that any solution of

$$bZ^\tau Z = XZ \tag{3.27}$$

is an LSE of $\beta$. If the rank of the matrix $Z$ is $p$, in which case $(Z^\tau Z)^{-1}$ exists and $Z$ is said to be of full rank, then there is a unique LSE which is

$$\hat{\beta} = XZ(Z^\tau Z)^{-1}. \tag{3.28}$$

If $Z$ is not of full rank, then there are infinitely many LSE's of $\beta$. It can be shown (exercise) that any LSE of $\beta$ is of the form

$$\hat{\beta} = XZ(Z^\tau Z)^-, \tag{3.29}$$

where $(Z^\tau Z)^-$ is called a *generalized inverse* of $Z^\tau Z$ and satisfies

$$Z^\tau Z(Z^\tau Z)^- Z^\tau Z = Z^\tau Z.$$

Generalized inverse matrices are not unique unless $Z$ is of full rank, in which case $(Z^\tau Z)^- = (Z^\tau Z)^{-1}$ and (3.29) reduces to (3.28).

To study properties of LSE's of $\beta$, we need some assumptions on the distribution of $X$. Since $Z_i$'s are nonrandom, assumptions on the distribution of $X$ can be expressed in terms of assumptions on the distribution of $\varepsilon$. Several commonly adopted assumptions are stated as follows.

*Assumption* A1: $\varepsilon$ is distributed as $N_n(0, \sigma^2 I_n)$ with an unknown $\sigma^2 > 0$.

*Assumption* A2: $E(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = \sigma^2 I_n$ with an unknown $\sigma^2 > 0$.

*Assumption* A3: $E(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon)$ is an unknown matrix.

Assumption A1 is the strongest and implies a parametric model. We may assume a slightly more general assumption that $\varepsilon$ has the $N_n(0, \sigma^2 D)$ distribution with unknown $\sigma^2$ but a known positive definite matrix $D$. Let $D^{-1/2}$ be the inverse of the square root matrix of $D$. Then model (3.25) with assumption A1 holds if we replace $X$, $Z$, and $\varepsilon$ by the transformed variables $\tilde{X} = XD^{-1/2}$, $\tilde{Z} = ZD^{-1/2}$, and $\tilde{\varepsilon} = \varepsilon D^{-1/2}$, respectively. A similar conclusion can be made for assumption A2.

Under assumption A1, the distribution of $X$ is $N_n(\beta Z^\tau, \sigma^2 I_n)$, which is in an exponential family $\mathcal{P}$ with parameter $\theta = (\beta, \sigma^2) \in \mathcal{R}^p \times (0, \infty)$. However, if the matrix $Z$ is not of full rank, then $\mathcal{P}$ is not identifiable (see §2.1.2), since $\beta_1 Z^\tau = \beta_2 Z^\tau$ does not imply $\beta_1 = \beta_2$.

Suppose that the rank of $Z$ is $r \leq p$. Then there is an $n \times r$ submatrix $Z_*$ of $Z$ such that

$$Z = Z_* Q \tag{3.30}$$

and $Z_*$ is of rank $r$, where $Q$ is a fixed $r \times p$ matrix. Then

$$\beta Z^\tau = \beta Q^\tau Z_*^\tau$$

and $\mathcal{P}$ is identifiable if we consider the reparameterization $\tilde{\beta} = \beta Q^\tau$. Note that the new parameter $\tilde{\beta}$ is in a subspace of $\mathcal{R}^p$ with dimension $r$.

In many applications we are interested in estimating some linear functions of $\beta$, i.e., $\vartheta = \beta l^\tau$ for some $l \in \mathcal{R}^p$. From the previous discussion, however, estimation of $\beta l^\tau$ is meaningless unless $l = cQ$ for some $c \in \mathcal{R}^r$ so that

$$\beta l^\tau = \beta Q^\tau c^\tau = \tilde{\beta} c^\tau.$$

The following result shows that $\beta l^\tau$ is estimable if $l = cQ$, which is also necessary for $\beta l^\tau$ to be estimable under assumption A1.

**Theorem 3.6.** Assume model (3.25) with assumption A3.
(i) A necessary and sufficient condition for $l \in \mathcal{R}^p$ being $cQ$ for some $c \in \mathcal{R}^r$ is $l \in \mathcal{R}(Z) = \mathcal{R}(Z^\tau Z)$, where $Q$ is given by (3.30) and $\mathcal{R}(A)$ is the smallest linear subspace of $\mathcal{R}^p$ containing all rows of $A$.
(ii) If $l \in \mathcal{R}(Z)$, then the LSE $\hat{\beta} l^\tau$ is unique and unbiased for $\beta l^\tau$.
(iii) If $l \notin \mathcal{R}(Z)$ and assumption A1 holds, then $\beta l^\tau$ is not estimable.
**Proof.** (i) If $l = cQ$, then

$$l = cQ = c(Z_*^\tau Z_*)^{-1} Z_*^\tau Z_* Q = [c(Z_*^\tau Z_*)^{-1} Z_*^\tau] Z.$$

Hence $l \in \mathcal{R}(Z)$. If $l \in \mathcal{R}(Z)$, then $l = \zeta Z$ for some $\zeta$ and

$$l = \zeta Z_* Q = cQ$$

with $c = \zeta Z_*$.
(ii) If $l \in \mathcal{R}(Z)$, then $l = \zeta Z^\tau Z$ for some $\zeta$ and by (3.29),

$$\begin{aligned} E(\hat{\beta} l^\tau) &= E[XZ(Z^\tau Z)^- l^\tau] \\ &= \beta Z^\tau Z (Z^\tau Z)^- (Z^\tau Z) \zeta^\tau \\ &= \beta Z^\tau Z \zeta^\tau \\ &= \beta l^\tau. \end{aligned}$$

If $\tilde{\beta}$ is any other LSE of $\beta$, then, by (3.27),

$$\hat{\beta} l^\tau - \tilde{\beta} l^\tau = (\hat{\beta} - \tilde{\beta})(Z^\tau Z)\zeta^\tau = (XZ - XZ)\zeta^\tau = 0.$$

(iii) Under assumption A1, if there is an estimator $h(X, Z)$ unbiased for $\beta l^\tau$, then

$$\beta l^\tau = \int_{\mathcal{R}^n} h(x, Z) (2\pi)^{-n/2} \sigma^{-n} \exp\left\{-\tfrac{1}{2\sigma^2}\|x - \beta Z^\tau\|^2\right\} dx.$$

Differentiating w.r.t. $\beta$ and applying Theorem 2.1 leads to

$$l^\tau = Z^\tau \int_{\mathcal{R}^n} h(x, Z) (2\pi)^{-n/2} \sigma^{n-2} (x^\tau - Z\beta^\tau) \exp\left\{-\tfrac{1}{2\sigma^2}\|x - \beta Z^\tau\|^2\right\} dx,$$

which implies $l \in \mathcal{R}(Z)$. ∎

Theorem 3.6 shows that LSE's are unbiased for estimable parameters $\beta l^\tau$. If $Z$ is of full rank, then $\mathcal{R}(Z) = \mathcal{R}^p$ and, therefore, $\beta l^\tau$ is estimable for any $l \in \mathcal{R}^p$.

**Example 3.12** (Simple linear regression). Let $\beta = (\beta_0, \beta_1) \in \mathcal{R}^2$ and $Z_i = (1, t_i)$, $t_i \in \mathcal{R}$, $i = 1, ..., n$. Then model (3.24) or (3.25) is called a *simple linear regression* model. It turns out that

$$Z^\tau Z = \begin{pmatrix} n & \sum_{i=1}^n t_i \\ \sum_{i=1}^n t_i & \sum_{i=1}^n t_i^2 \end{pmatrix}.$$

This matrix is invertible if and only if some $t_i$'s are different. Thus, if some $t_i$'s are different, then the unique unbiased LSE of $\beta l^\tau$ for any $l \in \mathcal{R}^2$ is $XZ(Z^\tau Z)^{-1} l^\tau$, which has the normal distribution if assumption A1 holds.

The result can be easily extended to the case of *polynomial regression* of order $p$ in which $\beta = (\beta_0, \beta_1, ..., \beta_{p-1})$ and $Z_i = (1, t_i, ..., t_i^{p-1})$. ∎

**Example 3.13** (One-way ANOVA). Suppose that $n = \sum_{j=1}^m n_j$ with $m$ positive integers $n_1, ..., n_m$ and that

$$X_i = \mu_j + \varepsilon_i, \qquad i = n_{j-1} + 1, ..., n_j, j = 1, ..., m,$$

where $n_0 = 0$ and $(\mu_1, ..., \mu_m) = \beta$. Let $J_m$ be the $m$-vector of ones. Then the matrix $Z$ in this case is a block diagonal matrix with $J_{n_j}^\tau$ as the $j$th diagonal block. Consequently, $Z^\tau Z$ is an $m \times m$ diagonal matrix whose $j$th diagonal element is $n_j$. Thus, $Z^\tau Z$ is invertible and the unique LSE of $\beta$ is the $m$-vector whose $j$th component is $n_j^{-1} \sum_{i=n_{j-1}+1}^{n_j} X_i$, $j = 1, ..., m$.

Sometimes it is more convenient to use the following notation:

$$X_{ij} = X_{n_{i-1}+j}, \ \varepsilon_{ij} = \varepsilon_{n_{i-1}+j}, \qquad j = 1, ..., n_i, i = 1, ..., m,$$

and

$$\mu_i = \mu + \alpha_i, \qquad i = 1, ..., m.$$

Then our model becomes

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \qquad j = 1, ..., n_i, i = 1, ..., m, \tag{3.31}$$

which is called a *one-way analysis of variance* (ANOVA) model. Under model (3.31), $\beta = (\mu, \alpha_1, ..., \alpha_m) \in \mathcal{R}^{m+1}$. The matrix $Z$ under model (3.31) is not of full rank (exercise). The LSE of $\beta$ under model (3.31) is

$$\hat{\beta} = \left(\bar{X}, \bar{X}_{1\cdot} - \bar{X}, ..., \bar{X}_{m\cdot} - \bar{X}\right),$$

where $\bar{X}$ is still the sample mean of $X_{ij}$'s and $\bar{X}_{i\cdot}$ is the sample mean of the $i$th group $\{X_{ij}, j = 1, ..., n_i\}$. The problem of finding the form of $l^\tau \in \mathcal{R}(Z)$ under model (3.31) is left as an exercise. ∎

The notation used in model (3.31) allows us to generalize the one-way ANOVA model to any $s$-way ANOVA model with a positive integer $s$ under the so-called factorial experiments. The following example is for the two-way ANOVA model.

**Example 3.14** (Two-way balanced ANOVA). Suppose that

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \quad i = 1, ..., a, j = 1, ..., b, k = 1, ..., c, \tag{3.32}$$

where $a$, $b$, and $c$ are some positive integers. Model (3.32) is called a two-way balanced ANOVA model. If we view model (3.32) as a special case of model (3.25), then the parameter vector $\beta$ is

$$\beta = (\mu, \alpha_1, ..., \alpha_a, \beta_1, ..., \beta_b, \gamma_{11}, ..., \gamma_{1b}, ..., \gamma_{a1}, ..., \gamma_{ab}). \tag{3.33}$$

One can obtain the matrix $Z$ and show that it is $n \times p$, where $n = abc$ and $p = 1 + a + b + ab$, and is of rank $ab < p$ (exercise). It can also be shown (exercise) that the LSE of $\beta$ is given by the right-hand side of (3.33) with $\mu$, $\alpha_i$, $\beta_j$, and $\gamma_{ij}$ replaced by $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\beta}_j$, and $\hat{\gamma}_{ij}$, respectively, where $\hat{\mu} = \bar{X}_{\cdots}$, $\hat{\alpha}_i = \bar{X}_{i\cdot\cdot} - \bar{X}_{\cdots}$, $\hat{\beta}_j = \bar{X}_{\cdot j\cdot} - \bar{X}_{\cdots}$, $\hat{\gamma}_{ij} = \bar{X}_{ij\cdot} - \bar{X}_{i\cdot\cdot} - \bar{X}_{\cdot j\cdot} + \bar{X}_{\cdots}$, and a dot is used to denote averaging over the indicated subscript, e.g.,

$$\bar{X}_{\cdot j\cdot} = \frac{1}{ac}\sum_{i=1}^{a}\sum_{k=1}^{c} X_{ijk}$$

with a fixed $j$. ∎

## 3.3.2 The UMVUE and BLUE

We now study UMVUE's in model (3.25) with assumption A1.

**Theorem 3.7.** Consider model (3.25) with assumption A1.
(i) The LSE $\hat{\beta}l^\tau$ is the UMVUE of $\beta l^\tau$ for any estimable $\beta l^\tau$.
(ii) The UMVUE of $\sigma^2$ is $\hat{\sigma}^2 = (n-r)^{-1}\|X - \hat{\beta}Z^\tau\|^2$, where $r$ is the rank of $Z$.
(iii) The UMVUE's in (i) and (ii) attain the Cramér-Rao lower bound.
**Proof.** (i) Let $\hat{\beta}$ be an LSE of $\beta$. By (3.27),

$$(X - \hat{\beta}Z^\tau)Z(\hat{\beta} - \beta)^\tau = (XZ - XZ)(\hat{\beta} - \beta)^\tau = 0$$

and, hence,

$$\begin{aligned}
\|X - \beta Z^\tau\|^2 &= \|X - \hat{\beta}Z^\tau + \hat{\beta}Z^\tau - \beta Z^\tau\|^2 \\
&= \|X - \hat{\beta}Z^\tau\|^2 + \|\hat{\beta}Z^\tau - \beta Z^\tau\|^2 \\
&= \|X - \hat{\beta}Z^\tau\|^2 - 2XZ\beta^\tau + \|\beta Z^\tau\|^2 + \|\hat{\beta}Z^\tau\|^2.
\end{aligned}$$

Using this result and assumption A1, we obtain the following joint Lebesgue p.d.f. of $X$:

$$(2\pi\sigma^2)^{-n/2}\exp\left\{\tfrac{xZ\beta^\tau}{\sigma^2} - \tfrac{\|x-\hat\beta Z^\tau\|^2+\|\hat\beta Z^\tau\|^2}{2\sigma^2} - \tfrac{\|\beta Z^\tau\|^2}{2\sigma^2}\right\}.$$

By Proposition 2.1 and the fact that $\hat\beta Z^\tau = XZ(Z^\tau Z)^- Z^\tau$ is a function of $XZ$, $(XZ, \|X-\hat\beta Z^\tau\|^2)$ is complete and sufficient for $\theta = (\beta, \sigma^2)$. Note that $\hat\beta$ is a function of $XZ$ and, hence, a function of the complete sufficient statistic. If $\beta l^\tau$ is estimable, then $\hat\beta l^\tau$ is unbiased for $\beta l^\tau$ (Theorem 3.6) and, hence, $\hat\beta l^\tau$ is the UMVUE of $\beta l^\tau$.
(ii) Since each column of $Z^\tau \in \mathcal{R}(Z)$, $\hat\beta Z^\tau$ does not depend on the choice of $\hat\beta$ and $E(\hat\beta Z^\tau) = \beta Z^\tau$ (Theorem 3.6). Then

$$\mathrm{Cov}(X-\hat\beta Z^\tau, \hat\beta Z^\tau) = E(X-\hat\beta Z^\tau)Z\hat\beta^\tau = E(XZ - XZ)\hat\beta^\tau = 0 \quad (3.34)$$

and

$$\begin{aligned} E\|X-\hat\beta Z^\tau\|^2 &= E(X-\beta Z^\tau)(X-\beta Z^\tau)^\tau - E[(\beta-\hat\beta)Z^\tau Z(\beta-\hat\beta)^\tau] \\ &= \mathrm{tr}\left(\mathrm{Var}(X) - \mathrm{Var}(\hat\beta Z^\tau)\right) \\ &= \sigma^2[n - \mathrm{tr}\left(Z(Z^\tau Z)^- Z^\tau Z(Z^\tau Z)^- Z^\tau\right)] \\ &= \sigma^2[n - \mathrm{tr}\left((Z^\tau Z)^- Z^\tau Z\right)]. \end{aligned}$$

Since the previous result does not depend on the particular choice of $\hat\beta$ or $(Z^\tau Z)^-$, we can evaluate $\mathrm{tr}((Z^\tau Z)^- Z^\tau Z)$ using a particular $(Z^\tau Z)^-$. From the theory of linear algebra, there exists a $p\times p$ matrix $C$ such that $CC^\tau = I_p$ and

$$C(Z^\tau Z)C^\tau = \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix},$$

where $\Lambda$ is an $r\times r$ diagonal matrix whose diagonal elements are positive. Then a particular choice of $(Z^\tau Z)^-$ is

$$(Z^\tau Z)^- = C\begin{pmatrix} \Lambda^{-1} & 0 \\ 0 & 0 \end{pmatrix} C^\tau \quad (3.35)$$

and

$$(Z^\tau Z)^- Z^\tau Z = C\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} C^\tau$$

whose trace is $r$. Hence $\hat\sigma^2$ is the UMVUE of $\sigma^2$, since it is a function of the complete sufficient statistic and

$$E\hat\sigma^2 = (n-r)^{-1}E\|X-\hat\beta Z^\tau\|^2 = \sigma^2.$$

(iii) The result follows from Proposition 3.2. ∎

The vector $X - \hat{\beta}Z^\tau$ is called the *residual vector* and $\|X - \hat{\beta}Z^\tau\|^2$ is called the *sum of squared residuals* and is denoted by $SSR$. The estimator $\hat{\sigma}^2$ is then equal to $SSR/(n-r)$.

Since $X - \hat{\beta}Z^\tau$ and $\hat{\beta}l^\tau$ are linear in $X$, they are normally distributed under assumption A1. Then (3.34) and assumption A1 imply that $\hat{\sigma}^2$ and $\hat{\beta}l^\tau$ are independent for any estimable $\beta l^\tau$. Furthermore, using the generalized inverse matrix in (3.35), we obtain that

$$SSR = X[I_n - Z(Z^\tau Z)^- Z^\tau]X^\tau, \tag{3.36}$$

where $P_n = I_n - Z(Z^\tau Z)^- Z^\tau$ is a projection matrix of rank $n - r$. Then, there exists an $n \times n$ matrix $G$ such that $GG^\tau = I_n$ and

$$P_n G = \left(G_1^\tau, ..., G_{n-r}^\tau, 0, ..., 0\right),$$

where $G_j$ is the $j$th row of $G^\tau$. This and (3.36) imply that

$$SSR = \sum_{j=1}^{n-r} Y_j^2,$$

where $Y_j = XG_j^\tau$. Let $Y = (Y_1, ..., Y_{n-r})$. Under assumption A1, $Y$ is normal; $\mathrm{Var}(Y) = \sigma^2 I_{n-r}$; and $EY = 0$ since

$$EY_j = E(XG_j^\tau) = \beta Z^\tau P_n G_j^\tau = 0,$$

which follows from the fact that $Z^\tau P_n P_n Z = Z^\tau P_n Z = 0$ by the definition of the generalized inverse. Thus, we have the following result.

**Theorem 3.8.** Consider model (3.25) with assumption A1. For any estimable parameter $\beta l^\tau$, the UMVUE's $\hat{\beta}l^\tau$ and $\hat{\sigma}^2$ are independent; the distribution of $\hat{\beta}l^\tau$ is $N(\beta l^\tau, \sigma^2 l(Z^\tau Z)^- l^\tau)$; and $(n-r)\hat{\sigma}^2/\sigma^2$ has the chi-square distribution $\chi^2_{n-r}$. ∎

**Example 3.15.** In Examples 3.12-3.14, UMVUE's of estimable $\beta l^\tau$ are the LSE's $\hat{\beta}l^\tau$, under assumption A1. In Example 3.13,

$$SSR = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i\cdot})^2;$$

in Example 3.14, if $c > 1$,

$$SSR = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(X_{ijk} - \bar{X}_{ij\cdot})^2. \quad ∎$$

We now study properties of $\hat{\beta}l^\tau$ and $\hat{\sigma}^2$ under assumption A2, i.e., without the normality assumption on $\varepsilon$. From Theorem 3.6 and the proof of Theorem 3.7(ii), $\hat{\beta}l^\tau$ (with an $l \in \mathcal{R}(Z)$) and $\hat{\sigma}^2$ are still unbiased without the normality assumption. In what sense are $\hat{\beta}l^\tau$ and $\hat{\sigma}^2$ optimal beyond being unbiased? We have the following result for the LSE $\hat{\beta}l^\tau$. Some discussion about $\hat{\sigma}^2$ can be found, for example, in Rao (1973, p. 228).

**Theorem 3.9.** Consider model (3.25) with assumption A2.
(i) A necessary and sufficient condition for the existence of a linear function of $X$ that is unbiased for $\beta l^\tau$ is $l \in \mathcal{R}(Z)$.
(ii) (The Gauss-Markov theorem). If $l \in \mathcal{R}(Z)$, then the LSE $\hat{\beta}l^\tau$ is the *best linear unbiased estimator* (BLUE) of $\beta l^\tau$ in the sense that it has the minimum variance in the class of linear unbiased estimators of $\beta l^\tau$.
**Proof.** (i) The sufficiency has been established in Theorem 3.6. Suppose now a linear function of $X$, $Xc^\tau$ with $c \in \mathcal{R}^n$, is unbiased for $\beta l^\tau$. Then

$$\beta l^\tau = E(Xc^\tau) = (EX)c^\tau = \beta Z^\tau c^\tau.$$

Since this equality holds for all $\beta$, $l = cZ$, i.e., $l \in \mathcal{R}(Z)$.
(ii) Let $l \in \mathcal{R}(Z) = \mathcal{R}(Z^\tau Z)$. Then $l = \zeta(Z^\tau Z)$ for some $\zeta$ and $\hat{\beta}l^\tau = \hat{\beta}(Z^\tau Z)\zeta^\tau = XZ\zeta^\tau$ by (3.27). Let $Xc^\tau$ be any linear unbiased estimator of $\beta l^\tau$. From the proof of (i), $cZ = l$. Then

$$\begin{aligned}
\mathrm{Cov}(XZ\zeta^\tau, Xc^\tau - XZ\zeta^\tau) &= E(XZ\zeta^\tau cX^\tau) - E(XZ\zeta^\tau\zeta Z^\tau X^\tau) \\
&= \sigma^2\mathrm{tr}(Z\zeta^\tau c) - \sigma^2\mathrm{tr}(Z\zeta^\tau\zeta Z^\tau) \\
&= \sigma^2[\mathrm{tr}(\zeta^\tau cZ) - \mathrm{tr}(\zeta^\tau l)] = 0.
\end{aligned}$$

Hence

$$\begin{aligned}
\mathrm{Var}(Xc^\tau) &= \mathrm{Var}(Xc^\tau - XZ\zeta^\tau + XZ\zeta^\tau) \\
&= \mathrm{Var}(Xc^\tau - XZ\zeta^\tau) + \mathrm{Var}(XZ\zeta^\tau) \\
&\quad + 2\mathrm{Cov}(XZ\zeta^\tau, Xc^\tau - XZ\zeta^\tau) \\
&= \mathrm{Var}(Xc^\tau - XZ\zeta^\tau) + \mathrm{Var}(\hat{\beta}l^\tau) \\
&\geq \mathrm{Var}(\hat{\beta}l^\tau). \quad \blacksquare
\end{aligned}$$

### 3.3.3 Robustness of LSE's

Consider now model (3.25) under assumption A3. An interesting question is under what conditions on $\mathrm{Var}(\varepsilon)$, the LSE of $\beta l^\tau$ with $l \in \mathcal{R}(Z)$ is still the BLUE. If $\hat{\beta}l^\tau$ is still the BLUE, then we say that $\hat{\beta}l^\tau$, considered as a BLUE, is *robust* against violation of assumption A2. In general, a statistical procedure having certain properties under an assumption is said to be robust against violation of the assumption if the statistical procedure still has the same properties if the assumption is (slightly) violated. For example, the LSE of $\beta l^\tau$ with $l \in \mathcal{R}(Z)$, as an unbiased estimator, is robust against violation of assumption A1 or A2, since the LSE is unbiased as long as $E(\varepsilon) = 0$, which can be always assumed without loss of generality. On the other hand, the LSE as a UMVUE may not be robust against violation of assumption A1 (see §3.5).

**Theorem 3.10.** Consider model (3.25) with assumption A3. The following are equivalent.
(a) $\hat{\beta} l^\tau$ is the BLUE of $\beta l^\tau$ for any $l \in \mathcal{R}(Z)$.
(b) $E(\hat{\beta} l^\tau X \eta^\tau) = 0$ for any $l \in \mathcal{R}(Z)$ and any $\eta$ such that $E X \eta^\tau = 0$.
(c) $Z^\tau \mathrm{Var}(\varepsilon) U = 0$, where $U$ is a matrix such that $Z^\tau U = 0$ and $\mathcal{R}(U^\tau) + \mathcal{R}(Z^\tau) = \mathcal{R}^n$.
(d) $\mathrm{Var}(\varepsilon) = Z \Lambda_1 Z^\tau + U \Lambda_2 U^\tau$ for some $\Lambda_1$ and $\Lambda_2$.
(e) The matrix $Z(Z^\tau Z)^- Z^\tau \mathrm{Var}(\varepsilon)$ is symmetric.
**Proof.** We first show that (a) and (b) are equivalent, which is an analogue of Theorem 3.2(i). Suppose that (b) holds. Let $l \in \mathcal{R}(Z)$. If $Xc^\tau$ is another unbiased estimator of $\beta l^\tau$, then $E(X\eta^\tau) = 0$ with $\eta = c - l(Z^\tau Z)^- Z^\tau$. Hence

$$
\begin{aligned}
\mathrm{Var}(Xc^\tau) &= \mathrm{Var}(Xc^\tau - \hat{\beta} l^\tau + \hat{\beta} l^\tau) \\
&= \mathrm{Var}(Xc^\tau - XZ(Z^\tau Z)^- l^\tau + \hat{\beta} l^\tau) \\
&= \mathrm{Var}(X\eta^\tau + \hat{\beta} l^\tau) \\
&= \mathrm{Var}(X\eta^\tau) + \mathrm{Var}(\hat{\beta} l^\tau) + 2\mathrm{Cov}(X\eta^\tau, \hat{\beta} l^\tau) \\
&= \mathrm{Var}(X\eta^\tau) + \mathrm{Var}(\hat{\beta} l^\tau) + 2E(\hat{\beta} l^\tau X \eta^\tau) \\
&= \mathrm{Var}(X\eta^\tau) + \mathrm{Var}(\hat{\beta} l^\tau) \\
&\geq \mathrm{Var}(\hat{\beta} l^\tau).
\end{aligned}
$$

Suppose now that there are $l \in \mathcal{R}(Z)$ and $\eta$ such that $E(X\eta^\tau) = 0$ but $\delta = E(\hat{\beta} l^\tau X \eta^\tau) \neq 0$. Let $c_t = t\eta + l(Z^\tau Z)^- Z^\tau$. From the previous proof we obtain that

$$
\mathrm{Var}(Xc_t^\tau) = t^2 \mathrm{Var}(X\eta^\tau) + \mathrm{Var}(\hat{\beta} l^\tau) + 2\delta t.
$$

As long as $\delta \neq 0$, there exists a $t$ such that $\mathrm{Var}(Xc_t^\tau) < \mathrm{Var}(\hat{\beta} l^\tau)$. This shows that $\hat{\beta} l^\tau$ cannot be a BLUE and, therefore, (a) implies (b).

We next show that (b) implies (c). Suppose that (b) holds. Since $l \in \mathcal{R}(Z)$, $l = \gamma Z$ for some $\gamma$. For any $\eta$ such that $E(X\eta^\tau) = 0$,

$$
0 = E(l \hat{\beta}^\tau X^\tau \eta) = E[\gamma Z (Z^\tau Z)^- Z^\tau X X^\tau \eta] = \gamma Z (Z^\tau Z)^- Z^\tau \mathrm{Var}(\varepsilon) \eta.
$$

Since this equality holds for all $\gamma$, $Z(Z^\tau Z)^- Z^\tau \mathrm{Var}(\varepsilon)\eta = 0$. Note that $E(X\eta^\tau) = \beta Z \eta^\tau = 0$ for all $\beta$. Hence $Z\eta^\tau = 0$, i.e., $\eta \in \mathcal{R}(U)$. Since this is true for all $\eta$,

$$Z(Z^\tau Z)^- Z^\tau \text{Var}(\varepsilon) U = 0,$$

which implies

$$Z^\tau Z(Z^\tau Z)^- Z^\tau \text{Var}(\varepsilon) U = Z^\tau \text{Var}(\varepsilon) U = 0,$$

since $Z^\tau Z(Z^\tau Z)^- Z^\tau = Z^\tau$. Thus, (c) holds.

To show that (c) implies (d), we need to use the following facts from the theory of linear algebra: there exists nonsingular matrix $C$ such that $\text{Var}(\varepsilon) = CC^\tau$ and $C = ZC_1 + UC_2$ for some matrices $C_j$ (since $\mathcal{R}(U^\tau) + \mathcal{R}(Z^\tau) = \mathcal{R}^n$). Let $\Lambda_1 = C_1 C_1^\tau$, $\Lambda_2 = C_2 C_2^\tau$, and $\Lambda_3 = C_1 C_2^\tau$. Then

$$\text{Var}(\varepsilon) = Z\Lambda_1 Z^\tau + U\Lambda_2 U^\tau + Z\Lambda_3 U^\tau + U\Lambda_3^\tau Z^\tau \tag{3.37}$$

and $Z^\tau \text{Var}(\varepsilon) U = Z^\tau Z \Lambda_3 U^\tau U$, which is 0 if (c) holds. Hence, (c) implies

$$0 = Z(Z^\tau Z)^- Z^\tau Z \Lambda_3 U^\tau U (U^\tau U)^- U^\tau = Z\Lambda_3 U^\tau,$$

which with (3.37) implies (d).

If (d) holds, then $Z(Z^\tau Z)^- Z^\tau \text{Var}(\varepsilon) = Z\Lambda_1 Z^\tau$, which is symmetric. Hence (d) implies (e). To complete the proof we need to show that (e) implies (b), which is left as an exercise. ∎

As a corollary of this theorem, the following result shows when the UMVUE's in model (3.25) with assumption A1 is robust against the violation of $\text{Var}(\varepsilon) = \sigma^2 I_n$.

**Corollary 3.3.** Consider model (3.25) with normally distributed $\varepsilon$ and a full rank $Z$. Then $\hat{\beta} l^\tau$ and $\hat{\sigma}^2$ are still UMVUE's of $\beta l^\tau$ and $\sigma^2$, respectively, if and only if one of (b)-(e) in Theorem 3.10 holds. ∎

**Example 3.16.** Consider model (3.25) with $\beta$ replaced by a random vector $\boldsymbol{\beta}$ which is independent of $\varepsilon$. Such a model is called a linear model with random coefficients. Suppose that $\text{Var}(\varepsilon) = \sigma^2 I_n$, $E(\boldsymbol{\beta}) = \beta$. Then

$$X = \beta Z^\tau + (\boldsymbol{\beta} - \beta) Z^\tau + \varepsilon = \beta Z^\tau + e, \tag{3.38}$$

where $e = (\boldsymbol{\beta} - \beta) Z^\tau + \varepsilon$ satisfies $E(e) = 0$ and

$$\text{Var}(e) = Z \text{Var}(\boldsymbol{\beta}) Z^\tau + \sigma^2 I_n.$$

Since

$$Z(Z^\tau Z)^- Z^\tau \text{Var}(e) = Z \text{Var}(\boldsymbol{\beta}) Z^\tau + \sigma^2 Z(Z^\tau Z)^- Z^\tau$$

is symmetric, by Theorem 3.10, the LSE $\hat{\beta} l^\tau$ under model (3.38) is the BLUE for any $\beta l^\tau$, $l \in \mathcal{R}(Z)$. If $Z$ is of full rank and $\varepsilon$ is normal, then, by Corollary 3.3, $\hat{\beta} l^\tau$ is the UMVUE. ∎

**Example 3.17** (Random effects models). Suppose that

$$X_{ij} = \mu + A_i + e_{ij}, \quad j = 1, ..., n_i, i = 1, ..., m, \tag{3.39}$$

where $\mu \in \mathcal{R}$ is an unknown parameter, $A_i$'s are i.i.d. random variables having mean 0 and variance $\sigma_a^2$, $e_{ij}$'s are i.i.d. random errors with mean 0 and variance $\sigma^2$, and $A_i$'s and $e_{ij}$'s are independent. Model (3.39) is called a one-way *random effects* model and $A_i$'s are unobserved random effects. Let $\varepsilon_{ij} = A_i + e_{ij}$. Then (3.39) is a special case of the general model (3.25) with

$$\mathrm{Var}(\varepsilon) = \sigma_a^2 \Sigma + \sigma^2 I_n,$$

where $\Sigma$ is a block diagonal matrix whose $i$th block is $J_{n_i}^\tau J_{n_i}$ and $J_k$ is the $k$-vector of ones. Under this model, $Z = J_n^\tau$, $n = \sum_{i=1}^m n_i$, and $Z(Z^\tau Z)^- Z^\tau = n^{-1} J_n^\tau J_n$. Note that

$$J_n^\tau J_n \Sigma = \begin{pmatrix} n_1 J_{n_1}^\tau J_{n_1} & n_2 J_{n_1}^\tau J_{n_2} & \cdots & n_m J_{n_1}^\tau J_{n_m} \\ n_1 J_{n_2}^\tau J_{n_1} & n_2 J_{n_2}^\tau J_{n_2} & \cdots & n_m J_{n_2}^\tau J_{n_m} \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \\ n_1 J_{n_m}^\tau J_{n_1} & n_2 J_{n_m}^\tau J_{n_2} & \cdots & n_m J_{n_m}^\tau J_{n_m} \end{pmatrix},$$

which is symmetric if and only if $n_1 = n_2 = \cdots = n_m$. Since $J_n^\tau J_n \mathrm{Var}(\varepsilon)$ is symmetric if and only if $J_n^\tau J_n \Sigma$ is symmetric, a necessary and sufficient condition for the LSE of $\mu$ to be the BLUE is that all $n_i$'s are the same. This condition is also necessary and sufficient for the LSE of $\mu$ to be the UMVUE when $\varepsilon_{ij}$'s are normal. ∎

In some cases we are interested in some (not all) linear functions of $\beta$. For example, consider $\beta l^\tau$ with $l \in \mathcal{R}(H)$, where $H$ is an $n \times p$ matrix such that $\mathcal{R}(H) \subset \mathcal{R}(Z)$. We have the following result.

**Proposition 3.4.** Consider model (3.25) with assumption A3. Suppose that $H$ is a matrix such that $\mathcal{R}(H) \subset \mathcal{R}(Z)$. A necessary and sufficient condition for the LSE $\hat{\beta} l^\tau$ to be the BLUE for any $l \in \mathcal{R}(H)$ is $H(Z^\tau Z)^- Z^\tau \mathrm{Var}(\varepsilon) U = 0$, where $U$ is the same as that in (c) of Theorem 3.10. ∎

**Example 3.18.** Consider model (3.25) with assumption A3 and $Z = (H_1, H_2)$, where $H_1^\tau H_2 = 0$. Suppose that under the reduced model

$$X = \beta_1 H_1^\tau + \varepsilon,$$

$\hat{\beta}_1 l^\tau$ is the BLUE for any $\beta_1 l^\tau$, $l \in \mathcal{R}(H_1)$, and that under the reduced model

$$X = \beta_2 H_2^\tau + \varepsilon,$$

$\hat{\beta}_2 l^\tau$ is not a BLUE for some $\beta_2 l^\tau$, $l \in \mathcal{R}(H_2)$, where $\beta = (\beta_1, \beta_2)$ and $\hat{\beta}_j$'s are LSE's under the reduced models. Let $H = (H_1, 0)$ be $n \times p$. Note that

$$H(Z^\tau Z)^- Z^\tau \mathrm{Var}(\varepsilon) U = H_1 (H_1^\tau H_1)^- H_1^\tau \mathrm{Var}(\varepsilon) U,$$

which is 0 by Theorem 3.10 for the $U$ given in (c) of Theorem 3.10, and

$$Z(Z^\tau Z)^- Z^\tau \mathrm{Var}(\varepsilon) U = H_2 (H_2^\tau H_2)^- H_2^\tau \mathrm{Var}(\varepsilon) U,$$

which is not 0 by Theorem 3.10. This implies that some LSE $\hat{\beta} l^\tau$ is not a BLUE but $\hat{\beta} l^\tau$ is the BLUE if $l \in \mathcal{R}(H)$. ∎

Finally, we consider model (3.25) with $\mathrm{Var}(\varepsilon)$ being a diagonal matrix whose $i$th diagonal element is $\sigma_i^2$, i.e., $\varepsilon_i$'s are uncorrelated but have unequal variances. A straightforward calculation shows that condition (e) in Theorem 3.10 holds if and only if, for all $i \neq j$, $\sigma_i^2 \neq \sigma_j^2$ only when $h_{ij} = 0$, where $h_{ij}$ is the $(i,j)$th element of the projection matrix $Z(Z^\tau Z)^- Z^\tau$. Thus, the LSE's are not BLUE's in general.

Suppose that the unequal variances of $\varepsilon_i$'s are caused by some small perturbations, i.e., $\varepsilon_i = e_i + u_i$, where $\mathrm{Var}(e_i) = \sigma^2$, $\mathrm{Var}(u_i) = \delta_i$, and $e_i$ and $u_i$ are independent so that $\sigma_i^2 = \sigma^2 + \delta_i$. If $\delta_i = 0$ for all $i$ (no perturbations), then assumption A2 holds and any LSE $\hat{\beta} l^\tau$ is the BLUE with variance

$$\mathrm{Var}(\hat{\beta} l^\tau) = \sigma^2 l (Z^\tau Z)^- l^\tau.$$

When $\delta_i > 0$, $\hat{\beta} l^\tau$ is still unbiased for $\beta l^\tau$, $l \in \mathcal{R}(Z)$, and

$$\mathrm{Var}(\hat{\beta} l^\tau) = l (Z^\tau Z)^- \sum_{i=1}^n \sigma_i^2 Z_i^\tau Z_i (Z^\tau Z)^- l^\tau.$$

Suppose that $\delta_i \leq \sigma^2 \delta$. Then

$$\mathrm{Var}(\hat{\beta} l^\tau) \leq (1 + \delta) \sigma^2 l (Z^\tau Z)^- l^\tau.$$

This indicates that the LSE is robust in the sense that its variance increases slightly when there is a slight violation of the equal variance assumption (small $\delta$).

### 3.3.4 Asymptotic properties of LSE's

We consider first the consistency of the LSE $\hat{\beta} l^\tau$ with $l \in \mathcal{R}(Z)$ for every $n$.

**Theorem 3.11.** Consider model (3.25) with assumption A3. Suppose that $\sup_n \lambda_+[\mathrm{Var}(\varepsilon)] < \infty$, where $\lambda_+[A]$ is the largest eigenvalue of the matrix

$A$, and that $\lim_{n\to\infty}\lambda_+[(Z^\tau Z)^-]=0$. Then $\hat\beta l^\tau$ is consistent in mse for any $l\in\mathcal{R}(Z)$.
**Proof.** The result follows from the fact that $\hat\beta l^\tau$ is unbiased and

$$\begin{aligned}\mathrm{Var}(\hat\beta l^\tau)&=l(Z^\tau Z)^-Z^\tau\mathrm{Var}(\varepsilon)Z(Z^\tau Z)^-l^\tau\\&\le\lambda_+[\mathrm{Var}(\varepsilon)]l(Z^\tau Z)^-l^\tau.\quad\blacksquare\end{aligned}$$

Without the normality assumption on $\varepsilon$, the exact distribution of $\hat\beta l^\tau$ is very hard to obtain. The asymptotic distribution of $\hat\beta l^\tau$ is derived in the following result.

**Theorem 3.12.** Consider model (3.25) with assumption A3. Suppose that $0<\inf_n\lambda_-[\mathrm{Var}(\varepsilon)]$, where $\lambda_-[A]$ is the smallest eigenvalue of the matrix $A$, and that

$$\lim_{n\to\infty}\max_{1\le i\le n}Z_i(Z^\tau Z)^-Z_i^\tau=0.\tag{3.40}$$

Suppose further that $n=\sum_{j=1}^k m_j$ for some integers $k$, $m_j$, $j=1,...,k$, with $m_j$'s bounded by a fixed integer $m$, $\varepsilon=(\xi_1,...,\xi_k)$, $\xi_j\in\mathcal{R}^{m_j}$, and $\xi_j$'s are independent.
(i) If $\sup_i E|\varepsilon_i|^{2+\delta}<\infty$, then for any $l\in\mathcal{R}(Z)$,

$$(\hat\beta-\beta)l^\tau\Big/\sqrt{\mathrm{Var}(\hat\beta l^\tau)}\to_d N(0,1).\tag{3.41}$$

(ii) If $\xi_i$'s are i.i.d., then result (3.41) holds.
**Proof.** Let $l\in\mathcal{R}(Z)$. Then

$$\beta Z^\tau Z(Z^\tau Z)^-l^\tau-\beta l^\tau=0$$

and

$$(\hat\beta-\beta)l^\tau=\varepsilon Z(Z^\tau Z)^-l^\tau=\sum_{j=1}^k\xi_j c_{nj}^\tau,$$

where $c_{nj}$ is the $m_j$-vector whose components are $l(Z^\tau Z)^-Z_i^\tau$, $i=m_{j-1}+1,...,m_j$, $m_0=0$. Note that

$$\sum_{j=1}^k\|c_{nj}\|^2=l(Z^\tau Z)^-Z^\tau Z(Z^\tau Z)^-l^\tau=l(Z^\tau Z)^-l^\tau.\tag{3.42}$$

Also,

$$\begin{aligned}\max_{1\le j\le k}\|c_{nj}\|^2&\le m\max_{1\le i\le n}[l(Z^\tau Z)^-Z_i^\tau]^2\\&\le ml(Z^\tau Z)^-l^\tau\max_{1\le i\le n}Z_i(Z^\tau Z)^-Z_i^\tau,\end{aligned}$$

which, together with (3.42) and condition (3.40), implies that

$$\lim_{n\to\infty}\left(\max_{1\le j\le k}\|c_{nj}\|^2 \Big/ \sum_{j=1}^{k}\|c_{nj}\|^2\right)=0.$$

The results then follow from Corollary 1.3. ∎

Under the conditions of Theorem 3.12, $\mathrm{Var}(\varepsilon)$ is a diagonal block matrix with $\mathrm{Var}(\xi_j)$ as the $j$th diagonal block, which includes the case of independent $\varepsilon_i$'s as a special case.

The following lemma tells us how to check condition (3.40).

**Lemma 3.3.** The following are sufficient conditions for (3.40).
(a) $\lambda_+[(Z^\tau Z)^-]\to 0$ and $Z_n(Z^\tau Z)^- Z_n^\tau \to 0$, as $n\to\infty$.
(b) There is an increasing sequence $\{a_n\}$ such that $a_n\to\infty$ and $Z^\tau Z/a_n$ converges to a positive definite matrix. ∎

If $n^{-1}\sum_{i=1}^n t_i^2\to c$ in the simple linear regression model (Example 3.12), where $c$ is a positive constant, then condition (b) in Lemma 3.3 is satisfied with $a_n=n$ and, therefore, Theorem 3.12 applies. In the one-way ANOVA model (Example 3.13),

$$\max_{1\le i\le n} Z_i(Z^\tau Z)^- Z_i^\tau = \lambda_+[(Z^\tau Z)^-]=\max_{1\le j\le m} n_j^{-1}.$$

Hence conditions related to $Z$ in Theorem 3.12 are satisfied if and only if $\min_j n_j\to\infty$. Some similar conclusions can be drawn in the two-way ANOVA model (Example 3.14).

## 3.4 Unbiased Estimators in Survey Problems

In this section we consider unbiased estimation for another type of non-i.i.d. data often encountered in applications: survey data from finite populations. A description of the problem is given in Example 2.3 of §2.1.1. Examples and a fuller account of theoretical aspects of survey sampling can be found in, for example, Cochran (1977) and Särndal, Swensson, and Wretman (1992).

### 3.4.1 UMVUE's of population totals

We use the same notation as in Example 2.3. Let $X=(X_1,...,X_n)$ be a sample from a finite population $\mathcal{P}=\{y_1,...,y_N\}$ with

$$P(X_1=y_{i_1},...,X_n=y_{i_n})=p(\boldsymbol{s}),$$

where $\boldsymbol{s} = \{i_1, ..., i_n\}$ is a subset of distinct elements of $\{1, ..., N\}$ and $p$ is a selection probability measure. We consider univariate $y_i$, although most of our conclusions are valid for the case of multivariate $y_i$. In many survey problems the parameter to be estimated is $Y = \sum_{i=1}^{N} y_i$, the population total.

In Example 2.27, it is shown that $\hat{Y} = N\bar{X} = \frac{N}{n}\sum_{i\in\boldsymbol{s}} y_i$ is unbiased for $Y$ if $p(\boldsymbol{s})$ is constant (simple random sampling); a formula of $\mathrm{Var}(\hat{Y})$ is also given. We now show that $\hat{Y}$ is in fact the UMVUE of $Y$ under simple random sampling. Let $\mathcal{Y}$ be the range of $y_i$, $\theta = (y_1, ..., y_N)$ and $\Theta = \prod_{i=1}^{N} \mathcal{Y}$. Under simple random sampling, the population under consideration is a parametric family indexed by $\theta \in \Theta$.

**Theorem 3.13.** (i) (Watson and Royall). If $p(\boldsymbol{s}) > 0$ for all $\boldsymbol{s}$, then the vector of order statistics $X_{(1)} \le \cdots \le X_{(n)}$ is complete for $\theta \in \Theta$.
(ii) Under simple random sampling, the vector of order statistics is sufficient for $\theta \in \Theta$.
(iii) Under simple random sampling, for any estimable function of $\theta$, its unique UMVUE is the unbiased estimator $h(X_1, ..., X_n)$, where $h$ is symmetric in its $n$ arguments.
**Proof.** (i) Let $h(X)$ be a function of the order statistics. Then $h$ is symmetric in its $n$ arguments. We need to show that if

$$E[h(X)] = \sum_{\boldsymbol{s}=\{i_1,...,i_n\}\subset\{1,...,N\}} p(\boldsymbol{s})h\left(y_{i_1}, ..., y_{i_n}\right) = 0 \qquad (3.43)$$

for all $\theta \in \Theta$, then $h(y_{i_1}, ..., y_{i_n}) = 0$ for all $y_{i_1}, ..., y_{i_n}$. First, suppose that all $N$ elements of $\theta$ are equal to $a \in \mathcal{Y}$. Then (3.43) implies $h(a, ..., a) = 0$. Next, suppose that $N-1$ elements in $\theta$ are equal to $a$ and one is $b > a$. Then (3.43) reduces to

$$q_1 h(a, ..., a) + q_2 h(a, ..., a, b),$$

where $q_1$ and $q_2$ are some known numbers in $(0, 1)$. Since $h(a, ..., a) = 0$ and $q_2 \ne 0$, $h(a, ..., a, b) = 0$. Using the same argument, we can show that $h(a, ..., a, b, ..., b) = 0$ for any $k$ $a$'s and $n-k$ $b$'s. Suppose next that elements of $\theta$ are equal to $a$, $b$, or $c$, $a < b < c$. Then we can show that $h(a, ..., a, b, ..., b, c, ..., c) = 0$ for any $k$ $a$'s, $l$ $b$'s, and $n-k-l$ $c$'s. Continuing inductively, we see that $h(y_1, ..., y_n) = 0$ for all possible $y_1, ..., y_n$. This completes the proof of (i).
(ii) The result follows from the factorization theorem (Theorem 2.2), the fact that $p(\boldsymbol{s})$ is constant under simple random sampling, and

$$P(X_1 = y_{i_1}, ..., X_n = y_{i_n}) = P(X_{(1)} = y_{(i_1)}, ..., X_{(n)} = y_{(i_n)})/n!,$$

where $y_{(i_1)} \le \cdots \le y_{(i_n)}$ are the ordered values of $y_{i_1}, ..., y_{i_n}$.
(iii) The result follows directly from (i) and (ii). ∎

It is interesting to note the following two issues. (1) Although we have a parametric problem under simple random sampling, the sufficient and complete statistic is the same as that in a nonparametric problem (Example 2.17). (2) For the completeness of the order statistics, we do not need the assumption of simple random sampling.

**Example 3.19.** From Example 2.27, $\hat{Y} = N\bar{X}$ is unbiased for $Y$. Since $\hat{Y}$ is symmetric in its arguments, it is the UMVUE of $Y$. We now derive the UMVUE for $\mathrm{Var}(\hat{Y})$. From Example 2.27,

$$\mathrm{Var}(\hat{Y}) = \frac{N^2}{n}\left(1 - \frac{n}{N}\right)\sigma^2, \qquad \sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \frac{Y}{N}\right)^2. \tag{3.44}$$

It can be shown (exercise) that $E(S^2) = \sigma^2$, where $S^2$ is the usual sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i\in \boldsymbol{s}}\left(y_i - \frac{\hat{Y}}{N}\right)^2.$$

Since $S^2$ is symmetric in its arguments, $\frac{N^2}{n}\left(1 - \frac{n}{N}\right)S^2$ is the UMVUE of $\mathrm{Var}(\hat{Y})$. ∎

Simple random sampling is rarely used in practice, since it is inefficient unless the population is fairly homogeneous w.r.t. the $y_i$'s. A sampling plan often used in practice is the *stratified sampling* plan which can be described as follows. The population $\mathcal{P}$ is divided into nonoverlapping subpopulations $\mathcal{P}_1, ..., \mathcal{P}_H$ called strata; a sample is drawn from each stratum $\mathcal{P}_h$, independently across the strata. There are many reasons for stratification: (1) it may produce a gain in precision in parameter estimation when a heterogeneous population is divided into strata, each of which is internally homogeneous; (2) sampling problems may differ markedly in different parts of the population; and (3) administrative considerations may also lead to stratification. More discussions can be found, for example, in Cochran (1977).

In stratified sampling, if a simple random sample (without replacement), $X_h = (X_{h1}, ..., X_{hn_h})$, is drawn from each stratum, where $n_h$ is the sample size in stratum $h$, then the joint distribution of $X = (X_1, ..., X_H)$ is in a parametric family indexed by $\theta = (h, \theta_1, ..., \theta_H)$, where $h = 1, ..., H$ and $\theta_h = (y_i, i \in \mathcal{P}_h)$. Let $\mathcal{Y}_h$ be the range of $y_i$'s in stratum $h$ and $\Theta_h = \prod_{i=1}^{N_h}\mathcal{Y}_h$, where $N_h$ is the size of $\mathcal{P}_h$. We assume that the parameter space is $\Theta = \{1, ..., H\} \times \prod_{i=1}^{H}\Theta_h$. The following result is similar to Theorem 3.13.

**Theorem 3.14.** Let $X$ be a sample obtained using the stratified simple random sampling plan described previously.
(i) For each $h$, let $Z_h$ be the vector of the ordered values of the sample in stratum $h$. Then $(Z_1, ..., Z_H)$ is sufficient and complete for $\theta \in \Theta$.
(ii) For any estimable function of $\theta$, its unique UMVUE is the unbiased estimator $h(X)$ which is symmetric in its first $n_1$ arguments, symmetric in its second $n_2$ arguments,..., and symmetric in its last $n_H$ arguments. ∎

**Example 3.20.** Consider the estimation of the population total $Y$ based on a sample $X = (X_{hi}, i = 1, ..., n_h, h = 1, ..., H)$ obtained by stratified simple random sampling. Let $Y_h$ be the population total of the $h$th stratum and let $\hat{Y}_h = N_h \bar{X}_{h\cdot}$, where $\bar{X}_{h\cdot}$ is the sample mean of the sample from stratum $h$, $h = 1, ..., H$. From Example 2.27, each $\hat{Y}_h$ is an unbiased estimator of $Y_h$. Let

$$\hat{Y}_{st} = \sum_{h=1}^{H} \hat{Y}_h = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \frac{N_h}{n_h} X_{hi}.$$

Then, by Theorem 3.14, $\hat{Y}_{st}$ is the UMVUE of $Y$. Since $\hat{Y}_1, ..., \hat{Y}_H$ are independent, it follows from (3.44) that

$$\text{Var}(\hat{Y}_{st}) = \sum_{h=1}^{H} \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) \sigma_h^2, \tag{3.45}$$

where $\sigma_h^2 = (N_h - 1)^{-1} \sum_{i \in \mathcal{P}_h} (y_i - Y_h/N_h)^2$. A similar argument to that in Example 3.19 shows that the UMVUE of $\text{Var}(\hat{Y}_{st})$ is

$$S_{st}^2 = \sum_{h=1}^{H} \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) S_h^2, \tag{3.46}$$

where $S_h^2$ is the usual sample variance based on $X_{h1}, ..., X_{hn_h}$.

It is interesting to compare the mse of the UMVUE $\hat{Y}_{st}$ with the mse of the UMVUE $\hat{Y}$ under simple random sampling. Let $\sigma^2$ be given by (3.44). Then

$$(N-1)\sigma^2 = \sum_{h=1}^{H} (N_h - 1)\sigma_h^2 + \sum_{h=1}^{H} N_h (\mu_h - \mu)^2,$$

where $\mu_h = Y_h/N_h$ is the population mean of the $h$th stratum and $\mu = Y/N$ is the overall population mean. By (3.44), (3.45), and (3.46), $\text{Var}(\hat{Y}) \geq \text{Var}(\hat{Y}_{st})$ if and only if

$$\sum_{h=1}^{H} \tfrac{N^2 N_h}{n(N-1)} \left(1 - \tfrac{n}{N}\right) (\mu_h - \mu)^2 \geq \sum_{h=1}^{H} \left[\tfrac{N_h^2}{n_h} \left(1 - \tfrac{n_h}{N_h}\right) - \tfrac{N^2 (N_h - 1)}{n(N-1)} \left(1 - \tfrac{n}{N}\right)\right] \sigma_h^2.$$

This means that stratified simple random sampling is better than simple random sampling if the deviations $\mu_h - \mu$ are sufficiently large. If $\frac{n_h}{N_h} \equiv \frac{n}{N}$ (proportional allocation), then this condition simplifies to

$$\sum_{h=1}^{H} N_h(\mu_h - \mu)^2 \geq \sum_{h=1}^{H} \left(1 - \frac{N_h}{N}\right) \sigma_h^2, \tag{3.47}$$

which is usually true when $\mu_h$'s are different and some $N_h$'s are large. ∎

### 3.4.2 Horvitz-Thompson estimators

If some elements of the finite population $\mathcal{P}$ are groups (called clusters) of subunits, then sampling from $\mathcal{P}$ is *cluster sampling*. Cluster sampling is used often because of administrative convenience or economic considerations. Although sometimes the first intention may be to use the subunits as sampling units, it is found that no reliable list of the subunits in the population is available. For example, in many countries there are no complete lists of the people or houses in a region. From the maps of the region, however, it can be divided into units such as cities or blocks in the cities.

In cluster sampling, one may greatly increase the precision of estimation by using sampling with probability proportional to cluster size. Thus, unequal probability sampling is often used.

Suppose that a sample of clusters is obtained. If subunits within a selected cluster give similar results, then it may be uneconomical to measure them all. A sample of the subunits in any chosen cluster may be selected. This is called two-stage sampling. One can continue this process to have a multistage sampling (e.g., cities $\rightarrow$ blocks $\rightarrow$ houses $\rightarrow$ people). Of course, at each stage one may use stratified sampling and/or unequal probability sampling.

When the sampling plan is complex, so is the structure of the observations. We now introduce a general method of deriving unbiased estimators of population totals, which are called *Horvitz-Thompson estimators*.

**Theorem 3.15.** Let $X = \{y_i, i \in \boldsymbol{s}\}$ denote a sample from $\mathcal{P} = \{y_1, ..., y_N\}$ which is selected, without replacement, by some method. Define

$$\pi_i = \text{probability that } i \in \boldsymbol{s}, \quad i = 1, ..., N.$$

(i) (Horvitz-Thompson). If $\pi_i > 0$ for $i = 1, ..., N$ and $\pi_i$ is known when $i \in \boldsymbol{s}$, then $\hat{Y}_{ht} = \sum_{i \in \boldsymbol{s}} y_i/\pi_i$ is an unbiased estimator of the population total $Y$.
(ii) Define

$$\pi_{ij} = \text{probability that } i \in \boldsymbol{s} \text{ and } j \in \boldsymbol{s}, \quad i = 1, ..., N,\ j = 1, ..., N.$$

Then

$$\begin{aligned}\mathrm{Var}(\hat{Y}_{ht}) &= \sum_{i=1}^{N} \frac{1-\pi_i}{\pi_i} y_i^2 + 2\sum_{i=1}^{N}\sum_{j=i+1}^{N} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j} y_i y_j && (3.48)\\ &= \sum_{i=1}^{N}\sum_{j=i+1}^{N} (\pi_i\pi_j - \pi_{ij})\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2. && (3.49)\end{aligned}$$

**Proof.** (i) Let $a_i = 1$ if $i \in \boldsymbol{s}$ and $a_i = 0$ if $i \notin \boldsymbol{s}$, $i = 1, ..., N$. Then $E(a_i) = \pi_i$ and

$$E(\hat{Y}_{ht}) = E\left(\sum_{i=1}^{N} \frac{a_i y_i}{\pi_i}\right) = \sum_{i=1}^{N} y_i = Y.$$

(ii) Since $a_i^2 = a_i$,

$$\mathrm{Var}(a_i) = E(a_i) - [E(a_i)]^2 = \pi_i(1-\pi_i).$$

For $i \neq j$,

$$\mathrm{Cov}(a_i, a_j) = E(a_i a_j) - E(a_i)E(a_j) = \pi_{ij} - \pi_i\pi_j.$$

Then

$$\begin{aligned}\mathrm{Var}(\hat{Y}_{ht}) &= \mathrm{Var}\left(\sum_{i=1}^{N} \frac{a_i y_i}{\pi_i}\right)\\ &= \sum_{i=1}^{N} \frac{y_i^2}{\pi_i^2}\mathrm{Var}(a_i) + 2\sum_{i=1}^{N}\sum_{j=i+1}^{N} \frac{y_i y_j}{\pi_i\pi_j}\mathrm{Cov}(a_i, a_j)\\ &= \sum_{i=1}^{N} \frac{1-\pi_i}{\pi_i} y_i^2 + 2\sum_{i=1}^{N}\sum_{j=i+1}^{N} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j} y_i y_j.\end{aligned}$$

Hence (3.48) follows. To show (3.49), note that

$$\sum_{i=1}^{N} \pi_i = n \qquad \text{and} \qquad \sum_{j=1,\dots,N, j\neq i} \pi_{ij} = (n-1)\pi_i,$$

which implies

$$\sum_{j=1,\dots,N, j\neq i} (\pi_{ij} - \pi_i\pi_j) = (n-1)\pi_i - \pi_i(n-\pi_i) = -\pi_i(1-\pi_i).$$

Hence

$$\sum_{i=1}^{N} \frac{1-\pi_i}{\pi_i} y_i^2 = \sum_{i=1}^{N} \sum_{j=1,\ldots,N, j\neq i} (\pi_i \pi_j - \pi_{ij}) \frac{y_i^2}{\pi_i^2}$$
$$= \sum_{i=1}^{N} \sum_{j=i+1}^{N} (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i^2}{\pi_i^2} + \frac{y_j^2}{\pi_j^2} \right)$$

and, by (3.48),

$$\mathrm{Var}(\hat{Y}_{ht}) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} (\pi_{ij} - \pi_i \pi_j) \left( \frac{y_i^2}{\pi_i^2} + \frac{y_j^2}{\pi_j^2} - \frac{2 y_i y_j}{\pi_i \pi_j} \right)$$
$$= \sum_{i=1}^{N} \sum_{j=i+1}^{N} (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 . \quad \blacksquare$$

Using the same idea, we can obtain unbiased estimators of $\mathrm{Var}(\hat{Y}_{ht})$. Suppose that $\pi_{ij} > 0$ for all $i$ and $j$ and $\pi_{ij}$ is known when $i \in \boldsymbol{s}$ and $j \in \boldsymbol{s}$. By (3.48), an unbiased estimator of $\mathrm{Var}(\hat{Y}_{ht})$ is

$$v_1 = \sum_{i \in \boldsymbol{S}} \frac{1-\pi_i}{\pi_i^2} y_i^2 + 2 \sum_{i \in \boldsymbol{S}} \sum_{j \in \boldsymbol{S}, j>i} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j \pi_{ij}} y_i y_j . \tag{3.50}$$

By (3.49), an unbiased estimator of $\mathrm{Var}(\hat{Y}_{ht})$ is

$$v_2 = \sum_{i \in \boldsymbol{S}} \sum_{j \in \boldsymbol{S}, j>i} \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 . \tag{3.51}$$

Variance estimators $v_1$ and $v_2$ may not be the same in general, but they are the same in some special cases (Exercise 84). A more serious problem is that they may take negative values. Some discussions about deriving better estimators of $\mathrm{Var}(\hat{Y}_{st})$ are provided in Cochran (1977, Chapter 9A).

Some special cases of Theorem 3.15 are considered as follows.

Under simple random sampling, $\pi_i = n/N$. Thus, $\hat{Y}$ in Example 3.19 is the Horvitz-Thompson estimator.

Under stratified simple random sampling, $\pi_i = n_h/N_h$ if unit $i$ is in stratum $h$. Hence, the estimator $\hat{Y}_{st}$ in Example 3.20 is the Horvitz-Thompson estimator.

Suppose now each $y_i \in \mathcal{P}$ is a cluster, i.e., $y_i = (y_{i1}, ..., y_{iM_i})$, where $M_i$ is the size of the $i$th cluster, $i = 1, ..., N$. The total number of units in $\mathcal{P}$ is then $M = \sum_{i=1}^{N} M_i$. Consider a single-stage sampling plan, i.e., if $y_i$ is selected, then every $y_{ij}$ is observed. If simple random sampling is used, then $\pi_i = k/N$, where $k$ is the first-stage sample size (the total sample size is $n = \sum_{i=1}^k M_i$), and the Horvitz-Thompson estimator is

$$\hat{Y}_s = \frac{N}{k} \sum_{i \in \boldsymbol{s}_1} \sum_{j=1}^{M_i} y_{ij} = \frac{N}{k} \sum_{i \in \boldsymbol{s}_1} Y_i,$$

where $\boldsymbol{s}_1$ is the index set of first-stage sampled clusters and $Y_i$ is the total of the $i$th cluster. In this case,

$$\mathrm{Var}(\hat{Y}_s) = \frac{N^2}{k(N-1)} \left(1 - \frac{k}{N}\right) \sum_{i=1}^{N} \left(Y_i - \frac{Y}{N}\right)^2.$$

If the selection probability is proportional to the cluster size, then $\pi_i = kM_i/M$ and the Horvitz-Thompson estimator is

$$\hat{Y}_{pps} = \frac{M}{k} \sum_{i \in \boldsymbol{s}_1} \frac{1}{M_i} \sum_{j=1}^{M_i} y_{ij} = \frac{M}{k} \sum_{i \in \boldsymbol{s}_1} \frac{Y_i}{M_i}$$

whose variance is given by (3.48) or (3.49). Usually $\mathrm{Var}(\hat{Y}_{pps})$ is smaller than $\mathrm{Var}(\hat{Y}_s)$; see the discussions in Cochran (1977, Chapter 9A).

Consider next a two-stage sampling in which $k$ first-stage clusters are selected and a simple random sample of size $m_i$ is selected from each sampled cluster $y_i$, where sampling is independent across clusters. If the first-stage sampling plan is simple random sampling, then $\pi_i = km_i/(NM_i)$ and the Horvitz-Thompson estimator is

$$\hat{Y}_s = \frac{N}{k} \sum_{i \in \boldsymbol{s}_1} \frac{M_i}{m_i} \sum_{j \in \boldsymbol{s}_{2i}} y_{ij},$$

where $\boldsymbol{s}_{2i}$ denotes the second-stage sample from cluster $i$. If the first-stage selection probability is proportional to the cluster size, then $\pi_i = km_i/M$ and the Horvitz-Thompson estimator is

$$\hat{Y}_{pps} = \frac{M}{k} \sum_{i \in \boldsymbol{s}_1} \frac{1}{m_i} \sum_{j \in \boldsymbol{s}_{2i}} y_{ij}.$$

Finally, let us consider another popular sampling method called *systematic sampling*. Suppose that $\mathcal{P} = \{y_1, ..., y_N\}$ and the population size $N = nk$ for two integers $n$ and $k$. To select a sample of size $n$, we first draw a $j$ randomly from $\{1, ..., k\}$. Our sample is then

$$\{y_j, y_{j+k}, y_{j+2k}, ..., y_{j+(n-1)k}\}.$$

Systematic sampling is used mainly because it is easier to draw a systematic sample and often easier to execute without mistakes. It is also likely that systematic sampling provides more efficient point estimators than simple random sampling or even stratified sampling, since the sample units are spread more evenly over the population. Under systematic sampling, $\pi_i = k^{-1}$ for every $i$ and the Horvitz-Thompson estimator of the population total is

$$\hat{Y}_{sy} = k \sum_{t=1}^{n} y_{j+(t-1)k}.$$

The unbiasedness of this estimator is a direct consequence of Theorem 3.15, but it can be easily shown as follows. Since $j$ takes value $i \in \{1, ..., k\}$ with probability $k^{-1}$,

$$E(\hat{Y}_{sy}) = k\left(\frac{1}{k}\sum_{i=1}^{k}\sum_{t=1}^{n} y_{i+(t-1)k}\right) = \sum_{i=1}^{N} y_i = Y.$$

The variance of $\hat{Y}_{sy}$ is simply

$$\mathrm{Var}(\hat{Y}_{sy}) = \frac{N^2}{k}\sum_{i=1}^{k}(\mu_i - \mu)^2,$$

where $\mu_i = n^{-1}\sum_{t=1}^{n} y_{i+(t-1)k}$ and $\mu = k^{-1}\sum_{i=1}^{k}\mu_i = Y/N$. Let $\sigma^2$ be given by (3.44) and

$$\sigma_{sy}^2 = \frac{1}{k(n-1)}\sum_{i=1}^{k}\sum_{t=1}^{n}(y_{i+(t-1)k} - \mu_i)^2.$$

Then

$$(N-1)\sigma^2 = n\sum_{i=1}^{k}(\mu_i - \mu)^2 + \sum_{i=1}^{k}\sum_{t=1}^{n}(y_{i+(t-1)k} - \mu_i)^2.$$

Thus,

$$(N-1)\sigma^2 = N^{-1}\mathrm{Var}(\hat{Y}_{sy}) + k(n-1)\sigma_{sy}^2$$

and

$$\mathrm{Var}(\hat{Y}_{sy}) = N(N-1)\sigma^2 - N(N-k)\sigma_{sy}^2.$$

Since the variance of the Horvitz-Thompson estimator of the population total under simple random sampling is, by (3.44),

$$\frac{N^2}{n}\left(1 - \frac{n}{N}\right)\sigma^2 = N(k-1)\sigma^2,$$

the Horvitz-Thompson estimator under systematic sampling has a smaller variance if and only if $\sigma_{sy}^2 > \sigma^2$.

## 3.5 Asymptotically Unbiased Estimators

As we discussed in §2.5, we often need to consider biased but asymptotically unbiased estimators. A large and useful class of such estimators are smooth functions of some exactly unbiased estimators such as UMVUE's, U-statistics, and LSE's. Some other methods of constructing asymptotically unbiased estimators are also introduced in this section.

### 3.5.1 Functions of unbiased estimators

If the parameter to be estimated is $\vartheta = g(\theta)$ with a vector-valued parameter $\theta$ and $U_n$ is a vector of unbiased estimators of components of $\theta$ (i.e., $EU_n = \theta$), then $T_n = g(U_n)$ is asymptotically unbiased for $\vartheta$. Assume that $g$ is second-order differentiable and $\|U_n - \theta\| = o_p(1)$. Then

$$\tilde{b}_{T_n}(P) = \mathrm{tr}(\nabla^2 g(\theta)\mathrm{Var}(U_n))/2$$

and

$$\mathrm{amse}_{T_n}(P) = \nabla g(\theta)\mathrm{Var}(U_n)[\nabla g(\theta)]^\tau$$

(Theorem 2.6). Hence, $T_n$ has a good performance in terms of amse if $U_n$ is optimal in terms of mse (such as the UMVUE).

The following are some examples.

**Example 3.21** (Ratio estimators). Let $(X_1, Y_1), ..., (X_n, Y_n)$ be i.i.d. random 2-vectors. Consider the estimation of the ratio of two population means: $\vartheta = \mu_y/\mu_x$ ($\mu_x \neq 0$). Note that $(\bar{Y}, \bar{X})$, the vector of sample means, is unbiased for $(\mu_y, \mu_x)$. The sample means are UMVUE's under some statistical models (§3.1 and §3.2) and are BLUE's in general (Example 2.22). The *ratio* estimator is $T_n = \bar{Y}/\bar{X}$. Assume that $\sigma_x^2 = \mathrm{Var}(X_1)$, $\sigma_y^2 = \mathrm{Var}(Y_1)$, and $\sigma_{xy} = \mathrm{Cov}(X_1, Y_1)$ exist. A direct calculation shows that

$$\tilde{b}_{T_n}(P) = \frac{\vartheta\sigma_x^2 - \sigma_{xy}}{\mu_x^2 n}, \tag{3.52}$$

and

$$\sqrt{n}(T_n - \vartheta) \to_d N\left(0, \frac{\sigma_y^2 - 2\vartheta\sigma_{xy} + \vartheta^2\sigma_x^2}{\mu_x^2}\right), \tag{3.53}$$

which implies

$$\underline{\mathrm{amse}}_{T_n}(P) = \frac{\sigma_y^2 - 2\vartheta\sigma_{xy} + \vartheta^2\sigma_x^2}{\mu_x^2 n}. \tag{3.54}$$

Results (3.52) and (3.54) still hold when $(X_1, Y_1), ..., (X_n, Y_n)$ is a sample from a finite bivariate population of size $N$ (exercise). In some problems we are not interested in the ratio, but the use of a ratio estimator to improve an estimator of a marginal mean. For example, suppose that $\mu_x$ is known and we are interested in estimating $\mu_y$. Consider the following estimator

$$\hat{\mu}_y = (\bar{Y}/\bar{X})\mu_x.$$

Note that $\hat{\mu}_y$ is not unbiased; its $n^{-1}$ order asymptotic bias is

$$\tilde{b}_{\hat{\mu}_y}(P) = \frac{\vartheta\sigma_x^2 - \sigma_{xy}}{\mu_x n};$$

and

$$\underline{\text{amse}}_{\hat{\mu}_y}(P) = \frac{\sigma_y^2 - 2\vartheta\sigma_{xy} + \vartheta^2\sigma_x^2}{n}.$$

Comparing $\hat{\mu}_y$ with the unbiased estimator $\bar{Y}$, we find that $\hat{\mu}_y$ is asymptotically more efficient if and only if

$$2\vartheta\sigma_{xy} > \vartheta^2\sigma_x^2,$$

which means that $\hat{\mu}_y$ is a better estimator if and only if the correlation between $X_1$ and $Y_1$ is large enough to pay off the extra variability caused by using $\mu_x/\bar{X}$. ∎

Another example related to a bivariate sample is the sample correlation coefficient defined in Exercise 19 in §2.6.

**Example 3.22.** Consider a polynomial regression of order $p$:

$$X_i = \beta Z_i^\tau + \varepsilon_i, \qquad i = 1, ..., n,$$

where $\beta = (\beta_0, \beta_1, ..., \beta_{p-1})$, $Z_i = (1, t_i, ..., t_i^{p-1})$, and $\varepsilon_i$'s are i.i.d. with mean 0 and variance $\sigma^2 > 0$. Suppose that the parameter to be estimated is $t_\beta \in \mathcal{R}$ such that

$$\sum_{j=0}^{p-1} \beta_j t_\beta^j = \max_{t\in\mathcal{R}} \sum_{j=0}^{p-1} \beta_j t^j.$$

Note that $t_\beta = g(\beta)$ for some function $g$. Let $\hat{\beta}$ be the LSE of $\beta$. Then the estimator $\hat{t}_\beta = g(\hat{\beta})$ is asymptotically unbiased and its amse can be derived under some conditions (exercise).

**Example 3.23.** In the study of the reliability of a system component, we assume that

$$X_{ij} = z(t_j)\boldsymbol{\theta}_i^\tau + \varepsilon_{ij}, \quad i = 1, ..., k, \ j = 1, ..., m.$$

Here $X_{ij}$ is the measurement of the $i$th sample component at time $t_j$; $z(t)$ is a $q$-vector whose components are known functions of the time $t$; $\boldsymbol{\theta}_i$'s are unobservable random $q$-vectors that are i.i.d. from $N_q(\theta, \Sigma)$, where $\theta$ and $\Sigma$ are unknown; $\varepsilon_{ij}$'s are i.i.d. measurement errors with mean zero and variance $\sigma^2$; and $\boldsymbol{\theta}_i$'s and $\varepsilon_{ij}$'s are independent. As a function of $t$, $z(t)\boldsymbol{\theta}^\tau$ is the degradation curve for a particular component and $z(t)\theta^\tau$ is the mean degradation curve. Suppose that a component will fail to work if $z(t)\boldsymbol{\theta}^\tau < \eta$, a given critical value. Assume that $z(t)'\boldsymbol{\theta}$ is always a decreasing function of $t$. Then the reliability function of a component is

$$R(t) = P(z(t)\boldsymbol{\theta}^\tau > \eta) = \Phi\left(\frac{z(t)\theta^\tau - \eta}{s(t)}\right),$$

where $s(t) = \sqrt{z(t)\Sigma[z(t)]^\tau}$ and $\Phi$ is the standard normal distribution function. For a fixed $t$, estimators of $R(t)$ can be obtained by estimating $\theta$ and $\Sigma$, since $\Phi$ is a known function. It can be shown (exercise) that the BLUE of $\theta$ is the LSE

$$\hat{\theta} = \bar{X}Z(Z^\tau Z)^{-1},$$

where $Z = ([z(t_1)]^\tau, ...., [z(t_k)]^\tau)^\tau$, $X_i = (X_{i1}, ..., X_{im})$, and $\bar{X}$ is the sample mean of $X_i$'s. The estimation of $\Sigma$ is more difficult. An asymptotically unbiased (as $k \to \infty$) estimator of $\Sigma$ is

$$\hat{\Sigma} = \frac{1}{k}\sum_{i=1}^{k}(Z^\tau Z)^{-1}Z^\tau(X_i - \bar{X})^\tau(X_i - \bar{X})Z(Z^\tau Z)^{-1} - \hat{\sigma}^2(Z^\tau Z)^{-1},$$

where

$$\hat{\sigma}^2 = \frac{1}{k(m-q)}\sum_{i=1}^{k}[X_iX_i^\tau - X_iZ(Z^\tau Z)^{-1}Z^\tau X_i^\tau].$$

Hence an estimator of $R(t)$ is

$$\hat{R}(t) = \Phi\left(\frac{z(t)\hat{\theta}^\tau - \eta}{\hat{s}(t)}\right),$$

where

$$\hat{s}(t) = \left\{z(t)\hat{\Sigma}[z(t)]^\tau\right\}^{1/2}.$$

If we define $Y_{i1} = X_iZ(Z^\tau Z)^{-1}[z(t)]^\tau$, $Y_{i2} = \{X_iZ(Z^\tau Z)^{-1}[z(t)]^\tau\}^2$, $Y_{i3} = [X_iX_i^\tau - X_iZ(Z^\tau Z)^{-1}Z^\tau X_i^\tau]/(m-q)$ and $Y_i = (Y_{i1}, Y_{i2}, Y_{i3})'$, then it is apparent that $\hat{R}(t)$ can be written as $g(\bar{Y})$ for a function

$$g(y_1, y_2, y_3) = \Phi\left(\frac{y_1 - \eta}{\sqrt{y_2 - y_1^2 - y_3 z(t)(Z^\tau Z)^{-1}[z(t)]^\tau}}\right).$$

Suppose that $\varepsilon_{ij}$ has a finite fourth moment, which implies the existence of $\mathrm{Var}(Y_i)$. The amse of $\hat{R}(t)$ can be derived (exercise). ∎

### 3.5.2 The method of moments

The method of moments is the oldest method of deriving point estimators. It almost always produces some asymptotically unbiased estimators, although they may not be the best estimators.

Consider a parametric problem where $X_1, ..., X_n$ are i.i.d. random variables from $P_\theta$, $\theta \in \Theta \subset \mathcal{R}^k$, and $E|X_1|^k < \infty$. Let $\mu_j = EX_1^j$ be the $j$th moment of $P$ and let

$$\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j$$

be the $j$th *sample moment*, which is an unbiased estimator of $\mu_j$, $j = 1, ..., k$. Typically,

$$\mu_j = h_j(\theta), \qquad j = 1, ..., k, \tag{3.55}$$

for some functions $h_j$ on $\mathcal{R}^k$. By substituting $\mu_j$'s on the left-hand side of (3.55) by the sample moments $\hat{\mu}_j$, we obtain a *moment estimator* $\hat{\theta}$, i.e., $\hat{\theta}$ satisfies

$$\hat{\mu}_j = h_j(\hat{\theta}), \qquad j = 1, ..., k,$$

which is a sample analogue of (3.55). This method of deriving estimators is called the *method of moments.* Note that an important statistical principle, the *substitution principle*, is applied in this method.

Let $\hat{\mu} = (\hat{\mu}_1, ..., \hat{\mu}_k)$ and $h = (h_1, ..., h_k)$. Then $\hat{\mu} = h(\hat{\theta})$. If $h^{-1}$ exists, then the unique moment estimator of $\theta$ is $\hat{\theta} = h^{-1}(\hat{\mu})$. When $h^{-1}$ does not exist (i.e., $h$ is not one-to-one), any solution of $\hat{\mu} = h(\hat{\theta})$, denoted by $\hat{\theta} = g(\hat{\mu})$, is a moment estimator of $\theta$.

By the SLLN, $\hat{\mu}_j \to_{a.s.} \mu_j$. Assume that $h$ is one-to-one and let $g = h^{-1}$. Typically, the function $g$ in $\hat{\theta} = g(\hat{\mu})$ is continuous and, therefore, $\hat{\theta}$ is strongly consistent for $\theta$. If $g$ is differentiable and $E|X_1|^{2k} < \infty$, then $\hat{\theta}$ is asymptotically normal, by the CLT and Theorem 2.11, and

$$\underline{\text{amse}}_{\hat{\theta}}(\theta) = n^{-1}\nabla g(\mu) V_\mu [\nabla g(\mu)]^\tau, \tag{3.56}$$

where $\mu = (\mu_1, ..., \mu_k)$ and $V_\mu$ is a $k \times k$ matrix whose $(i, j)$th element is $\mu_{i+j} - \mu_i\mu_j$.

**Example 3.24.** Let $X_1, ..., X_n$ be i.i.d. from a population $P_\theta$ indexed by the parameter $\theta = (\mu, \sigma^2)$, where $\mu = EX_1 \in \mathcal{R}$ and $\sigma^2 = \text{Var}(X_1) \in (0, \infty)$. This includes cases like the family of normal distributions, double exponential distributions, or logistic distributions (Table 1.2, page 20). Since $EX_1 = \mu$ and $EX_1^2 = \text{Var}(X_1) + (EX_1)^2 = \sigma^2 + \mu^2$, setting $\hat{\mu}_1 = \mu$ and $\hat{\mu}_2 = \sigma^2 + \mu^2$ we obtain the moment estimators

$$\hat{\theta} = \left(\bar{X},\ \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = \left(\bar{X},\ \frac{n-1}{n}S^2\right).$$

Note that $\bar{X}$ is unbiased, but $\frac{n-1}{n}S^2$ is not. If $X_i$ is normal, then $\hat{\theta}$ is sufficient and is nearly the same as an optimal estimator such as the UMVUE. On the other hand, if $X_i$ is from a double exponential or logistic distribution, then $\hat{\theta}$ is not sufficient and can often be improved.

Consider now the estimation of $\sigma^2$ when we know that $\mu = 0$. Obviously we cannot use the equation $\hat{\mu}_1 = \mu$ to solve the problem. Using $\hat{\mu}_2 = \mu_2 = \sigma^2$, we obtain the moment estimator $\hat{\sigma}^2 = \hat{\mu}_2 = n^{-1}\sum_{i=1}^n X_i^2$. This is still a good estimator when $X_i$ is normal, but is not a function of sufficient statistic when $X_i$ is from a double exponential distribution. For the double exponential case one can argue that we should first make a transformation $Y_i = |X_i|$ and then obtain the moment estimator based on the transformed data. The moment estimator of $\sigma^2$ based on the transformed data is $\bar{Y} = n^{-1}\sum_{i=1}^n |X_i|$, which is sufficient for $\sigma^2$. Note that this estimator can also be obtained based on absolute moment equations. ∎

**Example 3.25.** Let $X_1, ..., X_n$ be i.i.d. from the uniform distribution on $(\theta_1, \theta_2)$, $-\infty < \theta_1 < \theta_2 < \infty$. Note that

$$EX_1 = (\theta_1 + \theta_2)/2$$

and

$$EX_1^2 = (\theta_1^2 + \theta_2^2 + \theta_1\theta_2)/3.$$

Setting $\hat{\mu}_1 = EX_1$ and $\hat{\mu}_2 = EX_1^2$ and substituting $\theta_1$ in the second equation by $2\hat{\mu}_1 - \theta_2$ (the first equation), we obtain that

$$(2\hat{\mu}_1 - \theta_2)^2 + \theta_2^2 + (2\hat{\mu}_1 - \theta_2)\theta_2 = 3\hat{\mu}_2,$$

which is the same as

$$(\theta_2 - \hat{\mu}_1)^2 = 3(\hat{\mu}_2 - \hat{\mu}_1^2).$$

Since $\theta_2 \geq \hat{\mu}_1$, we obtain that

$$\hat{\theta}_2 = \hat{\mu}_1 + \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)} = \bar{X} + \sqrt{\tfrac{3(n-1)}{n}S^2}$$

and

$$\hat{\theta}_1 = \hat{\mu}_1 - \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)} = \bar{X} - \sqrt{\tfrac{3(n-1)}{n}S^2}.$$

These estimators are not functions of the sufficient and complete statistic $(X_{(1)}, X_{(n)})$. ∎

**Example 3.26.** Let $X_1, ..., X_n$ be i.i.d. from the binomial distribution $Bi(p, k)$ with unknown parameters $k \in \{1, 2, ...\}$ and $p \in (0, 1)$. Since

$$EX_1 = kp$$

and

$$EX_1^2 = kp(1-p) + k^2p^2,$$

we obtain the moment estimators

$$\hat{p} = (\hat{\mu}_1 + \hat{\mu}_1^2 - \hat{\mu}_2)/\hat{\mu}_1 = 1 - \tfrac{n-1}{n}S^2/\bar{X}$$

and

$$\hat{k} = \hat{\mu}_1^2/(\hat{\mu}_1 + \hat{\mu}_1^2 - \hat{\mu}_2) = \bar{X}/(1 - \tfrac{n-1}{n}S^2/\bar{X}).$$

The estimator $\hat{p}$ is in the range of $(0,1)$. But $\hat{k}$ may not be an integer. It can be improved by an estimator which is $\hat{k}$ rounded to the nearest positive integer. ∎

**Example 3.27.** Suppose that $X_1, ..., X_n$ are i.i.d. from the Pareto distribution $Pa(a,\theta)$ with unknown $a > 0$ and $\theta > 2$ (Table 1.2, page 20). Note that

$$EX_1 = \theta a/(\theta - 1)$$

and

$$EX_1^2 = \theta a^2/(\theta - 2).$$

From the moment equation,

$$\tfrac{(\theta-1)^2}{\theta(\theta-2)} = \hat{\mu}_2/\hat{\mu}_1^2.$$

Note that $\frac{(\theta-1)^2}{\theta(\theta-2)} - 1 = \frac{1}{\theta(\theta-2)}$. Hence

$$\theta(\theta-2) = \hat{\mu}_1^2/(\hat{\mu}_2 - \hat{\mu}_1^2).$$

Since $\theta > 2$, there is a unique solution

$$\hat{\theta} = 1 + \sqrt{\hat{\mu}_2/(\hat{\mu}_2 - \hat{\mu}_1^2)} = 1 + \sqrt{1 + \tfrac{n}{n-1}\bar{X}^2/S^2}$$

and

$$\begin{aligned} \hat{a} &= \frac{\hat{\mu}_1(\hat{\theta}-1)}{\hat{\theta}} \\ &= \bar{X}\sqrt{1 + \tfrac{n}{n-1}\bar{X}^2/S^2} \Big/ \left(1 + \sqrt{1 + \tfrac{n}{n-1}\bar{X}^2/S^2}\right). \end{aligned}$$ ∎

The method of moments can also be applied to nonparametric problems. Consider, for example, the estimation of the central moments

$$c_j = E(X_1 - \mu)^j, \qquad j = 2, ..., k.$$

Since

$$c_j = \sum_{t=0}^{j} \binom{j}{t} (-\mu)^t \mu_{j-t},$$

the moment estimator of $c_j$ is

$$\hat{c}_j = \sum_{t=0}^{j} \binom{j}{t} (-\bar{X})^t \hat{\mu}_{j-t},$$

where $\hat{\mu}_0 = 1$. It can be shown (exercise) that

$$\hat{c}_j = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^j, \qquad j = 2, ..., k, \tag{3.57}$$

which are sample central moments. From the SLLN, $\hat{c}_j$'s are strongly consistent. If $E|X_1|^{2k} < \infty$, then

$$\sqrt{n}\,(\hat{c}_2 - c_2, ..., \hat{c}_k - c_k) \to_d N_{k-1}(0, D) \tag{3.58}$$

(exercise), where the $(i, j)$th element of the $(k-1) \times (k-1)$ matrix $D$ is

$$c_{i+j+2} - c_{i+1}c_{j+1} - (i+1)c_i c_{j+2} - (j+1)c_{i+2}c_j + (i+1)(j+1)c_i c_j c_2.$$

### 3.5.3 V-statistics

Let $X_1, ..., X_n$ be i.i.d. from $P$. For every U-statistic defined in (3.11) as an estimator of $\vartheta = E[h(X_1, ..., X_m)]$, there is a closely related *V-statistic* defined by

$$V_n = \frac{1}{n^m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} h(X_{i_1}, ..., X_{i_m}). \tag{3.59}$$

As an estimator of $\vartheta$, $V_n$ is biased; but the bias is small asymptotically as the following results show. For a fixed sample size $n$, $V_n$ may be better than $U_n$ in terms of their mse's. Consider, for example, the kernel $h(x_1, x_2) = (x_1 - x_2)^2/2$ in §3.2.1, which leads to $\vartheta = \sigma^2 = \mathrm{Var}(X_1)$ and $U_n = S^2$, the sample variance. The corresponding V-statistic is

$$\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(X_i - X_j)^2}{2} = \frac{1}{n^2} \sum_{1 \le i < j \le n} (X_i - X_j)^2 = \frac{n-1}{n} S^2,$$

which is the moment estimator of $\sigma^2$ discussed in Example 3.24. In Exercise 52 in §2.6, $\frac{n-1}{n} S^2$ is shown to have a smaller mse than $S^2$ in some cases. Of course, there are situations where U-statistics are better than their corresponding V-statistics.

The following result provides orders of magnitude of the bias and variance of a V-statistic as an estimator of $\vartheta$.

**Proposition 3.5.** Let $V_n$ be defined by (3.59).
(i) Assume that $E|h(X_{i_1}, ..., X_{i_m})| < \infty$ for all $1 \le i_1 \le \cdots \le i_m \le m$. Then the bias of $V_n$ satisfies

$$b_{V_n}(P) = O(n^{-1}).$$

(ii) Assume that $E[h(X_{i_1}, ..., X_{i_m})]^2 < \infty$ for all $1 \le i_1 \le \cdots \le i_m \le m$. Then the variance of $V_n$ satisfies

$$\mathrm{Var}(V_n) = \mathrm{Var}(U_n) + O(n^{-2}),$$

where $U_n$ is given by (3.11).
**Proof.** (i) Note that

$$U_n - V_n = \left[1 - \tfrac{n!}{n^m(n-m)!}\right](U_n - W_n), \tag{3.60}$$

where $W_n$ is the average of all terms $h(X_{i_1}, ..., X_{i_m})$ with at least one equality $i_m = i_l$, $m \ne l$. The result follows from $E(U_n - W_n) = O(1)$.
(ii) The result follows from $E(U_n - W_n)^2 = O(1)$, $E[W_n(U_n - \vartheta)] = O(n^{-1})$ (exercise), and (3.60). ∎

To study the asymptotic behavior of a V-statistic, we consider the following representation of $V_n$ in (3.59):

$$V_n = \sum_{j=1}^{m} \binom{m}{j} V_{nj},$$

where

$$V_{nj} = \vartheta + \frac{1}{n^j} \sum_{i_1=1}^{n} \cdots \sum_{i_j=1}^{n} g_j(X_{i_1}, ..., X_{i_j})$$

is a "V-statistic" with

$$\begin{aligned} g_j(x_1, ..., x_j) = h_j(x_1, ..., x_j) &- \sum_{i=1}^{j} \int h_j(x_1, ..., x_j) dP(x_i) \\ &+ \sum_{1 \le i_1 < i_2 \le j} \int\int h_j(x_1, ..., x_j) dP(x_{i_1}) dP(x_{i_2}) - \cdots \\ &+ (-1)^j \int \cdots \int h_j(x_1, ..., x_j) dP(x_1) \cdots dP(x_j) \end{aligned}$$

and $h_j(x_1, ..., x_j) = E[h(x_1, ..., x_j, X_{j+1}, ..., X_m)]$. Using a similar argument to the proof of Theorem 3.4, we can show (exercise) that

$$E(V_{nj})^2 = O(n^{-j}), \qquad j = 1, ..., m, \tag{3.61}$$

provided that $E[h(X_{i_1},...,X_{i_m})]^2 < \infty$ for all $1 \le i_1 \le \cdots \le i_m \le m$. Thus,

$$V_n - \vartheta = mV_{n1} + \tfrac{m(m-1)}{2} V_{n2} + o_p(n^{-1}),$$

which leads to the following result similar to Theorem 3.5.

**Theorem 3.16.** Let $V_n$ be given by (3.59) with $E[h(X_{i_1},...,X_{i_m})]^2 < \infty$ for all $1 \le i_1 \le \cdots \le i_m \le m$.
(i) If $\zeta_1 = \mathrm{Var}(h_1(X_1)) > 0$, then

$$\sqrt{n}(V_n - \vartheta) \to_d N(0, m^2\zeta_1).$$

(ii) If $\zeta_1 = 0$ but $\zeta_2 = \mathrm{Var}(h_2(X_1, X_2)) > 0$, then

$$n(V_n - \vartheta) \to_d \frac{m(m-1)}{2} \sum_{j=1}^{\infty} \lambda_j \chi^2_{1j},$$

where $\chi^2_{1j}$'s and $\lambda_j$'s are the same as those in (3.21). ∎

Theorem 3.16 indicates that if $\zeta_1 > 0$, then the asymptotic biases and amse's of $U_n$ and $V_n$ are the same. If $\zeta_1 = 0$ but $\zeta_2 > 0$, then a similar argument to that in the proof of Lemma 3.2 leads to

$$\begin{aligned} \underline{\mathrm{amse}}_{V_n}(P) &= \frac{m^2(m-1)^2\zeta_2}{2n^2} + \frac{m^2(m-1)^2}{2n^2}\left(\sum_{j=1}^{\infty}\lambda_j\right)^2 \\ &= \underline{\mathrm{amse}}_{U_n}(P) + \frac{m^2(m-1)^2}{2n^2}\left(\sum_{j=1}^{\infty}\lambda_j\right)^2 \end{aligned}$$

(see Lemma 3.2). Hence $U_n$ is asymptotically more efficient than $V_n$, unless $\sum_{j=1}^{\infty} \lambda_j = 0$. Technically, the proof of the asymptotic results for $V_n$ also requires moment conditions stronger than those for $U_n$.

**Example 3.28.** Consider the estimation of $\mu^2$, where $\mu = EX_1$. From the results in §3.2, the U-statistic $U_n = \frac{1}{n(n-1)} \sum_{1 \le i < j \le n} X_i X_j$ is unbiased for $\mu^2$. The corresponding V-statistic is simply $V_n = \bar{X}^2$. If $\mu \ne 0$, then $\zeta_1 \ne 0$ and the asymptotic relative efficiency of $V_n$ w.r.t. $U_n$ is 1. If $\mu = 0$, then

$$nV_n \to_d \sigma^2\chi^2_1 \qquad \text{and} \qquad nU_n \to_d \sigma^2(\chi^2_1 - 1),$$

where $\chi^2_1$ is a random variable having the chi-square distribution $\chi^2_1$. Hence the asymptotic relative efficiency of $V_n$ w.r.t. $U_n$ is

$$E(\chi^2_1 - 1)^2 / E(\chi^2_1)^2 = 2/3. \quad ∎$$

### 3.5.4 The weighted LSE

In linear model (3.25), the unbiased LSE of $\beta l^\tau$ may be improved by a slightly biased estimator when $\mathrm{Var}(\varepsilon)$ is not $\sigma^2 I_n$ and the LSE is not BLUE.

Assume that $Z$ in (3.25) is of full rank so that every $\beta l^\tau$ is estimable. For simplicity, let us denote $\mathrm{Var}(\varepsilon)$ by $V$. If $V$ is known, then the BLUE of $\beta l^\tau$ is $\breve{\beta} l^\tau$, where

$$\breve{\beta} = XV^{-1}Z(Z^\tau V^{-1}Z)^{-1} \tag{3.62}$$

(see the discussion after the statement of assumption A3 in §3.3.1). If $V$ is unknown and $\hat{V}$ is an estimator of $V$, then an application of the substitution principle leads to a *weighted least squares estimator*

$$\hat{\beta}_w = X\hat{V}^{-1}Z(Z^\tau \hat{V}^{-1}Z)^{-1}. \tag{3.63}$$

The weighted LSE is not linear in $X$ and not necessarily unbiased for $\beta$. It is unbiased if $-\varepsilon$ and $\varepsilon$ have the same distribution, $E[\lambda_+(\hat{V})]^2 < \infty$, and $\hat{V} = u(\varepsilon)$ for some function $u$ satisfying $u(-\varepsilon) = u(\varepsilon)$. In such a case the LSE $\hat{\beta} l^\tau$ may not be a UMVUE, since $\hat{\beta}_w l^\tau$ may be better than $\hat{\beta} l^\tau$.

Asymptotic properties of the weighted LSE depend on the asymptotic behavior of $\hat{V}$. We say that $\hat{V}$ is consistent for $V$ if and only if

$$\|\hat{V}^{-1}V - I_n\| \to_p 0, \tag{3.64}$$

where $\|A\| = [\mathrm{tr}(A^\tau A)]^{1/2}$ for a matrix $A$.

**Theorem 3.17.** Consider model (3.25) with a full rank $Z$. Let $\breve{\beta}$ and $\hat{\beta}_w$ be defined by (3.62) and (3.63), respectively, with a $\hat{V}$ consistent in the sense of (3.64). Assume the conditions in Theorem 3.12. Then

$$(\hat{\beta}_w l^\tau - \beta l^\tau)/a_n \to_d N(0,1),$$

where $l \in \mathcal{R}^p$, $l \neq 0$, and

$$a_n^2 = \mathrm{Var}(\breve{\beta} l^\tau) = l(Z^\tau V^{-1}Z)^{-1} l^\tau.$$

**Proof.** Using the same argument as in the proof of Theorem 3.12, we obtain that

$$(\breve{\beta} l^\tau - \beta l^\tau)/a_n \to_d N(0,1).$$

By Slutsky's theorem, the result follows from

$$\hat{\beta}_w l^\tau - \breve{\beta} l^\tau = o_p(a_n).$$

Note that

$$\begin{aligned} \hat{\beta}_w l^\tau - \breve{\beta} l^\tau &= \varepsilon\hat{V}^{-1}Z(Z^\tau\hat{V}^{-1}Z)^{-1}l^\tau - \varepsilon V^{-1}Z(Z^\tau V^{-1}Z)^{-1}l^\tau \\ &= \varepsilon(\hat{V}^{-1} - V^{-1})Z(Z^\tau\hat{V}^{-1}Z)^{-1}l^\tau \qquad (3.65) \\ &\quad + \varepsilon V^{-1}Z[(Z^\tau\hat{V}^{-1}Z)^{-1} - (Z^\tau V^{-1}Z)^{-1}]l^\tau. \qquad (3.66) \end{aligned}$$

Let $\xi_n$ be the term in (3.65) and $A_n = V\hat{V}^{-1} - I_n$. Using inequality (1.34), we obtain that

$$\begin{aligned}
\xi_n^2 &= [\varepsilon V^{-1/2}V^{-1/2}A_n Z(Z^\tau \hat{V}^{-1}Z)^{-1}l^\tau]^2 \\
&\leq \varepsilon V^{-1}\varepsilon^\tau l(Z^\tau \hat{V}^{-1}Z)^{-1}Z^\tau A_n^\tau V^{-1}A_n Z(Z^\tau \hat{V}^{-1}Z)^{-1}l^\tau \\
&\leq O_p(1)o_p(a_n^2),
\end{aligned}$$

since $\|A_n\| = o_p(1)$ by condition (3.64). This proves that $\xi_n = o_p(a_n)$.

Let $\zeta_n$ be the term in (3.66), $B_n = Z^\tau V^{-1}Z(Z^\tau \hat{V}^{-1}Z)^{-1} - I_p$, and $C_n = \hat{V}V^{-1} - I_n$. By (3.64), $\|C_n\| = o_p(1)$. Then

$$\begin{aligned}
\|B_n\|^2 &= \|Z^\tau \hat{V}^{-1}C_n Z(Z^\tau \hat{V}^{-1}Z)^{-1}\|^2 \\
&= \operatorname{tr}\left((Z^\tau \hat{V}^{-1}Z)^{-1}Z^\tau C_n^\tau \hat{V}^{-1}ZZ^\tau \hat{V}^{-1}C_n Z(Z^\tau \hat{V}^{-1}Z)^{-1}\right) \\
&\leq \|C_n\| \operatorname{tr}\left(\hat{V}^{-1}ZZ^\tau \hat{V}^{-1}C_n Z(Z^\tau \hat{V}^{-1}Z)^{-1}(Z^\tau \hat{V}^{-1}Z)^{-1}Z^\tau\right) \\
&\leq \|C_n\|^2 \operatorname{tr}\left(Z(Z^\tau \hat{V}^{-1}Z)^{-1}(Z^\tau \hat{V}^{-1}Z)^{-1}Z^\tau \hat{V}^{-1}ZZ^\tau \hat{V}^{-1}\right) \\
&= o_p(1)\operatorname{tr}(I_p) \\
&= o_p(1).
\end{aligned}$$

Note that (exercise)

$$\varepsilon V^{-1}Z(Z^\tau V^{-1}Z)^{-1}l^\tau = O_p(a_n). \tag{3.67}$$

Then

$$\begin{aligned}
\zeta_n^2 &= [\varepsilon V^{-1}Z(Z^\tau V^{-1}Z)^{-1}B_n l^\tau]^2 \\
&\leq \|\varepsilon V^{-1}Z(Z^\tau V^{-1}Z)^{-1}\|^2 \|B_n l^\tau\|^2 \\
&= O_p(a_n^2)o_p(1).
\end{aligned}$$

This shows that $\zeta_n = o_p(a_n)$ and thus completes the proof. ∎

Theorem 3.17 shows that as long as $\hat{V}$ is consistent in the sense of (3.64), the weighted LSE $\hat{\beta}_w$ is asymptotically as efficient as $\breve{\beta}$, which is the BLUE if $V$ is known. If $V$ is known and $\varepsilon$ is normal, then $\mathrm{Var}(\breve{\beta}l^\tau)$ attains the Cramér-Rao lower bound (Theorem 3.7(iii)) and, thus, (3.10) holds with $T_n = \hat{\beta}_w l^\tau$.

By Theorems 3.12 and 3.17, the asymptotic relative efficiency of the LSE $\hat{\beta}l^\tau$ w.r.t. the weighted LSE $\hat{\beta}_w l^\tau$ is

$$\frac{l(Z^\tau V^{-1}Z)^{-1}l^\tau}{l(Z^\tau Z)^{-1}Z^\tau VZ(Z^\tau Z)^{-1}l^\tau},$$

which is always less than 1 and equals 1 if $\hat{\beta}l^\tau$ is a BLUE (in which case $\hat{\beta} = \breve{\beta}$).

Finding a consistent $\hat{V}$ is possible only when $V$ has certain structure. We consider two examples.

**Example 3.29.** Suppose that $V$ is a block diagonal matrix with the $i$th diagonal block

$$\sigma^2 I_{m_i} + U_i \Sigma U_i^\tau, \qquad i = 1, ..., k, \tag{3.68}$$

where $m_i$'s are integers bounded by a fixed integer $m$, $\sigma^2 > 0$ is an unknown parameter, $\Sigma$ is a $q \times q$ unknown nonnegative definite matrix, $U_i$ is an $m_i \times q$ full rank matrix whose columns are in $\mathcal{R}(W_i^\tau)$, $q < \inf_i m_i$, and $W_i$ is the $i$th block of $Z = (W_1^\tau, ..., W_k^\tau)^\tau$. Under (3.68), a consistent $\hat{V}$ can be obtained if we can obtain consistent estimators of $\sigma^2$ and $\Sigma$.

Let $X = (Y_1, ..., Y_k)$, where $Y_i$ is $m_i \times 1$, and let $R_i$ be the matrix containing linearly independent columns of $W_i$. Then

$$\hat{\sigma}^2 = \frac{1}{n - kq} \sum_{i=1}^k Y_i [I_{m_i} - R_i (R_i^\tau R_i)^{-1} R_i^\tau] Y_i^\tau \tag{3.69}$$

is an unbiased estimator of $\sigma^2$. Assume that $Y_i$'s are independent and that $\sup_i E|\varepsilon_i|^{2+\delta} < \infty$ for some $\delta > 0$. Then $\hat{\sigma}^2$ is consistent for $\sigma^2$ (exercise). A consistent estimator of $\Sigma$ is then (exercise)

$$\hat{\Sigma} = \frac{1}{k} \sum_{i=1}^k \left[ (U_i^\tau U_i)^{-1} U_i^\tau r_i^\tau r_i U_i (U_i^\tau U_i)^{-1} - \hat{\sigma}^2 (U_i^\tau U_i)^{-1} \right], \tag{3.70}$$

where $r_i = Y_i - \hat{\beta} W_i^\tau$. ∎

**Example 3.30.** Suppose that $V$ is diagonal with the $i$th diagonal element $\sigma_i^2 = \psi(Z_i)$, where $\psi$ is an unknown function. The simplest case is $\psi(t) = \theta_0 + \theta_1 v(Z_i)$ for a known function $v$ and some unknown $\theta_0$ and $\theta_1$. One can then obtain a consistent estimator $\hat{V}$ by using the LSE of $\theta_0$ and $\theta_1$ under the "model"

$$r_i^2 = \theta_0 + \theta_1 v(Z_i), \qquad i = 1, ..., n, \tag{3.71}$$

where $r_i = X_i - \hat{\beta} Z_i^\tau$ (exercise). If $\psi$ is nonlinear or nonparametric, some results are given in Carroll (1982) and Müller and Stadrmüller (1987). ∎

Finally, if $\hat{V}$ is not consistent (i.e., (3.64) does not hold), then the weighted LSE $\hat{\beta}_w l^\tau$ can still be consistent and asymptotically normal, but its asymptotic variance is not $l(Z^\tau V^{-1} Z)^{-1} l^\tau$; in fact, $\hat{\beta}_w l^\tau$ may not be asymptotically as efficient as the LSE $\hat{\beta} l^\tau$ (Carroll and Cline, 1988; Chen and Shao 1993).

## 3.6 Exercises

1. Let $X_1, ..., X_n$ be i.i.d. binary random variables with $P(X_i = 1) = p \in (0,1)$.
   (a) Find the UMVUE of $p^m$, $m \leq n$.
   (b) Find the UMVUE of $P(X_1 + \cdots + X_m = k)$, where $m$ and $k$ are positive integers $\leq n$.
   (c) Find the UMVUE of $P(X_1 + \cdots + X_{n-1} > X_n)$.

2. Let $X_1, ..., X_n$ be i.i.d. having the Poisson distribution $P(\theta)$ with $\theta > 0$. Find the UMVUE of $e^{-t\theta}$ with a fixed $t > 0$.

3. Let $X_1, ..., X_n$ be i.i.d. having the $N(\mu, \sigma^2)$ distribution with an unknown $\mu \in \mathcal{R}$ and a known $\sigma^2 > 0$.
   (a) Find the UMVUE's of $\mu^3$ and $\mu^4$.
   (b) Find the UMVUE's of $P(X_1 \leq t)$ and $\frac{d}{dt}P(X_1 \leq t)$ with a fixed $t \in \mathcal{R}$.

4. In Example 3.4,
   (a) show that the UMVUE of $\sigma^r$ is $k_{n-1,r}S^r$, where $r > 1 - n$;
   (b) prove that $(X_1 - \bar{X})/S$ has the p.d.f. given by (3.1);
   (c) show that $(X_1 - \bar{X})/S \to_d N(0,1)$ by using (i) the SLLN and (ii) Scheffé's theorem (Proposition 1.17).

5. Let $X_1, ..., X_m$ be i.i.d. having the $N(\mu_x, \sigma_x^2)$ distribution and let $Y_1, ..., Y_n$ be i.i.d. having the $N(\mu_y, \sigma_y^2)$ distribution. Assume that $X_i$'s and $Y_j$'s are independent.
   (a) Assume that $\mu_x \in \mathcal{R}$, $\mu_y \in \mathcal{R}$, $\sigma_x^2 > 0$, and $\sigma_y^2 > 0$. Find the UMVUE's of $\mu_x - \mu_y$ and $(\sigma_x/\sigma_y)^r$, $r > 0$.
   (b) Assume that $\mu_x \in \mathcal{R}$, $\mu_y \in \mathcal{R}$, and $\sigma_x^2 = \sigma_y^2 > 0$. Find the UMVUE's of $\sigma_x^2$ and $(\mu_x - \mu_y)/\sigma_x$.
   (c) Assume that $\mu_x = \mu_y \in \mathcal{R}$, $\sigma_x^2 > 0$, $\sigma_y^2 > 0$, and $\sigma_x^2/\sigma_y^2 = \gamma$ is known. Find the UMVUE of $\mu_x$.
   (d) Assume that $\mu_x = \mu_y \in \mathcal{R}$, $\sigma_x^2 > 0$, and $\sigma_y^2 > 0$. Show that a UMVUE of $\mu_x$ does not exist.
   (e) Assume that $\mu_x = \mu_y \in \mathcal{R}$, $\sigma_x^2 > 0$, and $\sigma_y^2 > 0$. Find the UMVUE of $P(X_1 \leq Y_1)$.
   (f) Repeat (e) under the assumption that $\sigma_x = \sigma_y$.

6. Let $X_1, ..., X_n$ be i.i.d. having the uniform distribution on the interval $(\theta_1 - \theta_2, \theta_1 + \theta_2)$, where $\theta_j \in \mathcal{R}$, $j = 1, 2$. Find the UMVUE's of $\theta_j$, $j = 1, 2$, and $\theta_1/\theta_2$.

7. Let $X_1, ..., X_n$ be i.i.d. having the exponential distribution $E(a, \theta)$ with parameters $\theta > 0$ and $a \in \mathcal{R}$.
   (a) Find the UMVUE of $a$ when $\theta$ is known.

(b) Find the UMVUE of $\theta$ when $a$ is known.
(c) Find the UMVUE's of $\theta$ and $a$.
(d) Assume that $\theta$ is known. Find the UMVUE of $P(X_1 \geq t)$ and $\frac{d}{dt}P(X_1 \geq t)$ for a fixed $t > 0$.
(e) Find the UMVUE of $P(X_1 \geq t)$ for a fixed $t > 0$.

8. Let $X_1, ..., X_n$ be i.i.d. having the Pareto distribution $Pa(a, \theta)$ with $\theta > 0$ and $a > 0$.
(a) Find the UMVUE of $\theta$ when $a$ is known.
(b) Find the UMVUE of $a$ when $\theta$ is known.
(c) Find the UMVUE's of $a$ and $\theta$.

9. Consider Exercise 41(a) of §2.6. Find the UMVUE of $\gamma$.

10. Let $X_1, ..., X_m$ be i.i.d. having the exponential distribution $E(a_x, \theta_x)$ with $\theta_x > 0$ and $a_x \in \mathcal{R}$ and $Y_1, ..., Y_n$ be i.i.d. having the exponential distribution $E(a_y, \theta_y)$ with $\theta_y > 0$ and $a_y \in \mathcal{R}$. Assume that $X_i$'s and $Y_j$'s are independent.
(a) Find the UMVUE's of $a_x - a_y$ and $\theta_x/\theta_y$.
(b) Suppose that $\theta_x = \theta_y$ but it is unknown. Find the UMVUE's of $\theta_x$ and $(a_x - a_y)/\theta_x$.
(c) Suppose that $a_x = a_y$ but it is unknown. Show that a UMVUE of $a_x$ does not exist.
(d) Suppose that $n = m$ and $a_x = a_y = 0$ and that our sample is $(Z_1, \Delta_1), ..., (Z_n, \Delta_n)$, where $Z_i = \min(X_i, Y_i)$ and $\Delta_i = 1$ if $X_i \geq Y_i$ and 0 otherwise, $i = 1, ..., n$. Find the UMVUE of $\theta_x - \theta_y$.

11. Let $X_1, ..., X_m$ be i.i.d. having the uniform distribution $U(0, \theta_x)$ and $Y_1, ..., Y_n$ be i.i.d. having the uniform distribution $U(0, \theta_y)$. Suppose that $X_i$'s and $Y_j$'s are independent and that $\theta_x > 0$ and $\theta_y > 0$. Find the UMVUE of $\theta_x/\theta_y$ when $n > 1$.

12. Let $X$ be a random variable having the negative binomial distribution $NB(p, r)$ with an unknown $p \in (0, 1)$ and a known $r$.
(a) Find the UMVUE of $p^t$, $t < r$.
(b) Find the UMVUE of $\mathrm{Var}(X)$.
(c) Find the UMVUE of $\log p$.

13. Let $X_1, ..., X_n$ be i.i.d. random variables having the Poisson distribution $P(\theta)$ truncated at 0, i.e., $P(X_i = x) = (e^\theta - 1)^{-1}\theta^x/x!$, $x = 1, 2, ..., \theta > 0$. Find the UMVUE of $\theta$ when $n = 1, 2$.

14. Let $X$ be a random variable having the negative binomial distribution $NB(p, r)$ truncated at $r$, where $r$ is known and $p \in (0, 1)$ is unknown. Let $k$ be a fixed positive integer $> r$.
(a) For $r = 1, 2, 3$, find the UMVUE of $p^k$.
(b) For $r = 1, 2, 3$, find the UMVUE of $P(X = k)$.

15. Let $X_1, ..., X_n$ be i.i.d. having the log-distribution $L(p)$ with an unknown $p \in (0,1)$. Let $k$ be a fixed positive integer.
(a) For $n = 1, 2, 3$, find the UMVUE of $p^k$.
(b) For $n = 1, 2, 3$, find the UMVUE of $P(X = k)$.

16. Suppose that $(X_0, X_1, ..., X_k)$ has the multinomial distribution in Example 2.7 with $p_i \in (0,1)$, $\sum_{j=0}^{k} p_j = 1$. Find the UMVUE of $p_0^{r_0} \cdots p_k^{r_k}$, where $r_j$'s are nonnegative integers with $r_0 + \cdots + r_k \leq n$.

17. Let $X_1, ..., X_n$ be i.i.d. from $P \in \mathcal{P}$ containing all symmetric c.d.f.'s with finite means and with Lebesgue p.d.f.'s on $\mathcal{R}$. Show that there is no UMVUE of $\mu = EX_1$.

18. Let $(X_1, Y_1), ..., (X_n, Y_n)$ be i.i.d. random 2-vectors from a population $P \in \mathcal{P}$ which is the family of all bivariate populations with Lebesgue p.d.f.'s.
(a) Show that the set of $n$ pairs $(X_i, Y_i)$ ordered according to the value of their first coordinate constitute a sufficient and complete statistic for $P \in \mathcal{P}$.
(b) A statistic $T$ is a function of the complete and sufficient statistic if and only if $T$ is invariant under permutation of the $n$ pairs.
(c) Show that $(n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$ is the UMVUE of $\mathrm{Cov}(X_1, Y_1)$.
(d) Find the UMVUE's of $P(X_i \leq Y_i)$ and $P(X_i \leq X_j \text{ and } Y_i \leq Y_j)$, $i \neq j$.

19. Prove Corollary 3.1.

20. Consider the problem in Exercise 68 of §2.6. Use Theorem 3.2 to show that $I_{\{0\}}(X)$ is a UMVUE of $(1-p)^2$ and that there is no UMVUE of $p$.

21. Let $X_1, ..., X_n$ be i.i.d. from a discrete distribution with

$$P(X_i = \theta - 1) = P(X_i = \theta) = P(X_i = \theta + 1) = \tfrac{1}{3},$$

where $\theta$ is an unknown integer. Show that no nonconstant function of $\theta$ has a UMVUE.

22. Let $X$ be a random variable having the Lebesgue p.d.f.

$$[(1-\theta) + \theta/(2\sqrt{x})] I_{(0,1)}(x),$$

where $\theta \in [0,1]$. Show that there is no UMVUE of $\theta$.

23. Let $X$ be a discrete random variable with $P(X = -1) = 2p(1-p)$ and $P(X = k) = p^k(1-p)^{3-k}$, $k = 0, 1, 2, 3$, where $p \in (0,1)$.
(a) Determine whether there is a UMVUE of $p$.
(b) Determine whether there is a UMVUE of $p(1-p)$.

24. Let $X_1, ..., X_n$ be i.i.d. having the exponential distribution $E(a, \theta)$ with a known $\theta$ and an unknown $a \leq 0$. Obtain a UMVUE of $a$.

25. Let $X_1, ..., X_n$ be i.i.d. having the Pareto distribution $Pa(a, \theta)$ with a known $\theta > 1$ and an unknown $a \in (0, 1]$. Obtain a UMVUE of $a$.

26. Prove Theorem 3.3 for the multivariate case ($k > 1$).

27. Let $X$ be a single sample from $P_\theta$. Find the Fisher information $I(\theta)$ in the following cases.
(a) $P_\theta$ is the $N(\mu, \sigma^2)$ distribution with $\theta = \mu \in \mathcal{R}$.
(b) $P_\theta$ is the $N(\mu, \sigma^2)$ distribution with $\theta = \sigma^2 > 0$.
(c) $P_\theta$ is the $N(\mu, \sigma^2)$ distribution with $\theta = \sigma > 0$.
(d) $P_\theta$ is the $N(\sigma, \sigma^2)$ distribution with $\theta = \sigma > 0$.
(e) $P_\theta$ is the $N(\mu, \sigma^2)$ distribution with $\theta = (\mu, \sigma^2) \in \mathcal{R} \times (0, \infty)$.
(f) $P_\theta$ is negative binomial distribution $NB(\theta, r)$ with $\theta \in (0, 1)$.
(g) $P_\theta$ is the gamma distribution $\Gamma(\alpha, \gamma)$ with $\theta = (\alpha, \gamma) \in (0, \infty) \times (0, \infty)$;
(h) $P_\theta$ is the beta distribution $B(\alpha, \beta)$ with $\theta = (\alpha, \beta) \in (0, 1) \times (0, 1)$.

28. Find a function of $\theta$ for which the amount of information is independent of $\theta$, when $P_\theta$ is
(a) the Poisson distribution $P(\theta)$ with $\theta > 0$;
(b) the binomial distribution $Bi(\theta, r)$ with $\theta \in (0, 1)$;
(c) the gamma distribution $\Gamma(\alpha, \theta)$ with $\theta > 0$.

29. Prove the result in Example 3.9. Show that if $\mu$ (or $\sigma$) is known, then $I_1(\mu)$ (or $I_2(\sigma)$) is the first (or second) diagonal element of $I(\theta)$.

30. Obtain the Fisher information matrix for
(a) the Cauchy distribution $C(\mu, \sigma)$, $\mu \in \mathcal{R}$, $\sigma > 0$;
(b) the double exponential distribution $DE(\mu, \theta)$, $\mu \in \mathcal{R}$, $\theta > 0$;
(c) the logistic distribution $LG(\mu, \sigma)$, $\mu \in \mathcal{R}$, $\sigma > 0$;
(d) $F_r\left(\frac{x-\mu}{\sigma}\right)$, where $F_r$ is the c.d.f. of the t-distribution $t_r$ with a known $r$, $\mu \in \mathcal{R}$, $\sigma > 0$.

31. Let $\phi$ be the standard normal p.d.f. Find the Fisher information contained in $X$ which has the Lebesgue p.d.f.
$$f_\theta(x) = (1 - \epsilon)\phi(x - \mu) + \tfrac{\epsilon}{\sigma}\phi\left(\tfrac{x-\mu}{\sigma}\right),$$
$\theta = (\mu, \sigma, \epsilon) \in \mathcal{R} \times (0, \infty) \times (0, 1)$.

32. Let $X_1, ..., X_n$ be i.i.d. from the uniform distribution $U(0, \theta)$ with $\theta > 0$.
(a) Show that condition (3.3) does not hold for $h(X) = X_{(n)}$.
(b) Show that the inequality (3.6) does not apply to the UMVUE of $\theta$.

33. Prove Proposition 3.3.

34. Let $X$ be a single sample from the double exponential distribution $DE(\mu, \theta)$ with $\mu = 0$ and $\theta > 0$. Find the UMVUE's of the following parameters and, in each case, determine whether the variance of the UMVUE attains the Cramér-Rao lower bound.
(a) $\vartheta = \theta$;
(b) $\vartheta = \theta^r$, where $r > 1$;
(c) $\vartheta = (1 + \theta)^{-1}$.

35. Let $X_1, ..., X_n$ be i.i.d. binary random variables with $P(X_i = 1) = p \in (0, 1)$.
(a) Show that the UMVUE of $p(1 - p)$ is $T_n = n\bar{X}(1 - \bar{X})/(n - 1)$.
(b) Show that $\mathrm{Var}(T_n)$ does not attain the Cramér-Rao lower bound.
(c) Show that (3.10) holds.

36. Let $X_1, ..., X_n$ be i.i.d. having the Poisson distribution $P(\theta)$ with $\theta > 0$. Find the amse of the UMVUE of $e^{-t\theta}$ with a fixed $t > 0$ and show that (3.10) holds.

37. Let $X_1, ..., X_n$ be i.i.d. having the $N(\mu, \sigma^2)$ distribution with an unknown $\mu \in \mathcal{R}$ and a known $\sigma^2 > 0$.
(a) Find the UMVUE of $\vartheta = e^{t\mu}$ with a fixed $t \neq 0$.
(b) Determine whether the variance of the UMVUE in (a) attains the Cramér-Rao lower bound.
(c) Show that (3.10) holds.

38. Show that if $X_1, ..., X_n$ are i.i.d. binary random variables, $U_n$ in (3.12) equals $T(T - 1) \cdots (T - m + 1)/[n(n - 1) \cdots (n - m + 1)]$, where $T = \sum_{i=1}^n X_i$.

39. Show that if $T_n = \bar{X}$, then $U_n$ in (3.13) is the same as the sample variance $S^2$ in (2.2). Show that (3.23) holds for $T_n$ given by (3.22) with $E(R_n^2) = o(n^{-1})$.

40. Prove (3.14) and (3.17).

41. Let $\zeta_k$ be given in Theorem 3.4. Show that $\zeta_1 \leq \zeta_2 \leq \cdots \leq \zeta_m$.

42. Prove Corollary 3.2.

43. Prove (3.20) and show that $U_n - \check{U}_n$ is also a U-statistic.

44. Let $T_n$ be a symmetric statistic with $\mathrm{Var}(T_n) < \infty$ for every $n$ and $\check{T}_n$ be the projection of $T_n$ on $\binom{n}{k}$ random vectors $\{X_{i_1}, ..., X_{i_k}\}$, $1 \leq i_1 < \cdots < i_k \leq n$. Show that $E(T_n) = E(\check{T}_n)$ and calculate $E(T_n - \check{T}_n)^2$.

45. Let $Y_k$ be defined in Lemma 3.2. Show that $\{Y_k^2\}$ is uniformly integrable.

46. Show that (3.22) with $E(R_n^2) = o(n^{-1})$ is satisfied for $T_n$ being a U-statistic with $E[h(X_1, ..., X_m)]^2 < \infty$.

47. Let $S^2$ be the sample variance given by (2.2), which is also a U-statistic (§3.2.1). Find the corresponding $h_1$, $h_2$, $\zeta_1$, and $\zeta_2$. Discuss how to apply Theorem 3.5 to this case.

48. Let $h(x_1, x_2, x_3) = I_{(-\infty,0)}(x_1 + x_2 + x_3)$. Define the U-statistic with this kernel and find $h_k$ and $\zeta_k$, $k = 1, 2, 3$.

49. Show that any $\hat{\beta}$ given by (3.29) is an LSE of $\beta$.

50. Obtain explicit forms for the LSE's of $\beta_j$, $j = 0, 1$, and $SSR$, under the simple linear regression model in Example 3.11, assuming that some $t_i$'s are different.

51. Consider the polynomial model

$$X_i = \beta_0 + \beta_1 t_i + \beta_2 t_i^2 + \varepsilon_i, \quad i = 1, ..., n.$$

Find explicit forms for the LSE's of $\beta_j$, $j = 0, 1, 2$, and $SSR$, assuming that some $t_i$'s are different.

52. Suppose that

$$X_{ij} = \alpha_i + \beta t_{ij} + \varepsilon_{ij}, \quad i = 1, ..., a,\ j = 1, ..., b.$$

Find explicit forms for the LSE's of $\beta$, $\alpha_i$, $i = 1, ..., a$, and $SSR$.

53. Find the matrix $Z$, $Z^\tau Z$, and the form of $l \in \mathcal{R}(Z)$ under the one-way ANOVA model (3.31).

54. Obtain the matrix $Z$ under the two-way balanced ANOVA model (3.32). Show that the rank of $Z$ is $ab$. Verify the form of the LSE of $\beta$ given in Example 3.14. Find the form of $l \in \mathcal{R}(Z)$.

55. Consider the following model as a special case of model (3.25):

$$X_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}, \quad i = 1, ..., a, j = 1, ..., b, k = 1, ..., c.$$

Obtain the matrix $Z$, the parameter vector $\beta$, and the form of LSE's of $\beta$. Discuss conditions under which $l \in \mathcal{R}(Z)$.

56. Under model (3.25) and assumption A1, find the UMVUE's of $(\beta l^\tau)^2$, $\beta l^\tau/\sigma$, and $(\beta l^\tau/\sigma)^2$ for an estimable $\beta l^\tau$.

57. Verify the formulas for $SSR$'s in Example 3.15.

58. Consider model (3.25) with assumption A2. Show that $\mathrm{Var}(\hat{\beta}l^\tau) = \sigma^2 l(Z^\tau Z)^- l^\tau$ for $l \in \mathcal{R}(Z)$.

59. Consider the one-way random effects model in Example 3.17. Assume that $n_i = n$ for all $i$ and that $A_i$'s and $e_{ij}$'s are normally distributed. Show that the family of populations is an exponential family with sufficient and complete statistics $\bar{X}_{\cdot\cdot}$, $S_A = n\sum_{i=1}^m (\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$, and $S_E = \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{X}_{i\cdot})^2$. Find the UMVUE's of $\mu$, $\sigma_a^2$, and $\sigma^2$.

60. Consider model (3.25). Suppose that $\varepsilon_i$'s are i.i.d. with a Lebesgue p.d.f. $\sigma^{-1}f(x/\sigma)$, where $f$ is a known Lebesgue p.d.f. and $\sigma > 0$ is unknown.
(a) Show that $X$ is from a subfamily of the location-scale family given by (2.10).
(b) Find the Fisher information about $(\beta, \sigma)$ contained in $X_i$.
(c) Find the Fisher information about $(\beta, \sigma)$ contained in $X$.

61. Consider model (3.25) with assumption A2. Let $c \in \mathcal{R}^p$. Show that if the equation $c = yZ^\tau$ has a solution, then there is a unique solution $y_0 \in \mathcal{R}(Z)$ such that $\mathrm{Var}(Xy_0^\tau) \le \mathrm{Var}(Xy^\tau)$ for any other solution of $c = yZ^\tau$.

62. Consider model (3.25). Show that the number of independent linear functions of $X$ with mean 0 is $n - r$, where $r$ is the rank of $Z$.

63. Consider model (3.25) with assumption A2. Let $\hat{X}_i = \hat{\beta}Z_i^\tau$, which is called the least squares *prediction* of $X_i$. Let $h_{ij}$ be the $(i,j)$th element of $Z(Z^\tau Z)^- Z^\tau$. Show that
(a) $\mathrm{Var}(\hat{X}_i) = \sigma^2 h_{ii}$;
(b) $\mathrm{Var}(X_i - \hat{X}_i) = \sigma^2(1 - h_{ii})$;
(c) $\mathrm{Cov}(\hat{X}_i, \hat{X}_j) = \sigma^2 h_{ij}$;
(d) $\mathrm{Cov}(X_i - \hat{X}_i, X_j - \hat{X}_j) = -\sigma^2 h_{ij}$, $i \ne j$;
(e) $\mathrm{Cov}(\hat{X}_i, X_j - \hat{X}_j) = 0$.

64. Prove that (e) implies (b) in Theorem 3.10.

65. Show that (a) in Theorem 3.10 is equivalent to either
(f) $\mathrm{Var}(\varepsilon)Z = ZB$ for some matrix $B$, or
(g) $\mathcal{R}(Z)$ is generated by $r$ eigenvectors of $\mathrm{Var}(\varepsilon)$, where $r$ is the rank of $Z$.

66. Prove Corollary 3.3.

67. Suppose that
$$X = \mu J_n + H\xi + e,$$

where $\mu \in \mathcal{R}$ is an unknown parameter, $H$ is an $n \times p$ known matrix of full rank, $\xi$ is a random $p$-vector with $E(\xi) = 0$ and $\text{Var}(\xi) = \sigma_\xi^2 I_p$, $e$ is a random $n$-vector with $E(e) = 0$ and $\text{Var}(e) = \sigma^2 I_n$, and $\xi$ and $e$ are independent. Show that the LSE of $\mu$ is the BLUE if and only if the row totals of $HH^\tau$ are the same.

68. Consider a special case of model (3.25):

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad i = 1, ..., a, j = 1, ..., b,$$

where $\mu$, $\alpha_i$'s and $\beta_j$'s are unknown parameters, $E(\varepsilon_{ij}) = 0$, $\text{Var}(\varepsilon_{ij}) = \sigma^2$, $\text{Cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = 0$ if $i \neq i'$, and $\text{Cov}(\varepsilon_{ij}, \varepsilon_{ij'}) = \sigma^2 \rho$ if $j \neq j'$. Show that the LSE of $\beta l^\tau$ is the BLUE for any $l \in \mathcal{R}(Z)$.

69. Consider model (3.25) under assumption A3 with $\text{Var}(\varepsilon) =$ a block diagonal matrix whose $i$th block diagonal $V_i$ is $n_i \times n_i$ and has a single eigenvalue $\lambda_i$ with eigenvector $J_{n_i}$ and a repeated eigenvalue $\rho_i$ with multiplicity $n_i - 1$, $i = 1, ..., k$, $\sum_{i=1}^k n_i = n$. Let $U = (U_1^\tau, ..., U_k^\tau)$, where $U_1 = (J_{n_1}, 0, ..., 0)$, $U_2 = (0, J_{n_2}, ..., 0)$,..., $U_k = (0, 0, ..., J_{n_k})$.
(a) If $\mathcal{R}(Z^\tau) \subset \mathcal{R}(U^\tau)$ and $\lambda_i \equiv \lambda$, show that $\hat{\beta} l^\tau$ is the BLUE for any $l \in \mathcal{R}(Z)$.
(b) If $Z^\tau U_i = 0$ for all $i$ and $\rho_i \equiv \rho$, show that $\hat{\beta} l^\tau$ is the BLUE for any $l \in \mathcal{R}(Z)$.

70. Prove Proposition 3.4.

71. Show that the condition $\sup_n \lambda_+[\text{Var}(\varepsilon)] < \infty$ is equivalent to the condition $\sup_i \text{Var}(\varepsilon_i) < \infty$.

72. Find a condition under which the mse of $\hat{\beta} l^\tau$ is of the order $n^{-1}$. Apply it to problems in Exercises 50-53.

73. Consider model (3.25) with i.i.d. $\varepsilon_1, ..., \varepsilon_n$ having $E(\varepsilon_i) = 0$ and $\text{Var}(\varepsilon_i) = \sigma^2$. Let $\hat{X}_i = \hat{\beta} Z_i^\tau$ and $h_{ii} = Z_i (Z^\tau Z)^- Z_i^\tau$.
(a) Show that for any $\epsilon > 0$,

$$P(|\hat{X}_i - E\hat{X}_i| \geq \epsilon) \geq \min[P(\varepsilon_i \geq \epsilon/h_i), P(\varepsilon_i \leq -\epsilon/h_i)].$$

(Hint: for independent random variables $X$ and $Y$, $P(|X + Y| \geq \epsilon) \geq P(X \geq \epsilon)P(Y \geq 0) + P(X \leq -\epsilon)P(Y < 0)$.)
(b) Show that $\hat{X}_i - E\hat{X}_i \to_p 0$ if and only if $h_{ii} \to 0$.

74. Prove Lemma 3.3 and show that condition (a) is implied by $\{\|Z_i\|\}$ is bounded and $\lambda_+(Z^\tau Z)^- \to 0$.

75. Consider the problem in Exercise 52. Suppose that $\{t_{ij}\}$ is bounded. Find a condition under which (3.40) holds.

76. Consider the one-way random effects model in Example 3.17. Assume that $\{n_i\}$ is bounded and $E|e_{ij}|^{2+\delta} < \infty$ for some $\delta > 0$. Show that the LSE $\hat{\mu}$ of $\mu$ is asymptotically normal and derive an explicit form of $\mathrm{Var}(\hat{\mu})$.

77. Suppose that
$$X_i = \rho t_i + \varepsilon_i, \quad i = 1, ..., n,$$
where $\rho \in \mathcal{R}$ is an unknown parameter, $t_i \in (a, b)$, $i = 1, ..., n$, $a$ and $b$ are known positive constants, and $\varepsilon_i$'s are independent random variables satisfying $E(\varepsilon_i) = 0$, $E|\varepsilon_i|^{2+\delta} < \infty$ for some $\delta > 0$ and $\mathrm{Var}(\varepsilon_i) = \sigma^2 t_i$ with an unknown $\sigma^2 > 0$.
(a) Obtain the LSE of $\rho$.
(b) Obtain the BLUE of $\rho$.
(c) Show that both the LSE and BLUE are asymptotically normal and obtain the asymptotic relative efficiency of the BLUE w.r.t. the LSE.

78. In Example 3.19, show that $E(S^2) = \sigma^2$ given by (3.44).

79. Suppose that $X = (X_1, ..., X_n)$ is a simple random sample (without replacement) from a finite population $\mathcal{P} = \{y_1, ..., y_N\}$ with univariate $y_i$.
(a) Show that a necessary condition for $h(\theta)$ to be estimable is that $h$ is symmetric in its $N$ arguments.
(b) Find the UMVUE of $Y^m$, where $m$ is a fixed positive integer $< n$ and $Y$ is the population total.
(c) Find the UMVUE of $P(X_i \leq X_j)$, $i \neq j$.
(d) Find the UMVUE of $\mathrm{Cov}(X_i, X_j)$, $i \neq j$.

80. Prove Theorem 3.14.

81. Under stratified simple random sampling described in §3.4.1, show that the vector of ordered values of all $X_{hi}$'s is neither sufficient nor complete for $\theta \in \Theta$.

82. Let $\mathcal{P} = \{y_1, ..., y_N\}$ be a population with univariate $y_i$. Define the population c.d.f. by
$$F(t) = \frac{1}{N} \sum_{i=1}^{N} I_{(-\infty,t)}(y_i).$$
Find the UMVUE of $F(t)$ under (a) simple random sampling and (b) stratified simple random sampling.

83. Consider the estimation of $F(t)$ in the previous exercise. Suppose that a sample of size $n$ is selected with $\pi_i > 0$. Find the Horvitz-Thompson estimator of $F(t)$. Is it a c.d.f.?

84. Show that $v_1$ in (3.50) and $v_2$ in (3.51) are unbiased estimators of $\text{Var}(\hat{Y}_{ht})$. Prove that $v_1 = v_2$ under (a) simple random sampling and (b) stratified simple random sampling.

85. Consider the following two-stage stratified sampling plan. In the first stage, the population is stratified into $H$ strata and $k_h$ clusters are selected from stratum $h$ with probability proportional to cluster size, where sampling is independent across strata. In the second stage, a sample of $m_{hi}$ units are selected from sampled cluster $i$ in stratum $h$, and sampling is independent across clusters. Find $\pi_i$ and the Horvitz-Thompson estimator $\hat{Y}_{ht}$ of the population total.

86. In the previous exercise, prove the unbiasedness of $\hat{Y}_{ht}$ directly (without using Theorem 3.15).

87. Under systematic sampling, show that $\text{Var}(\hat{Y}_{sy})$ is equal to

$$\left(1 - \frac{1}{N}\right)\frac{\sigma^2}{n} + \frac{2}{nN}\sum_{i=1}^{k}\sum_{1\le t<u\le n}\left(y_{i+(t-1)k} - \frac{Y}{N}\right)\left(y_{i+(u-1)k} - \frac{Y}{N}\right).$$

88. Prove (3.52)-(3.54) in Example 3.21. Show that (3.52) and (3.54) still hold if $(X_1, Y_1), ..., (X_n, Y_n)$ is a simple random sample from a finite bivariate population of size $N$, as $n \to N$.

89. Derive the $n^{-1}$ order asymptotic bias of the sample correlation coefficient defined in Exercise 19 in §2.6.

90. Derive the $n^{-1}$ order asymptotic bias and amse of $\hat{t}_\beta$ in Example 3.22, assuming that $\sum_{j=0}^{p-1}\beta_j t^j$ is convex in $t$.

91. Consider Example 3.23.
(a) Show that $\hat{\theta}$ is the BLUE of $\theta$.
(b) Show that $\hat{\sigma}^2$ is unbiased for $\sigma^2$.
(c) Show that $\hat{\Sigma}$ is consistent for $\Sigma$ as $k \to \infty$.
(d) Derive an amse of $\hat{R}(t)$.

92. Let $X_1, ..., X_n$ be i.i.d. from $N(\mu, \sigma^2)$, where $\mu \in \mathcal{R}$ and $\sigma^2 > 0$. Consider the estimation of $\vartheta = E\Phi(a + bX_1)$, where $\Phi$ is the standard normal c.d.f. and $a$ and $b$ are known constants. Obtain an explicit form of a function $g(\mu, \sigma^2) = \vartheta$ and an amse of $\hat{\vartheta} = g(\bar{X}, S^2)$.

93. Let $X_1, ..., X_n$ be i.i.d. with mean $\mu$, variance $\sigma^2$, and finite $\mu_j = EX_1^j$, $j = 2, 3, 4$. The sample coefficient of variation is defined to be $S/\bar{X}$, where $S$ is the squared root of the sample variance $S^2$.
(a) If $\mu \neq 0$, show that $\sqrt{n}(S/\bar{X} - \sigma/\mu) \to_d N(0, \tau)$ and obtain an explicit formula of $\tau$ in terms of $\mu$, $\sigma^2$, and $\mu_j$.
(b) If $\mu = 0$, show that $n^{-1/2}S/\bar{X} \to_d [N(0, 1)]^{-1}$.

94. Prove (3.56).

95. In Exercise 83, discuss how to obtain a consistent (as $n \to N$) estimator $\hat{F}(t)$ of $F(t)$ such that $\hat{F}$ is a c.d.f.

96. Let $X_1, ..., X_n$ be i.i.d. from $P$ in a parametric family. Obtain moment estimators of parameters in the following cases.
(a) $P$ is the gamma distribution $\Gamma(\alpha, \gamma)$, $\alpha > 0$, $\gamma > 0$.
(b) $P$ is the exponential distribution $E(a, \theta)$, $a \in \mathcal{R}$, $\theta > 0$.
(c) $P$ is the beta distribution $B(\alpha, \beta)$, $\alpha > 0$, $\beta > 0$.
(d) $P$ is the log-normal distribution $LN(\mu, \sigma^2)$, $\mu \in \mathcal{R}$, $\sigma > 0$.
(e) $P$ is the uniform distribution $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, $\theta \in \mathcal{R}$.
(f) $P$ is the negative binomial distribution $NB(p, r)$, $p \in (0, 1)$, $r = 1, 2, ....$
(g) $P$ is the log-distribution $L(p)$, $p \in (0, 1)$.
(h) $P$ is the chi-square distribution $\chi_k^2$ with an unknown $k = 1, 2, ....$

97. In part (b) of the previous exercise, obtain the asymptotic relative efficiencies of moment estimators w.r.t. UMVUE's.

98. Prove (3.57) and (3.58).

99. In the proof of Proposition 3.5, show that $E[W_n(U_n - \vartheta)] = O(n^{-1})$.

100. Prove (3.61).

101. Let $X_1, ..., X_n$ be i.i.d. with a c.d.f. $F$ and $U_n$ and $V_n$ be the U- and V-statistics with kernel $\int [I_{(-\infty,y)}(x_1) - F_0(y)][I_{(-\infty,y)}(x_2) - F_0(y)]dF_0$, where $F_0$ is a known c.d.f.
(a) Obtain the asymptotic distributions of $U_n$ and $V_n$ when $F \neq F_0$.
(b) Obtain the asymptotic relative efficiency of $U_n$ w.r.t. $V_n$ when $F = F_0$.

102. Let $X_1, ..., X_n$ be i.i.d. with a c.d.f. $F$ having a finite 6th moment. Consider the estimation of $\mu^3$, where $\mu = EX_1$. When $\mu = 0$, find $\text{amse}_{\bar{X}^3}(P)/\text{amse}_{U_n}(P)$, where $U_n = \binom{n}{3}^{-1} \sum_{1 \leq i < j < k \leq n} X_i X_j X_k$.

103. Prove (3.67).

104. Prove that $\hat{\sigma}^2$ in (3.69) is unbiased and consistent for $\sigma^2$ under model (3.25) with (3.68) and $\sup_i E|\varepsilon_i|^{2+\delta} < \infty$ for some $\delta > 0$. Under the same conditions, show that $\hat{\Sigma}$ in (3.70) is consistent for $\Sigma$.

105. Show how to use equation (3.71) to obtain consistent estimators of $\theta_0$ and $\theta_1$.