34. Statistical functionals and V-statistics

Lehmann §6.1, 6.2

Let S be a set of cumulative distribution functions and let $T\colon S \to R$ be a functional defined on that set. Then T is called a statistical functional. In general, we may have an iid sample from an unknown distribution F, and we want to learn the value of $\theta = T(F)$ for a (known) functional T. Some particular instances of statistical functionals are as follows:

- If $T(F) = F(c)$ for some constant c, then T is a statistical functional mapping each F to $P_F(X \le c)$.
- If $T(F) = F^{-1}(p)$ for some constant p, then T maps F to its pth quantile. Note that it is possible to define $F^{-1}(p)$ in such a way so that it is always defined even for distribution functions F that aren't invertible in the usual sense: For example, take $F^{-1}(p) \stackrel{\text{def}}{=} \inf\{x : F(x) \ge p\}$.
- If $T(F) = \operatorname{E}_F(X)$, then of course T maps a distribution function to its mean.

Suppose $X_1, \ldots, X_n$ is an iid sequence with distribution function $F(x)$. We define the empirical distribution function $\hat{F}_n$ to be the distribution function for a discrete uniform distribution on $\{X_1, \ldots, X_n\}$. In other words,

$$\hat{F}_n(x) \;=\; \frac{1}{n}\#\{i : X_i \le x\} \;=\; \frac{1}{n}\sum_{i=1}^n I\{X_i \le x\}.$$

Note that $\hat{F}_n(x)$ is a legitimate cumulative distribution function. Thus, if T is a statistical functional, we might consider $T(\hat{F}_n)$ as an estimator of $T(F)$. For obvious reasons, this estimator is called a *plug-in* estimator; we will discuss it in more detail when we consider V-estimators a bit later. For example, if $T(F) = \operatorname{E}_F(X)$, then the plug-in estimator given an iid sample $X_1, X_2, \ldots$ from F is

$$T(\hat{F}_n) \;=\; \operatorname{E}_{\hat{F}_n}(X) \;=\; \frac{1}{n}\sum_{i=1}^n X_i \;=\; \overline{X}_n.$$

Suppose that for some real-valued function $\phi(x)$, we define $T(F) = \operatorname{E}_F \phi(X)$. Note in this case that

$$T\{\alpha F_1 + (1-\alpha)F_2\} \;=\; \alpha \operatorname{E}_{F_1} \phi(X) + (1-\alpha) \operatorname{E}_{F_2} \phi(X) \;=\; \alpha T(F_1) + (1-\alpha) T(F_2).$$

For this reason, such a functional is sometimes called a linear functional.

To generalize this idea, we may take a real-valued function taking more than one real argument, say $\phi(x_1, \ldots, x_a)$ for some $a > 1$, and define

$$T(F) = \operatorname{E}_F \phi(X_1, \ldots, X_a), \tag{122}$$

which we take to mean the expectation of $\phi(X_1, \ldots, X_a)$ where $X_1, \ldots, X_a$ is an iid sample from the distribution function F. Since $X_1, \ldots, X_a$ are iid, it is clear that

$$\operatorname{E}_F \phi(X_1, \ldots, X_a) = \operatorname{E}_F \phi(X_{\pi(1)}, \ldots, X_{\pi(a)})$$

for any permutation π mapping $\{1, \ldots, a\}$ onto itself. Since there are $a!$ such permutations, consider the function

$$\phi^*(x_1, \ldots, x_a) = \frac{1}{a!} \sum_{\text{all } \pi} \phi(x_{\pi(1)}, \ldots, x_{\pi(a)}).$$

Since $\mathrm{E}_F\, \phi(X_1, \ldots, X_a) = \mathrm{E}_F\, \phi^*(X_1, \ldots, X_a)$ and ϕ^* is symmetric in its arguments (i.e., permuting its a arguments does not change its value), we see that in Equation (122) we may assume without loss of generality that ϕ is symmetric in its arguments. When $a > 1$, it may be shown that T is not a linear functional (this is beyond the scope of this course); nonetheless, we still consider it an expectation functional.

Definition 34.1 For some integer $a \geq 1$, suppose $X_1, \ldots, X_a$ are a random sample from the distribution F. Then given a function $\phi: R^a \to R$ that is symmetric in its a arguments, the map $T: F \mapsto \mathrm{E}\, \phi(X_1, \ldots, X_a)$ is called an expectation functional. We write $T(F) = \mathrm{E}_F\, \phi(X_1, \ldots, X_a)$. If $a = 1$, then T is also called a linear functional. The function ϕ is called the kernel function.

Suppose $T(F)$ is an expectation functional defined according to Equation (122) for some $a \geq 1$. If we have an iid sample of size n from F, then as noted earlier, a natural way to estimate $T(F)$ is by the use of the plug-in estimator $T(\hat{F}_n)$. This estimator is called a V-estimator or a V-statistic. It is possible to write down a V-statistic explicitly: Since $\hat{F}_n$ assigns probability $\frac{1}{n}$ to each X_i, we have

$$V_n \;=\; T(\hat{F}_n) \;=\; \mathrm{E}_{\hat{F}_n}\, \phi(X_1, \ldots, X_a) \;=\; \frac{1}{n^a} \sum_{i_1=1}^{n} \cdots \sum_{i_a=1}^{n} \phi(X_{i_1}, \ldots, X_{i_a}). \tag{123}$$

In the case $a = 1$, then Equation (123) becomes

$$V_n = \frac{1}{n} \sum_{i=1}^{n} \phi(X_i).$$

It is clear in this case that $\mathrm{E}\, V_n = T(F)$, which we denote by θ. Furthermore, if $\sigma^2 = \mathrm{Var}_F\, \phi(X) < \infty$, then the central limit theorem implies that

$$\sqrt{n}(V_n - \theta) \xrightarrow{\mathcal{L}} N(0, \sigma^2).$$

For $a > 1$, however, the situation becomes slightly more complicated. First of all, V_n is no longer unbiased, since the sum in equation (123) contains some terms in which $i_1, \ldots, i_a$ are not all distinct. The expectation of such terms is not necessarily equal to $\theta = T(F)$.

Example 34.1 Let $a = 2$ and $\phi(x_1, x_2) = |x_1 - x_2|$. It may be shown (Problem 34.2) that the functional $T(F) = \mathrm{E}_F\, |X_1 - X_2|$ is not linear in F. Furthermore, since $|X_{i_1} - X_{i_2}|$ is identically zero whenever $i_1 = i_2$, it may also be shown that the V-estimator of $T(F)$ is biased.

Since the bias in V_n is due to the duplication among the subscripts $i_1, \ldots, i_a$, the obvious way to correct this bias is to restrict the summation in Equation (123) to sets of subscripts $i_1, \ldots, i_a$ that contain no duplication. For example, we might sum instead over all $i_1 < \cdots < i_a$, an idea that leads in the next topic to U-statistics.

Problems

Problem 34.1 Let $X_1, \ldots, X_n$ be an iid sample from F. For a fixed x for which $0 < F(x) < 1$, find the asymptotic distribution of $\hat{F}_n(x)$.

Problem 34.2 Let $T(F) = \mathrm{E}_F\, |X_1 - X_2|$.

(a) Show that $T(F)$ is not a linear functional by exhibiting distributions F_1 and F_2 and a constant $\alpha \in (0, 1)$ such that

$$T\{\alpha F_1 + (1 - \alpha) F_2\} \neq \alpha T(F_1) + (1 - \alpha) T(F_2).$$

(b) For $n > 1$, demonstrate that the V-statistic V_n is biased in this case by finding $c_n \neq 1$ such that $\operatorname{E}_F V_n = c_n T(F)$.

Problem 34.3 Let $X_1, \ldots, X_n$ be a random sample from a distribution F with finite third absolute moment.

(a) Find $\phi(x_1, \ldots, x_a)$ for some a such that $\operatorname{E}_F \phi(X_1, \ldots, X_a) = \operatorname{Var}_F X$. Note that a is not allowed to depend on n, and you should try to find the smallest possible a. Furthermore, ϕ should be symmetric in its arguments.

(b) Let $\mu = \operatorname{E}_F(X)$. As in part (a), find $\phi(x_1, \ldots, x_a)$ such that $\operatorname{E}_F \phi(X_1, \ldots, X_a) = \operatorname{E}_F(X - \mu)^3$.

35. U-statistics

Lehmann §6.1

Because the V-statistic

$$V_n = \frac{1}{n^a} \sum_{i_1=1}^{n} \cdots \sum_{i_a=1}^{n} \phi(X_{i_1}, \ldots, X_{i_a})$$

is in general a biased estimator of $T(F)$ due to presence of summands in which there are duplicated indices on the X_{i_k}, the obvious way to produce an unbiased estimator is to sum only over those $(i_1, \ldots, i_a)$ in which no duplicates occur. Because ϕ is assumed to be symmetric in its arguments, we may without loss of generality restrict attention to the cases in which $1 \le i_1 < \cdots < i_a \le n$. Doing this, we obtain the U-statistic

$$U_n = \frac{1}{\binom{n}{a}} \sum_{1 \le i_1 < \cdots < i_a \le n} \phi(X_{i_1}, \ldots, X_{i_a}). \tag{124}$$

Note that the sample size n must be at least as large as a for a U-statistic to be defined. Naturally, the "U" in "U-statistic" stands for unbiased (the "V" in "V-statistic" stands for von Mises, who was one of the originators of this theory in the late 1940's). The fact that U_n is unbiased is obvious, since it is the average of $\binom{n}{a}$ terms, each with expectation $T(F) = \mathrm{E}_F\, \phi(X_1, \ldots, X_a)$. Because the subscripted F will always be understood, we omit it in what follows.

Example 35.1 Consider a random sample $X_1, \ldots, X_n$ from F, and let

$$R_n = \sum_{j=1}^{n} j I\{W_j > 0\}$$

be the Wilcoxon signed rank statistic, where $W_1, \ldots, W_n$ are simply $X_1, \ldots, X_n$ reordered in increasing absolute value. We showed in Example 24.1 that

$$R_n = \sum_{i=1}^{n} \sum_{j=1}^{i} I\{X_i + X_j > 0\}.$$

Letting $\phi(a, b) = I\{a + b > 0\}$, we see that ϕ is symmetric in its arguments and thus

$$\frac{1}{\binom{n}{2}} R_n = U_n + \frac{1}{\binom{n}{2}} \sum_{i=1}^{n} I\{X_i > 0\} = U_n + O_P\left(\tfrac{1}{n}\right),$$

where U_n is the U-statistic corresponding to the expectation functional $T(F) = P(X_1 + X_2 > 0)$. Therefore, many of the properties of the signed rank test that we have already derived elsewhere can also be obtained using the theory of U-statistics developed later in this topic.

In the special case $a = 1$, the V-statistic and the U-statistic coincide. In this case, we have already seen that both U_n and V_n are asymptotically normal by the central limit theorem. However, for $a > 1$ clearly the two statistics do not coincide in general. Furthermore, we may no longer use the central limit theorem to obtain asymptotic normality because the summands are not independent (each X_i appears in more than one summand).

To prove the asymptotic normality of U-statistics, we shall use a method sometimes known as the H-projection method after its inventor, Hoeffding. If $\phi(x_1, \ldots, x_a)$ is the kernel function of an expectation functional $T(F) = \mathrm{E}\, \phi(X_1, \ldots, X_a)$, suppose $X_1, \ldots, X_n$ is a random sample from the distribution F. Let $\theta = T(F)$ and let U_n be the U-statistic as defined in Equation (124). For $1 \le i \le a$, suppose that the values of $X_1, \ldots, X_i$ are held constant, say $X_1 = x_1, \ldots, X_i = x_i$, and the expectation of $\phi(X_1, \ldots, X_a)$ is taken. The result will of course be a function of $x_1, \ldots, x_i$, and we may view this procedure as projecting the random vector $(X_1, \ldots, X_a)$ onto the $(a-i)$-dimensional subspace in R^a given by $\{(x_1, \ldots, x_i, c_{i+1}, \ldots, c_a) : (c_{i+1}, \ldots, c_a) \in R^{a-i}\}$. To this end, we define for $i = 1, \ldots, a$

$$\phi_i(x_1, \ldots, x_i) \ = \ \mathrm{E}\ \phi(x_1, \ldots, x_i, X_{i+1}, \ldots, X_a). \tag{125}$$

Alternatively, we may define

$$\phi_i(X_1, \ldots, X_i) = \mathrm{E}\ \{\phi(X_1, \ldots, X_a) \mid X_1, \ldots, X_i\}. \tag{126}$$

From Equation (126), it is clear that $\mathrm{E}\ \phi_i(X_1, \ldots, X_i) = \theta$ for all i. Furthermore, we define

$$\sigma_i^2 = \mathrm{Var}\ \phi_i(X_1, \ldots, X_i). \tag{127}$$

The variance of U_n can be expressed in terms of the σ_i^2 as follows:

Theorem 35.1

$$\mathrm{Var}\ U_n = \frac{1}{\binom{n}{a}} \sum_{k=1}^{a} \binom{a}{k} \binom{n-a}{a-k} \sigma_k^2.$$

Proof: We start by noting that

$$\begin{aligned} \mathrm{Var} \binom{n}{a} U_n &= \sum_{1 \le j_1 < \cdots < j_a \le n} \ \sum_{1 \le i_1 < \cdots < i_a \le n} \mathrm{Cov}\left\{\phi(X_{j_1}, \ldots, X_{j_a}), \phi(X_{i_1}, \ldots, X_{i_a})\right\} \\ &= \binom{n}{a} \sum_{1 \le i_1 < \cdots < i_a \le n} \mathrm{Cov}\left\{\phi(X_1, \ldots, X_a), \phi(X_{i_1}, \ldots, X_{i_a})\right\} \end{aligned}$$

because the inner sum is the same for each of the $\binom{n}{a}$ possible choices of $1 \le j_1 < \cdots < j_a \le n$. Next, note that there are $\binom{a}{k}\binom{n-a}{a-k}$ ways to choose the values $1 \le i_1 < \cdots < i_a \le n$ so that exactly k of them are less than or equal to a. Therefore,

$$\mathrm{Var} \binom{n}{a} U_n = \binom{n}{a} \sum_{k=1}^{a} \binom{a}{k} \binom{n-a}{a-k} \mathrm{Cov}\left\{\phi(X_1, \ldots, X_a), \phi(X_1, \ldots, X_k, X_{a+1}, \ldots, X_{a+(a-k)})\right\}$$

(note that the sum does not include $k = 0$ because the covariance between $\phi(X_1, \ldots, X_a)$ and $\phi(X_{i_1}, \ldots, X_{i_a})$ is zero whenever none of the $i_1, \ldots, i_a$ is less than or equal to a). Finally, we evaluate

$$\begin{aligned} & \mathrm{E}\ \phi(X_1, \ldots, X_a)\phi(X_1, \ldots, X_k, X_{a+1}, \ldots, X_{a+(a-k)}) \\ = \ & \mathrm{E}\left(\mathrm{E}\left\{\phi(X_1, \ldots, X_a)\phi(X_1, \ldots, X_k, X_{a+1}, \ldots, X_{a+(a-k)}) \mid X_1, \ldots, X_k\right\}\right) \\ = \ & \mathrm{E}\left(\mathrm{E}\left\{\phi(X_1, \ldots, X_a) \mid X_1, \ldots, X_k\right\} \mathrm{E}\left\{\phi(X_1, \ldots, X_k, X_{a+1}, \ldots, X_{a+(a-k)}) \mid X_1, \ldots, X_k\right\}\right) \\ = \ & \mathrm{E}\ \{\phi_k(X_1, \ldots, X_k)\}^2 \\ = \ & \sigma_k^2 + \theta^2. \end{aligned}$$

Therefore,

$$\mathrm{Cov}\left\{\phi(X_1, \ldots, X_a), \phi(X_1, \ldots, X_k, X_{a+1}, \ldots, X_{a+(a-k)})\right\} = \sigma_k^2 \tag{128}$$

and the proof is complete. ■

Now we derive the asymptotic normality of U_n in a sequence of steps. The basic idea will be to show that $U_n - \theta$ has the same limiting distribution as the sum

$$\tilde{U}_n = \sum_{j=1}^{n} \mathrm{E}\,(U_n - \theta \mid X_j) \tag{129}$$

of projections. The asymptotic distribution of $\tilde{U}_n$ is easy to derive because it may be shown to be the sum of iid random variables.

Lemma 35.1 For all $1 \le j \le n$,

$$\mathrm{E}\,(U_n - \theta \mid X_j) = \frac{a}{n}\left\{\phi_1(X_j) - \theta\right\}.$$

Proof: Note that

$$\mathrm{E}\,(U_n - \theta \mid X_j) = \frac{1}{\binom{n}{a}} \sum_{1 \le i_1 < \cdots < i_a \le n} \mathrm{E}\,\{\phi(X_{i_1}, \ldots, X_{i_a}) - \theta \mid X_j\}.$$

Clearly, the term $\mathrm{E}\,\{\phi(X_{i_1}, \ldots, X_{i_a}) - \theta \mid X_j\}$ equals $\phi_1(X_j) - \theta$ whenever j is among $\{i_1, \ldots, i_a\}$ and 0 otherwise. The number of ways to choose $\{i_1, \ldots, i_a\}$ so that j is among them is $\binom{n-1}{a-1}$, so we obtain

$$\mathrm{E}\,(U_n - \theta \mid X_j) \;=\; \frac{\binom{n-1}{a-1}}{\binom{n}{a}}\{\phi_1(X_j) - \theta\} \;=\; \frac{a}{n}\{\phi_1(X_j) - \theta\}.$$

■

Lemma 35.2 If $\sigma_1^2 < \infty$ and $\tilde{U}_n$ is defined as in Equation (129), then

$$\sqrt{n}\tilde{U}_n \xrightarrow{\mathcal{L}} N(0, a^2\sigma_1^2).$$

Proof: Lemma 35.2 follows immediately from Lemma 35.1 and the central limit theorem since $a\phi_1(X_j)$ has mean $a\theta$ and variance $a^2\sigma_1^2$. ■

Now that we know the asymptotic distribution of $\tilde{U}_n$, it remains to show that $U_n - \theta$ and $\tilde{U}_n$ have the same limiting behavior.

Lemma 35.3

$$\mathrm{E}\left\{\tilde{U}_n(U_n - \theta)\right\} = \mathrm{E}\,\tilde{U}_n^2.$$

Proof: By Equation (129) and Lemma 35.1,

$$\begin{aligned}
\mathrm{E}\left\{\tilde{U}_n(U_n - \theta)\right\} &= \frac{a}{n}\sum_{j=1}^{n} \mathrm{E}\,\{(\phi_1(X_j) - \theta)(U_n - \theta)\} \\
&= \frac{a}{n}\sum_{j=1}^{n} \mathrm{E}\,\mathrm{E}\,\{(\phi_1(X_j) - \theta)(U_n - \theta) \mid X_j\} \\
&= \frac{a^2}{n^2}\sum_{j=1}^{n} \mathrm{E}\,\{\phi_1(X_j) - \theta\}^2 \\
&= \frac{a^2\sigma_1^2}{n}.
\end{aligned}$$

Since $\tilde{U}_n$ has mean zero, $\mathrm{E}\,\tilde{U}_n^2 = \mathrm{Var}\,\tilde{U}_n$ and this equals $a^2\sigma_1^2/n$ by Equation (129) and Lemma 35.1. ■

Lemma 35.4 If $\sigma_i^2 < \infty$ for $i = 1, \ldots, a$, then

$$\sqrt{n}\left(U_n - \theta - \tilde{U}_n\right) \stackrel{P}{\to} 0.$$

Proof: By Lemma 35.3, $n\,\mathrm{E}\left(U_n - \theta - \tilde{U}_n\right)^2 = n\left(\mathrm{Var}\ U_n - \mathrm{E}\ \tilde{U}_n^2\right)$. By Theorem 35.1,

$$n\,\mathrm{Var}\ U_n \;=\; an\sigma_1^2 \frac{\binom{n-a}{a-1}}{\binom{n}{a}} + \frac{n}{\binom{n}{a}} \sum_{k=2}^{a} \binom{a}{k}\binom{n-a}{a-k}\sigma_k^2.$$

Since

$$\frac{\binom{n-a}{a-k}}{\binom{n}{a}} \sim \frac{a!}{(a-k)!n^k},$$

we obtain $n\,\mathrm{Var}\ U_n \sim a^2\sigma_1^2 = n\,\mathrm{E}\ \tilde{U}_n^2$. This means that the relative error,

$$\frac{|n\,\mathrm{Var}\ U_n - n\,\mathrm{E}\ \tilde{U}_n^2|}{n\,\mathrm{E}\ \tilde{U}_n^2} = \frac{1}{a^2\sigma_1^2}|n\,\mathrm{Var}\ U_n - n\,\mathrm{E}\ \tilde{U}_n^2|,$$

converges to zero. We conclude that $\mathrm{E}\left\{\sqrt{n}(U_n - \theta - \tilde{U}_n)\right\}^2 \to 0$. Since convergence in quadratic mean implies convergence in probability, this proves the result. ■

Finally, since $\sqrt{n}(U_n - \theta) = \sqrt{n}\tilde{U}_n + \sqrt{n}(U_n - \theta - \tilde{U}_n)$, Lemmas 35.2 and 35.4 along with Slutsky's theorem result in the theorem we had set out to prove.

Theorem 35.2 If $\sigma_i^2 < \infty$ for $i = 1, \ldots, a$, then

$$\sqrt{n}(U_n - \theta) \stackrel{\mathcal{L}}{\to} N(0, a^2\sigma_1^2). \tag{130}$$

Problems

Problem 35.1 Suppose a kernel function $\phi(x_1, \ldots, x_a)$ satifies $\mathrm{E}\ |\phi(X_{i_1}, \ldots, X_{i_a})| < \infty$ for any (not necessarily distinct) $i_1, \ldots, i_a$. Prove that if U_n and V_n are the corresponding U- and V-statistics, then $\sqrt{n}(V_n - U_n) \stackrel{P}{\to} 0$ so that V_n has the same asymptotic distribution as U_n.

Problem 35.2 For the kernel function of Example 34.1, $\phi(a, b) = |a - b|$, the corresponding U-statistic is called Gini's mean difference and it is denoted G_n. For a random sample from uniform$(0, \tau)$, find the asymptotic distribution of G_n.

Problem 35.3 Let $\phi(x_1, x_2, x_3)$ have the property

$$\phi(a + bx_1, a + bx_2, a + bx_3) \;=\; \phi(x_1, x_2, x_3)\mathrm{sgn}(b) \quad \text{for all } a, b. \tag{131}$$

Let $\theta = \mathrm{E}\ \phi(X_1, X_2, X_3)$. The function $\mathrm{sgn}(x)$ is defined as $I\{x > 0\} - I\{x < 0\}$.

(a) We define the distribution F to be symmetric if for $X \sim F$, there exists some μ (the center of symmetry) such that $X - \mu$ and $\mu - X$ have the same distribution. Prove that if F is symmetric then $\theta = 0$.

(b) Let $\overline{x}$ and $\tilde{x}$ denote the mean and median of x_1, x_2, x_3. Let $\phi(x_1, x_2, x_3) = \mathrm{sgn}(\overline{x} - \tilde{x})$. Show that this function satisfies criterion (131), then find the asymptotic distribution for the corresponding U-statistic if F is the standard uniform distribution.

Problem 35.4 If the arguments of the kernel function $\phi(x_1, \ldots, x_a)$ of a U-statistic are vectors instead of scalars, note that Theorem 35.2 still applies with no modification. With this in mind, consider for $\underline{x}, \underline{y} \in R^2$ the kernel $\phi(\underline{x}, \underline{y}) = I\{(y_1 - x_1)(y_2 - x_2) > 0\}$.

(a) Given an iid sample $\underline{X}^{(1)}, \ldots, \underline{X}^{(n)}$, if U_n denotes the U-statistic corresponding to the kernel above, the statistic $2U_n - 1$ is called Kendall's tau statistic. Suppose the marginal distributions of $X_1^{(i)}$ and $X_2^{(i)}$ are both continuous, with $X_1^{(i)}$ and $X_2^{(i)}$ independent. Find the asymptotic distribution of $\sqrt{n}(U_n - \theta)$ for an appropriate value of θ.

(b) To test the null hypothesis that a sample $Z_1, \ldots, Z_n$ is iid against the alternative hypothesis that the Z_i are stochastically increasing in i, suppose we reject the null hypothesis if the number of pairs (Z_i, Z_j) with $Z_i < Z_j$ and $i < j$ is greater than c_n. This test is called Mann's test against trend. Based on your answer to part (a), find c_n so that the test has asymptotic level .05.

(c) Estimate the true level of the test in part (b) for an iid sample of size n from a standard normal distribution for each $n \in \{5, 15, 75\}$. Use 5000 samples in each case.