# U-Statistics

## 1.1 Definition

**Definition 1.1.1.** Suppose $(Y_i)_{i=1}^n$ is a random sample and $h : \mathbb{R}^m \to \mathbb{R}$ is a symmetric *kernel* function with:

$$\theta \equiv E\big\{h(Y_1, \cdots, Y_m)\big\}.$$

Here symmetric means $h(\cdot)$ is invariant with respect to permutations $(i_1, \cdots, i_m)$ of the indices $(1, \cdots, m)$. The **U-statistic** induced by $h$ is:

$$U_n(h) = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} h\big(Y_{i_1}, \cdots, Y_{i_m}\big). \tag{1.1.1}$$

■

**Remark 1.1.1.** Any unbiased estimator $u(Y_1, \cdots, Y_m)$ can be symmetrized via:

$$h(Y_1, \cdots, Y_m) \equiv \frac{1}{m!} \sum_P u(Y_{i_1}, \cdots, Y_{i_m}),$$

where $\sum_P$ denotes the sum across all permutations of the indices $(i_1, \cdots, i_m)$. ♦

**Proposition 1.1.1.** The definition of the $U$-statistic in (1.1.1) is equivalent to:

$$U_n = E\big\{h(Y_1, \cdots, Y_m)\big|(Y_{(1)}, \cdots, Y_{(n)})\big\}.$$

That is, $U_n(h)$ may be regarded as the expectation of $h(Y_1, \cdots, Y_m)$ conditional on the order statistics $(Y_{(1)}, \cdots, Y_{(n)})$. ♦

**Proof.** Given the order statistics, what remains random when evaluating the expectation of $h(Y_1, \cdots, Y_m)$ is the selection of the indices for the $m$ arguments from among the $n$ possible values. This can be done in $n$ choose $m$ possible ways. Let $(i_1, \cdots, i_m)$ denote one such sample of indices. Since $h$ is symmetric in its arguments, the sample of indices may always be arranged such that $1 \le i_1 < \cdots < i_m \le n$. Therefore:

$$E\big\{h(Y_1, \cdots, Y_m)\big|Y_{(1)}, \cdots, Y_{(n)}\big\} = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} h\big(Y_{i_1}, \cdots, Y_{i_m}\big).$$

■

## 1.2 Examples

**Example 1.1.1.**

- The sample mean is a $U$-statistic of order $m = 1$:

$$U_n = \frac{1}{n}\sum_{1\le i\le n} Y_i.$$

- The sample variance is a $U$-statistic of order $m = 2$:

$$U_n = \frac{2}{n(n-1)}\sum_{1\le i<j\le n}\frac{1}{2}(Y_i - Y_j)^2.$$

- The Gini mean absolute difference is a $U$-statistic of order $m = 2$:

$$U_n = \frac{2}{n(n-1)}\sum_{1\le i<j\le n}|Y_i - Y_j|.$$

♠

## 1.3 Optimality

**Theorem 1.1.1.** Suppose $S(Y_1, \cdots, Y_m)$ is unbiased for $\theta$. The $U$-statistic induced by $S$, defined by $U_n = E\big\{S(Y_1, \cdots, Y_m)|(Y_{(1)}, \cdots, Y_{(n)})\big\}$, is uniformly better than $S$ in that sense that $U_n$ is unbiased for $\theta$ and:

$$\text{Var}(U_n) \le \text{Var}(S).$$

Moreover, the inequality is strict unless $U_n = S$. □

**Proof.** By iterated expectation, $U_n$ is unbiased:

$$E(U_n) = EE\big\{S(Y_1, \cdots, Y_m)|(Y_{(1)}, \cdots, Y_{(n)})\big\} = E(S) = \theta.$$

By law of total variance:

$$\begin{aligned}\text{Var}(S) &= \text{Var}\Big[E\{S|(Y_{(1)}, \cdots, Y_{(n)})\}\Big] + E\Big[\text{Var}\{S|(Y_{(1)}, \cdots, Y_{(n)})\}\Big] \\ &= \text{Var}(U_n) + E\Big[\text{Var}\{S|(Y_{(1)}, \cdots, Y_{(n)})\}\Big].\end{aligned}$$

Since the latter term is non-negative, $\text{Var}(U_n) \le \text{Var}(S)$. ■

## 1.4 Properties

**Definition 1.1.2.** Consider $U_n(h)$ from (1.1.1). For $1 \leq k \leq m$, define the **projection**:

$$h_k(y_1, \cdots, y_k) = E\big\{h(Y_1, \cdots, Y_m)\big|Y_1 = y_1, \cdots, Y_k = y_k\big\}.$$

Let $\zeta_k$ denote the variance of the $k$th projection:

$$\zeta_k = \mathrm{Var}\big\{h_k(Y_1, \cdots, Y_k)\big\}.$$

■

**Remark 1.1.2.** Note that the $m$th projection is:

$$h_m = E\big\{h(Y_1, \cdots, Y_m)|Y_1 = y_1, \cdots, Y_m = y_m\big\} = h(y_1, \cdots, y_m),$$

and that $\zeta_m = \mathrm{Var}(h_m) = \mathrm{Var}\big\{h(Y_1, \cdots, Y_m)\big\}$. ♦

**Proposition 1.1.2.** Each projection $h_k$ has the same expectation as $U_n(h)$:

$$E\big\{h_k(Y_1, \cdots, Y_k)\big\} = \theta.$$

♦

**Proof.**

$$\begin{aligned} E\big\{h_k(Y_1, \cdots, Y_k)\big\} &= E\Big[E\big\{h(Y_1, \cdots, Y_m)\big|Y_1 = y_1, \cdots, Y_k = y_k\big\}\Big] \\ &= E\big\{h(Y_1, \cdots, Y_m)\big\} = \theta. \end{aligned}$$

■

**Proposition 1.1.3.** The variance of the projection $h_k$ is increasing in $k$:

$$0 \leq \zeta_1 \leq \cdots \leq \zeta_m = \mathrm{Var}(U).$$

♦

**Proof.**

$$\begin{aligned} \mathrm{Var}\big\{h_k(Y_1, \cdots, Y_k)\big\} &= \mathrm{Var}\Big[E\big\{h_k(Y_1, \cdots, Y_k)|(Y_1, \cdots, Y_{k-1})\big\}\Big] \\ &\quad + E\Big[\mathrm{Var}\big\{h_k(Y_1, \cdots, Y_k)|(Y_1, \cdots, Y_{k-1})\big\}\Big] \\ &\geq \mathrm{Var}\big\{h_{k-1}(Y_1, \cdots, Y_{k-1})\big\} = \zeta_{k-1}. \end{aligned}$$

■

**Proposition 1.1.4.** Suppose $(Y_i)_{i=1}^{n}$ are a random sample, then:

$$\begin{aligned}\zeta_k = \mathrm{Cov}\big\{&h(Y_1,\cdots,Y_k,Y_{k+1},\cdots,Y_m),\\ &h(Y_1,\cdots,Y_k,\tilde{Y}_{k+1},\cdots,\tilde{Y}_m)\big\}.\end{aligned}$$

Here $(Y_1,\cdots,Y_k)$ are in common, while $(Y_{k+1},\cdots,Y_m)$ are distinct from $(\tilde{Y}_{k+1},\cdots,\tilde{Y}_m)$. ♦

**Proof.** By definition:

$$\begin{aligned}\gamma &\equiv \mathrm{Cov}\big\{h(Y_1,\cdots,Y_k,Y_{k+1},\cdots,Y_m),h(Y_1,\cdots,Y_k,\tilde{Y}_{k+1},\cdots,\tilde{Y}_m)\big\}\\ &= E\big\{h(Y_1,\cdots,Y_k,Y_{k+1},\cdots,Y_m)h(Y_1,\cdots,Y_k,\tilde{Y}_{k+1},\cdots,\tilde{Y}_m)\big\}\\ &- E\big\{h(Y_1,\cdots,Y_k,Y_{k+1},\cdots,Y_m)\big\}E\big\{h(Y_1,\cdots,Y_k,\tilde{Y}_{k+1},\cdots,\tilde{Y}_m)\big\}.\end{aligned}$$

The expectation of each kernel function is $\theta$:

$$\gamma = E\big\{h(Y_1,\cdots,Y_k,Y_{k+1},\cdots,Y_m)h(Y_1,\cdots,Y_k,\tilde{Y}_{k+1},\cdots,\tilde{Y}_m)\big\} - \theta^2.$$

By iterated expectation:

$$\begin{aligned}&E\big\{h(Y_1,\cdots,Y_k,Y_{k+1},\cdots,Y_m)h(Y_1,\cdots,Y_k,\tilde{Y}_{k+1},\cdots,\tilde{Y}_m)\big\}\\ &\quad= E\big\{h(Y_1,\cdots,Y_k,Y_{k+1},\cdots,Y_m)h(Y_1,\cdots,Y_k,\tilde{Y}_{k+1},\cdots,\tilde{Y}_m)|(Y_1,\cdots,Y_k)\big\}\\ &\quad= E\big\{h_k(Y_1,\cdots,Y_k)h_k(Y_1,\cdots,Y_k)\big\}.\end{aligned}$$

Now:

$$\begin{aligned}\gamma &= E\big\{h_k(Y_1,\cdots,Y_m)h_k(Y_1,\cdots,Y_m)\big\} - \theta^2\\ &= E\big\{h_k^2(Y_1,\cdots,Y_m)\big\} - E^2\big\{h_k(Y_1,\cdots,Y_m)\big\} = \zeta_k.\end{aligned}$$

■

**Theorem 1.1.2 (Finite Sample Variance).** The finite sample variance of the $U$-statistic in (1.1.1) is:

$$\mathrm{Var}\big\{U_n(h)\big\} = \binom{n}{m}^{-1}\sum_{k=1}^{m}\binom{m}{k}\binom{n-m}{m-k}\zeta_k. \tag{1.1.2}$$

□

**Proof.**

$$\begin{aligned}\mathrm{Var}\big\{U_n(h)\big\} &= \mathrm{Var}\left\{\binom{n}{m}^{-1}\sum_{1\le i_1<\cdots<i_m\le n} h\big(Y_{i_1},\cdots,Y_{i_m}\big)\right\}\\ &= \binom{n}{m}^{-2}\mathrm{Cov}\left\{\sum_{1\le i_1<\cdots<i_m\le n} h\big(Y_{i_1},\cdots,Y_{i_m}\big),\sum_{1\le j_1<\cdots<j_m\le n} h\big(Y_{j_1},\cdots,Y_{j_m}\big)\right\}\\ &= \binom{n}{m}^{-2}\sum_{1\le i_1<\cdots<i_m\le n}\sum_{1\le j_1<\cdots<j_m\le n}\mathrm{Cov}\big\{h(Y_{i_1},\cdots,Y_{i_m}),h(Y_{j_1},\cdots,Y_{j_m})\big\}\end{aligned}$$

Consider cases. If $\{i_1, \cdots, i_m\}$ and $\{j_1, \cdots, j_m\}$ have no indices in common, then:

$$\text{Cov}\big\{h(Y_{i_1}, \cdots, Y_{i_m}), h(Y_{j_1}, \cdots, Y_{j_m})\big\} = 0.$$

Suppose $\{i_1, \cdots, i_m\}$ and $\{j_1, \cdots, j_m\}$ have $1 \leq k \leq m$ indices in common, then:

$$\text{Cov}\big\{h(Y_{i_1}, \cdots, Y_{i_m}), h(Y_{j_1}, \cdots, Y_{j_m})\big\} = \zeta_k$$

The number of covariance terms with $k$ indices in common is:

$$\binom{n}{m}\binom{m}{k}\binom{n-m}{m-k}.$$

$\binom{n}{m}$ is the number of ways to choose the first index set $\{i_1, \cdots, i_m\}$. Having fixed the first set, there are $\binom{m}{k}$ ways to select the $k$ indices $\{j_1, \cdots, j_k\}$ of the second set that are in common, times $\binom{n-m}{m-k}$ was to select the $m-k$ indices $\{j_{k+1}, \cdots, j_m\}$ of the second set that are different. Overall:

$$\text{Var}\big\{U_n(h)\big\} = \binom{n}{m}^{-2}\sum_{k=1}^{m}\binom{n}{m}\binom{m}{k}\binom{n-m}{m-k}\zeta_k.$$

■

**Proposition 1.1.5** (**Limiting Variance**). If $\zeta_1 > 0$ and $\zeta_m < \infty$, then:

$$\lim_{n\to\infty} \text{Var}\big\{\sqrt{n}(U_n - \theta)\big\} = m^2\zeta_1. \tag{1.1.3}$$

♦

**Proof.** From (1.1.2):

$$\text{Var}\big\{U_n(h)\big\} = \binom{n}{m}^{-1}\left\{\binom{m}{1}\binom{n-m}{m-1}\zeta_1 + \binom{m}{2}\binom{n-m}{m-2}\zeta_2 + \cdots\right\}.$$

For $n \to \infty$, the binomial coefficient is:

$$\binom{n-m}{a} = \frac{(n-m)!}{a!(n-m-a)!} = \frac{1}{a!}(n-m)(n-m-1)\cdots(n-m-a+1) \asymp \frac{n^a}{a!}$$

Thus, to leading order:

$$\text{Var}\big\{U_n(h)\big\} \asymp \frac{m!}{n^m}\left\{m \cdot \frac{n^{m-1}}{(m-1)!}\zeta_1 + \mathcal{O}(n^{m-2})\right\} = \frac{m^2\zeta_1}{n} + \mathcal{O}(n^{-2}).$$

Conclude that: $\text{Var}\big\{\sqrt{n}(U_n - \theta)\big\} = m^2\zeta_1 + \mathcal{O}(n^{-1})$. ■

**Theorem 1.1.3** (**Asymptotic Normality**). Consider the $U$-statistic in (1.1.1).

i. If $0 < \zeta_1 < \infty$, then as $n \to \infty$:

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

where $\sigma^2 = m^2\zeta_1$ is the limiting variance in (1.1.3).

ii. If $0 < \zeta_1 \leq \zeta_m < \infty$, then as $n \to \infty$:

$$\frac{U_n - \theta}{\sqrt{\text{Var}(U_n)}} \xrightarrow{\mathcal{L}} N\big(0, 1\big),$$

where $\text{Var}(U_n)$ is the finite sample variance in (1.1.2).

□

**Proof.** (ii.) Consider the Hajék projection representation of $U_n$ from (1.2.9):

$$\hat{S} - \theta = \frac{m}{n}\sum_{i=1}^{n}\big\{h_1(Y_i) - \theta\big\}$$

The variance of $(\hat{S} - \theta)$ is:

$$\text{Var}\big(\hat{S} - \theta\big) = \frac{m^2}{n}\text{Var}\big\{h_1(Y_i)\big\} = \frac{m^2}{n}\zeta_1.$$

By the central limit theorem:

$$\frac{\hat{S} - \theta}{\sqrt{m^2\zeta_1/n}} \xrightarrow{\mathcal{L}} N\big(0, 1\big).$$

By (1.2.8), if $\lim_{n\to\infty} \text{Var}(U_n - \theta)/\text{Var}(\hat{S} - \theta) = 1$, then:

$$\frac{\hat{S} - \theta}{\sqrt{\text{Var}(\hat{S} - \theta)}} = \frac{U_n - \theta}{\sqrt{\text{Var}(U_n - \theta)}} + o_p(1).$$

From (1.1.3), the variance of $(U_n - \theta)$ is:

$$\text{Var}\big(U_n - \theta\big) = \frac{m^2}{n}\zeta_1 + \mathcal{O}(n^{-2}).$$

The limiting variances are asymptotically equivalent, since:

$$\frac{\frac{m^2}{n}\zeta_1 + \mathcal{O}(n^{-2})}{\frac{m^2}{n}\zeta_1} = 1 + \mathcal{O}(n^{-1}).$$

Adding and subtracting the standardized projection:

$$\frac{U_n - \theta}{\sqrt{\mathrm{Var}(U_n - \theta)}} = \left( \frac{U_n - \theta}{\sqrt{\mathrm{Var}(U_n - \theta)}} - \frac{\hat{S} - \theta}{\sqrt{\mathrm{Var}(\hat{S} - \theta)}} \right) + \frac{\hat{S} - \theta}{\sqrt{\mathrm{Var}(\hat{S} - \theta)}}$$

By (1.2.8), the term in parentheses converges in probability to zero, and by the central limit theorem, the standardized projection converges in distribution to standard normal. From Slutsky's theorem, conclude that:

$$\frac{U_n - \theta}{\sqrt{\mathrm{Var}(U_n - \theta)}} \xrightarrow{\mathcal{L}} N\big(0, 1\big).$$

■

## 1.5 Wilcoxon One-Sample Statistic

**Example 1.1.2.** Suppose $(Y_i) \overset{\text{IID}}{\sim} F_Y$. Consider the parameter $\theta = P(Y_1 + Y_2 > 0)$. The corresponding $U$-statistic is:

$$U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} I(Y_i + Y_j > 0). \tag{1.1.4}$$

Identify $h(Y_1, Y_2) = I(Y_1 + Y_2 > 0)$ as the kernel. The first projection of $h$ is:

$$\begin{aligned} h_1(y_1) &= E\big\{h(Y_1, Y_2)|Y_1 = y_1\big\} = E\big\{I(y_1 + Y_2 > 0)\big\} \\ &= P(-y_1 < Y_2) = F_Y(-y_1). \end{aligned}$$

The Hajék representation of $U_n$ is:

$$\hat{S} - \theta = \frac{2}{n} \sum_{i=1}^{n} \big\{F_Y(-Y_i) - \theta\big\},$$

where $\theta = E\big\{F_Y(-Y)\big\}$. Suppose that, under $H_0$, $F_Y$ is *continuous* and *symmetric*, then by the probability integral theorem $F_Y(Y) \sim U(0, 1)$, and:

$$F_Y(-Y) = 1 - F_Y(Y) \overset{d}{=} 1 - U(0, 1) \overset{d}{=} U(0, 1).$$

Now, under $H_0$,

$$\begin{aligned} \theta &= E\big\{F_Y(-Y)\big\} = E\big\{U(0, 1)\big\} = \frac{1}{2}, \\ \zeta_1 &= \mathrm{Var}\big\{h_1(Y)\big\} = \mathrm{Var}\big\{U(0, 1)\big\} = \frac{1}{12}. \end{aligned}$$

Therefore, under $H_0$,

$$\sqrt{n}\left(U_n - \frac{1}{2}\right) \xrightarrow{\mathcal{L}} N\left(0, \frac{4}{12}\right) = N\left(0, \frac{1}{3}\right)$$

The Wilcoxon rank sum is:

$$W_n = \sum_{i=1}^{n} \text{rank}(|Y_i|)I(Y_i > 0), \qquad \text{rank}(|Y_i|) = \sum_{j=1}^{n} I(|Y_j| \leq |Y_i|).$$

Under the null hypothesis that $F_Y$ is *continuous*, such that there are no ties, the normalized Wilcoxon one-sample statistic $\binom{n}{2}^{-1}W_n$ is asymptotically equivalent to the $U$-statistic in (1.1.4). To see this, first rewrite $W_n$ using the indicator definition of rank:

$$W_n = \sum_{i=1}^{n}\sum_{j=1}^{n} I(|Y_j| \leq |Y_i|)I(Y_i > 0).$$

The summand evalutes to 1 if and only if $|Y_j| \leq Y_i$:

$$W_n = \sum_{i=1}^{n}\sum_{j=1}^{n} I(|Y_j| \leq Y_i).$$

Split the domain of summation:

$$W_n = \sum_{i<j} I(|X_j| < X_i) + \sum_{i<j} I(|X_i| < X_j) + \sum_{i=1}^{n} I(Y_i > 0).$$

The last term arises because $|Y_j| \leq Y_i$ only holds for positive $Y_i$. Let $P_n = \sum_{i=1}^{n} I(Y_i > 0)$ and combine the two leading sums:

$$\begin{aligned} W_n &= \sum_{i<j} \big\{I(|X_j| < X_i) + I(|X_i| < X_j)\big\} + P_n \\ &= \sum_{i<j} \big\{I(X_i - |X_j| > 0) + I(X_j - |X_i| > 0)\big\} + P_n. \end{aligned}$$

Observe that the term in braces $\{\}$ evaluates to 1 if and only if $X_i + X_j > 0$, therefore:

$$W_n = \sum_{1 \leq i < j \leq n} I(X_i + X_j > 0) + P_n.$$

Note that $P_n$ is bounded by $n$. Upon normalizing by $n$ choose 2:

$$\binom{n}{2}^{-1} W_n = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} I(X_i + X_j > 0) + \binom{n}{2}^{-1} P_n = U_n + \mathcal{O}_p(n^{-1}).$$

♠

## 1.6 Multi-sample Extension

**Definition 1.1.3.** Suppose $(X_i)_{i=1}^{n_1}$ and $(Y_i)_{i=1}^{n_2}$ are random samples, and $h : \mathbb{R}^m \to \mathbb{R}$ is a kernel function $(m = m_1 + m_2)$, symmetric within groups of arguments, and:

$$\theta \equiv E\big\{h(X_1, \cdots, X_{m_1}; Y_1, \cdots, Y_{m_2})\big\}.$$

The two-sample **U-statistic** $(n = n_1 + n_2)$ takes the form:

$$U_n(h) = \frac{1}{\binom{n_1}{m_1}\binom{n_2}{m_2}} \sum_{1 \le i_1 < \cdots < i_{m_1} \le n_1} \sum_{1 \le j_1 < \cdots < j_{m_2} \le n_2} h\big(X_{i_1}, \cdots, X_{i_{m_1}}; Y_{j_1}, \cdots, Y_{j_{m_2}}\big). \quad (1.1.5)$$

Extensions to addition samples are analogous. ■

**Definition 1.1.4.** Consider the multi-sample $U$-statistic in (1.1.5). Define $\zeta_{kl}$ as:

$$\begin{aligned}\zeta_{kl} = \mathrm{Cov}\big\{&h(X_1, \cdots, X_k, X_{k+1}, \cdots, X_{m_1}; Y_1, \cdots, Y_l, Y_{l+1}, \cdots, Y_{m_2}),\\ &h(X_1, \cdots, X_k, \tilde{X}_{k+1}, \cdots, \tilde{X}_{m_1}; Y_1, \cdots, Y_l, \tilde{Y}_{l+1}, \cdots, \tilde{Y}_{m_2})\big\}\end{aligned}$$

Here $(X_1, \cdots, X_k)$ and $(Y_1, \cdots, Y_l)$ are in common, while $(X_{k+1}, \cdots, X_{m_1})$ are distinct from $(\tilde{X}_{k+1}, \cdots, \tilde{X}_{m_1})$, and $(Y_{l+1}, \cdots, Y_{m_1})$ are distinct from $(\tilde{Y}_{l+1}, \cdots, \tilde{Y}_{m_2})$. ■

**Theorem 1.1.4 (Finite Sample Variance).** The finite sample variance of the $U$-statistic in (1.1.5) is:

$$\mathrm{Var}\big\{U_n(h)\big\} = \sum_{k=1}^{m_1} \sum_{l=1}^{m_2} \frac{\binom{m_1}{k}\binom{n_1-m_1}{m_1-k}}{\binom{n_1}{m_1}} \frac{\binom{m_2}{l}\binom{n_2-m_2}{m_2-l}}{\binom{n_2}{m_2}} \zeta_{kl}. \quad (1.1.6)$$

□

**Theorem 1.1.5 (Limiting Variance).** Consider the $U$-statistic in (1.1.5). Suppose $(n_1, n_2) \to \infty$ in such a way that:

$$\lim_{(n_1,n_2)\to\infty} \frac{n_1}{n_1 + n_2} = \rho, \qquad \lim_{(n_1,n_2)\to\infty} \frac{n_2}{n_1 + n_2} = 1 - \rho,$$

and let $n = n_1 + n_2$. If $0 < \zeta_{10}, \zeta_{01} \le \zeta_{m_1 m_2} < \infty$, then:

$$\sigma^2 = \lim_{n\to\infty} \mathrm{Var}\big\{\sqrt{n}(U_n - \theta)\big\} = \frac{m_1^2}{\rho}\zeta_{10} + \frac{m_2^2}{1-\rho}\zeta_{01}. \quad (1.1.7)$$

□

**Remark 1.1.3.** See Lehmann (1999), theorem 6.1.3. ♦

**Theorem 1.1.6 (Asymptotic Normality).** Consider the multi-sample $U$-statistic in (1.1.5).

i. If $0 < \zeta_{10}, \zeta_{01} < \infty$, then as $n \to \infty$:

$$\sqrt{n}\big(\hat{U}_n - \theta\big) \xrightarrow{\mathcal{L}} N\big(0, \sigma^2\big),$$

where $\sigma^2$ is the limiting variance in (1.1.7).

ii. If $0 < \zeta_{10}, \zeta_{01} \leq \zeta_{m_1 m_2} < \infty$, then:

$$\frac{\sqrt{n}\big(\hat{U}_n - \theta\big)}{\sqrt{\mathrm{Var}(U_n)}} \xrightarrow{\mathcal{L}} N\big(0, 1\big),$$

where $\mathrm{Var}(U_n)$ is the finite sample variance in (1.1.6).

□

**Remark 1.1.4.** See Lehmann (1999), theorem 6.1.4. ♦

## 1.7 Mann-Whitney Statistic

**Example 1.1.3.** Suppose $(X_i) \overset{\text{IID}}{\sim} F_X$ and $(Y_j) \overset{\text{IID}}{\sim} F_Y$. Consider the parameter $\theta = P(X \leq Y)$. The corresponding $U$-statistic is:

$$U_n = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(X_i \leq Y_j).$$

Identify $h(X, Y) = I(X \leq Y)$ as the kernel. The first projections of $h$ are:

$$h_{10}(x) = E\big\{h(X,Y)|X = x\big\} = P(x \leq Y) = 1 - F_Y(x)$$
$$h_{01}(y) = E\big\{h(X,Y)|Y = y\big\} = P(X \leq y) = F_X(y).$$

Suppose that, under $H_0$, $F_X = F_Y = F_0$, with $F_0$ continuous, then by the probability integral theorem $h_{10}(X) = 1 - F_0(X) \sim U(0,1)$ and $h_{01}(Y) = F_0(Y) \sim U(0,1)$.

Now, under $H_0$,

$$\theta = E\big\{I(X \leq Y)\big\} = E\big\{F_0(Y)\big\} = E\big\{U(0,1)\big\} = \frac{1}{2},$$
$$\zeta_{10} = \mathrm{Var}\big\{h_{10}(X)\big\} = \mathrm{Var}\big\{U(0,1)\big\} = \frac{1}{12},$$
$$\zeta_{01} = \mathrm{Var}\big\{h_{01}(Y)\big\} = \mathrm{Var}\big\{U(0,1)\big\} = \frac{1}{12}.$$

Therefore, under $H_0$,

$$\sqrt{n}\left(U_n - \frac{1}{2}\right) \xrightarrow{\mathcal{L}} N\left\{0, \frac{1}{12}\left(\frac{1}{\rho} + \frac{1}{1-\rho}\right)\right\},$$

where $n = n_1 + n_2$, and $\rho = \lim_{n\to\infty} n_1/n$. ♠

## 1.8 Joint Distribution

**Theorem 1.1.7** (**Joint Asymptotic Normality**). Suppose $(Y_i)_{i=1}^{n}$ is a random sample, and consider two distinct, one-sample $U$-statistics $(U_n^{(1)}, U_n^{(2)})$ of the form (1.1.1). Let $\theta_k = E\big(U_n^{(k)}\big)$, and define:

$$\begin{aligned}\gamma_k = \mathrm{Cov}\big\{&h^{(1)}(Y_1, \cdots, Y_k, Y_{k+1}, \cdots, Y_{m_1}),\\ &h^{(2)}(Y_1, \cdots, Y_k, Y_{k+1}, \cdots, Y_{m_2})\big\},\end{aligned}$$

where $h^{(k)}$ is the *kernel* of $U_n^{(k)}$. If

$$0 < \mathrm{Var}\big\{h^{(1)}(Y_1, \cdots, Y_{m_1})\big\} < \infty, \qquad 0 < \mathrm{Var}\big\{h^{(2)}(Y_1, \cdots, Y_{m_2})\big\} < \infty,$$

then:

$$\sqrt{n}\begin{pmatrix} U_n^{(1)} - \theta_1 \\ U_n^{(2)} - \theta_2 \end{pmatrix} \xrightarrow{\mathcal{L}} N\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} m_1^2 \zeta_1^{(1)} & m_1 m_2 \gamma_1 \\ m_1 m_2 \gamma_1 & m_2^2 \zeta_1^{(2)} \end{pmatrix}\right\}$$

where $\zeta_1^{(1)} = \mathrm{Var}\big\{h_1^{(1)}(Y_1)\big\}$ and $\zeta_2^{(2)} = \mathrm{Var}\big\{h_1^{(2)}(Y_1)\big\}$. □

**Remark 1.1.5.** See Lehmann (1999), theorem 6.1.6. ♦

**Example 1.1.4.** Suppose $(Y_i) \overset{\text{IID}}{\sim} F_Y$. Consider the joint asymptotic distribution of the sample mean and variance:

$$U_n^{(1)} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \qquad U_n^{(2)} = \frac{2}{n(n-1)}\sum_{i<j}\frac{1}{2}(Y_i - Y_j)^2.$$

For the mean sample mean $h_1^{(1)}(Y_1) = h^{(1)}(Y_1) = Y_1$, hence:

$$\begin{aligned}\theta_1 &= E\big\{h_1^{(1)}(Y_1)\big\} = E\big\{Y_1\big\} = \mu,\\ \zeta_1^{(1)} &= \mathrm{Var}\big\{h_1^{(1)}(Y_1)\big\} = \mathrm{Var}(Y_1) = \sigma^2.\end{aligned}$$

For the variance, the first projection is:

$$\begin{aligned}h_1^{(2)}(y) &= E\big\{h^{(2)}(Y_1, Y_2)|Y_1 = y\big\} = \frac{1}{2}E\big\{(y - Y_2)^2\big\} = \frac{1}{2}E\big\{(Y_2 - \mu + \mu - y)^2\big\}\\ &= \frac{1}{2}E\big\{(Y_2 - \mu)^2\big\} + \frac{1}{2}(y - \mu)^2 = \frac{1}{2}\sigma^2 + \frac{1}{2}(y - \mu)^2.\end{aligned}$$

Thus:

$$\begin{aligned}\theta_2 &= E\big\{h_1^{(2)}(Y_1)\big\} = \frac{1}{2}\sigma^2 + \frac{1}{2}E\big\{(Y_1 - \mu)^2\big\} = \sigma^2,\\ \zeta_1^{(2)} &= \mathrm{Var}\big\{h_1^{(2)}(Y_1)\big\} = \frac{1}{4}\mathrm{Var}\big\{(Y_1 - \mu)^2\big\} = \frac{1}{4}\big(\mu_4 - \sigma^4\big),\end{aligned}$$

where $\mu_4 = E\big\{(Y_1 - \mu)^4\big\}$.

The covariance term is:

$$\begin{aligned}\gamma = \text{Cov}\big\{h^{(1)}(Y_1), h^{(2)}(Y_1, Y_2)\big\} &= \frac{1}{2}\text{Cov}\big\{Y_1, (Y_1 - Y_2)^2\big\} \\ &= \frac{1}{2}\text{Cov}\big\{(Y_1 - \mu), (Y_1 - \mu + \mu - Y_2)^2\big\} \\ &= \frac{1}{2}\text{Cov}\big\{(Y_1 - \mu), (Y_1 - \mu)^2 + (Y_2 - \mu)^2 - 2(Y_1 - \mu)(Y_2 - \mu)\big\}\end{aligned}$$

By independence of $Y_1$ from $Y_2$:

$$\gamma = \frac{1}{2}\text{Cov}\big\{(Y_1 - \mu), (Y_1 - \mu)^2\big\} - \text{Cov}\big\{(Y_1 - \mu), (Y_1 - \mu)(Y_2 - \mu)\big\} = \frac{1}{2}\mu_3 - 0,$$

where $\mu_3 = E\big\{(Y_1 - \mu)^3\big\}$.

Overall, noting that $m_1 = 1$ and $m_2 = 2$, the joint asymptotic distribution of the sample mean and covariance is:

$$\sqrt{n}\begin{pmatrix} U_n^{(1)} - \mu \\ U_n^{(2)} - \sigma^2 \end{pmatrix} \xrightarrow{\mathcal{L}} N\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{pmatrix}\right\}.$$

This result may be confirmed using the $M$-estimation framework. ♠

# Projection

## 2.1 Projections of Random Variables

**Definition 1.2.1.** Let $\mathcal{V}$ denote the linear space of random variables with finite second moments, and let $\mathcal{U}$ denote a closed linear subspace. The projection of $V \in \mathcal{V}$ onto $\mathcal{U}$ is the random variable $\hat{U} = \Pi(V|\mathcal{U})$ satisfying:

i. $\hat{U} \in \mathcal{U}$,

ii. $E\big\{(V - \hat{U})U\big\} = 0$ for $\forall U \in \mathcal{U}$.

■

**Proposition 1.2.1.** Suppose $V \in \mathcal{V}$, then $\hat{U} = \Pi(V|\mathcal{U})$ is the closest element in $\mathcal{U}$ to $V$ in the sense that:

$$E(\hat{U} - V)^2 \leq E(U - V)^2 \text{ for } \forall U \in \mathcal{U}.$$

Moreover, the inequality is strict unless $E(U - \hat{U})^2 = 0$. ♦

**Proof.**

$E(U-V)^2 = E(U-\hat{U}+\hat{U}-V)^2 = E(U-\hat{U})^2 + E(\hat{U}-V)^2 + 2E\big\{(U-\hat{U})(\hat{U}-V)\big\}.$

Since $U \in \mathcal{U}$ and $\hat{U} \in \mathcal{U}$, the difference $(U-\hat{U}) \in \mathcal{U}^{\perp}$, thus:

$$E\big\{(U-\hat{U})(\hat{U}-V)\big\} = 0.$$

Thus, $E(U-V)^2 \geq E(\hat{U}-V)^2$, equality holding if and only if $E(U-\hat{U})^2 = 0$. ■

**Proposition 1.2.2.** Suppose the closed, linear subspace $\mathcal{U}$ contains the constant random variable, then:

i. $E(V) = E(\hat{U})$,

ii. $\mathrm{Cov}(V-\hat{U}, U) = 0$ for $\forall U \in \mathcal{U}$.

iii. $\mathrm{Cov}(V, \hat{U}) = \mathrm{Var}(\hat{U})$.

♦

**Proof.** (i.) If $\mathcal{U}$ contains the constant random variable $U = 1$, then the orthogonality condition implies $E\big\{(V-\hat{U})1\big\} = E(V-\hat{U}) = 0$, so $E(V) = E(\hat{U})$.

(ii.)

$$\mathrm{Cov}(V-\hat{U}, U) = E\big\{(V-\hat{U})U\big\} - E(V-\hat{U})E(U) = 0.$$

(iii.) Since $(V-\hat{U})$ is orthogonal to $\hat{U}$:

$$E\big\{(V-\hat{U})\hat{U}\big\} = 0 \implies E(V\hat{U}) = E(\hat{U}^2).$$

The covariance of $V$ and $\hat{U}$ is:

$$\mathrm{Cov}(V, \hat{U}) = E(V\hat{U}) - E(V)E(\hat{U}) = E(\hat{U}^2) - E^2(\hat{U}) = \mathrm{Var}(\hat{U}).$$

■

**Proposition 1.2.3.** Suppose $\mathcal{U}_n$ is a sequence of closed linear subspaces, each containing the constant random variable. Let denote $V_n$ is a sequence of random variables with $\hat{U}_n = \Pi(V_n|\mathcal{U})$. If $\lim_{n\to\infty} \mathrm{Var}(\hat{U}_n)/\mathrm{Var}(V_n) = 1$, then:

$$\frac{V_n - E(V_n)}{\sqrt{\mathrm{Var}(V_n)}} = \frac{\hat{U}_n - E(\hat{U}_n)}{\sqrt{\mathrm{Var}(\hat{U}_n)}} + o_p(1). \tag{1.2.8}$$

♦

**Proof.** Recall that convergence in mean square implies convergence in probability. The variance of the difference is:

$$\text{Var}\left\{\frac{V_n - E(V_n)}{\sqrt{\text{Var}(U_n)}} - \frac{\hat{U}_n - E(\hat{U}_n)}{\sqrt{\text{Var}\hat{U}_n}}\right\} = 1 + 1 - 2\frac{\text{Cov}(V_n, \hat{U}_n)}{\sqrt{\text{Var}(V_n)\text{Var}(\hat{U}_n)}}$$

Since $\text{Cov}(V_n, \hat{U}_n) = \text{Var}(\hat{U}_n)$:

$$\begin{aligned}\lim_{n\to\infty}\text{Var}\left\{\frac{V_n - E(V_n)}{\sqrt{\text{Var}(U_n)}} - \frac{\hat{U}_n - E(\hat{U}_n)}{\sqrt{\text{Var}\hat{U}_n}}\right\} &= 2 - 2\lim_{n\to\infty}\frac{\text{Var}(\hat{U}_n)}{\sqrt{\text{Var}(V_n)\text{Var}(\hat{U}_n)}} \\ &= 2 - 2\sqrt{\lim_{n\to\infty}\frac{\text{Var}(\hat{U}_n)}{\text{Var}(V_n)}} = 2 - 2 = 0.\end{aligned}$$

Thus the difference:

$$\frac{V_n - E(V_n)}{\sqrt{\text{Var}(V_n)}} - \frac{\hat{U}_n - E(\hat{U}_n)}{\sqrt{\text{Var}(\hat{U})_n}} \xrightarrow{L^2} 0.$$

■

## 2.2 Projection onto Sums

**Theorem 1.2.1.** Suppose $(Y_i)$ are independent. Consider the class $\mathcal{S}$ of random variables of the form:

$$S = \sum_{i=1}^{n} g_i(Y_i),$$

for measurable functions $g_i(\cdot)$ with finite second moments $E\{g_i^2(Y_i)\}$.The **Hájek projection** of $V \in \mathcal{V}$ onto this class is given by:

$$\hat{S} = \Pi(V|\mathcal{S}) = \sum_{i=1}^{n}\{E(V|Y_i)\} - (n-1)E(V).$$

□

**Proof.** Towards finding $E(\hat{S}|Y_j)$, consider the expectation of the random variable $E(V|Y_i)$ given $Y_j$. For $i = j$:

$$E\{E(V|Y_i)|Y_j\} = E(V|Y_i).$$

By independence, for $i \neq j$:

$$E\{E(V|Y_i)|Y_j\} = E\{E(V|Y_i)\} = E(V).$$

Thus the expectation of the Hajek projection $\hat{S}$ given $Y_j$ is:

$$E(\hat{S}|Y_j) = E(V|Y_j) + \sum_{i\neq j} E\big\{E(V|Y_i)|Y_j\big\} - (n-1)E(V) = E(V|Y_j).$$

For each $j$, the residual $(V - \hat{S})$ is orthogonal to $g_j(Y_j)$:

$$\begin{aligned} E\big\{(V-\hat{S})g_j(Y_j)\big\} &= E\Big[E\big\{(V-\hat{S})g_j(Y_j)|Y_j\big\}\Big] \\ &= E\Big[\big\{E(V|Y_j) - E(\hat{S}|Y_j)\big\}g_j(Y_j)\Big] = 0. \end{aligned}$$

Thus $(V - \hat{S})$ is orthogonal to random variables of the form $S = \sum_{i=1}^{n} g_i(Y_i)$, i.e. to $\mathcal{S}$. Finally, since $\hat{S} \in \mathcal{S}$ and $(V - \hat{S})$ is orthogonal to $\mathcal{S}$, conclude that $\hat{S}$ is the projection of $V$ onto $\mathcal{S}$. ■

**Proposition 1.2.4.** The Hajék projection of the $U$-statistic in (1.1.1) is:

$$\hat{S} - \theta = \frac{m}{n}\sum_{i=1}^{n}\big\{h_1(Y_i) - \theta\big\}. \tag{1.2.9}$$

♦

**Proof.** Recall that the $U$-statistic takes the form:

$$U_n(h) = \binom{n}{m}^{-1} \sum_{1\leq i_1 < \cdots < i_m \leq n} h(Y_{i_1}, \cdots, Y_{i_m}).$$

Taking the expectation conditional on $Y_i$:

$$E\big(U_n|Y_i = y\big) = \binom{n}{m}^{-1} \sum_{1\leq i_1 < \cdots < i_m \leq n} E\big\{h(Y_{i_1}, \cdots, Y_{i_m})|Y_i = y\big\}$$

Consider cases. If $i \in \{i_1, \cdots, i_m\}$, then:

$$E\big\{h(Y_{i_1}, \cdots, Y_{i_m})|Y_i = y\big\} = h_1(y),$$

whereas if $i \notin \{i_1, \cdots, i_m\}$, then:

$$E\big\{h(Y_{i_1}, \cdots, Y_{i_m})|Y_i = y\big\} = \theta.$$

There are $\binom{n-1}{m-1}$ ways to choose $(i_1, \cdots, i_m)$ so as to include $i$, and $\binom{n-1}{m}$ to choose $(i_1, \cdots, i_m)$ so as to exclude $i$. Thus:

$$E\big(U_n|Y_i = y\big) = \binom{n}{m}^{-1}\left\{\binom{n-1}{m-1}h_1(y) + \binom{n-1}{m}\theta\right\} = \frac{m}{n}h_1(y) + \frac{n-m}{n}\theta.$$

The Hajek projection is:

$$\hat{S} = \sum_{i=1}^{n}\left\{\frac{m}{n}h_1(y) + \frac{n-m}{n}\theta\right\} - (n-1)\theta = \frac{m}{n}\sum_{i=1}^{n}\big\{h_1(Y_i) - \theta\big\} + n\theta - (n-1)\theta.$$

■

# References


- Serfling. *Approximation Theorems of Mathematical Statistics* (1980); Chapter 5.
- van der Vaart. *Asymptotic Statistics* (1998); Chapters 11 & 12.
- Lehmann. *Elements of Large Sample Theory* (1999); Chapter 6.