# 9 U-STATISTICS

Suppose $X_1, X_2, \ldots, X_n$ are $P \in \mathscr{P}$ i.i.d. with CDF $F$. Our goal is to estimate the expectation $t(P) = Eh(X_1, X_2, \ldots, X_m)$. Note that this expectation requires more than one $X$ in contrast to $EX$, $EX^2$, or $Eh(X)$. One example is $E|X_2 - X_1|$ or $P((X_1, X_2) \in S)$.

For $Eh(X_1)$, the empirical estimator is asymptotically optimal. Now, U-statistics generalize the idea. The advantage of using U-statistics is unbiasedness.

Notation. Let $t(P) = \int \cdots \int h(x_1, \ldots, x_m) dF(x_1) \cdots dF(x_m)$ where $h$ is a known function; $h$ is called kernel. Assume that, in the following, without loss of generality, $h$ is symmetric, i.e. $h(x_1, x_2) = h(x_2, x_1)$. If $h$ is not symmetric, we can replace it with $h^*(x_1, \ldots, x_m) = (m!)^{-1} \sum_{\pi \in \Pi} h(x_{\pi(1)}, \ldots, x_{\pi(m)})$ where $\pi : \mathbb{N} \mapsto \mathbb{N}$ and $\Pi$ is set of all possible permutations of $\{1, \ldots, m\}$. Note that $h^*$ is also unbiased,

$$Eh^* = (m!)^{-1} \sum_{\pi \in \Pi} Eh(X_{\pi(1)}, \ldots, X_{\pi(m)}) \overset{\text{i.i.d}}{=} (m!)^{-1} \sum_{\pi \in \Pi} Eh(X_1, \ldots, X_m) = t(P).$$

Definition 9.1. Suppose $h : \mathbb{R}^m \mapsto \mathbb{R}$ is symmetric in its arguments. The U-statistic for estimating $t(P) = Eh(X_1, \ldots, X_m)$ is a symmetric average

$$u = u(X_1, \ldots, X_m) = \frac{1}{\binom{n}{m}} \sum \cdots \sum_{1 \le i_1 < i_2 < \cdots < i_m \le n} h(X_{i_1}, \ldots, X_{i_m}).$$

Example. Suppose $m = 2$, then

$$u = u(X_1, X_2) = \frac{1}{\binom{n}{2}} \sum\sum_{i_1 < i_2} h(X_{i_1}, X_{i_2}) = \frac{2}{n(n-1)} \sum\sum_{i<j} h(X_i, X_j) \tag{1}$$

$$= \frac{1}{n(n-1)} \sum\sum_{i \ne j} h(X_i, X_j) \tag{2}$$

Remark.

$$Eu = \frac{1}{\binom{n}{m}} \sum \cdots \sum_{1 \le i_1 < i_2 < \cdots < i_m \le n} Eh(X_{i_1}, \ldots, X_{i_m}) = Eh(X_1, \ldots, X_m) = t(P).$$

Example 9.2.

(a) Suppose $t(P) = EX_1 = \int x_1 dF(x_1)$. Then, $h(x_1) = x_1$ and

$$u(X_1, \ldots, X_m) = \frac{1}{\binom{n}{1}} \sum_{i_1} h(X_{i_1}) = \frac{1}{n} \sum_i X_i = \bar{X}.$$

(b) Suppose $t(P) = (EX_1)^2 = \left(\int x_1 dF(x_1)\right)^2$. Then $u^2 = \bar{X}^2$ from (a) is biased since

$$E(\bar{X}^2) = n^{-2} \left[ \sum_{i=1}^{n} \sum_{j \ne i} \underbrace{EX_i X_j}_{=(EX_1)^2} + \sum_{i=1}^{n} \underbrace{E(X_i^2)}_{EX_1^2 \ne (EX_1)^2} \right] \ne (EX_1)^2.$$

Now, write $t(P) = \int\int x_1 x_2 dF(x_1)dF(x_2)$ and $h(x_1, x_2) = x_1 x_2$. Then

$$u = u(X_1, \ldots, X_n) = \frac{2}{n(n-1)} \sum\sum_{i<j} X_i X_j.$$

(c) Suppose $t(P) = P(X_1 \le t_0) = F(t_0) = \int h(x_1)dF(x_1)$. Then $h(x_1) = 1_{(-\infty, t_0]}(x_1)$ and

$$u = n^{-1} \sum_i 1\{X_i \le t_0\} = \hat{F}(t_0)$$

which is just the empirical CDF.

(d) Suppose $t(P) = \text{Var}X_1 = \int\int \frac{x_1^2 + x_2^2 - 2x_1x_2}{2} dF(x_1)dF(x_2)$. Then $h(x_1, x_2) = (x_1^2 + x_2^2 - 2x_1x_2)/2 = (x_1 - x_2)^2/2$. Note that

$$Eh(x_1, x_2) = \{\text{Var}(X_1) + [E(X_1)]^2 + \text{Var}(X_2) + (EX_2)^2 - 2E(X_1X_2)\}/2 \overset{\text{i.i.d.}}{=} \text{Var}(X_1).$$

Hence,

$$\begin{aligned} u &= \frac{2}{n(n-1)} \sum\sum_{i<j} h(X_i, X_j) \\ &= \frac{1}{n(n-1)} \sum\sum_{i \ne j} \frac{X_i^2 + X_j^2 - 2X_iX_j}{2} \\ &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \\ &\ne \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2. \end{aligned}$$

Note that $\hat{\sigma}^2$ is a biased estimator and $u$ is an unbiased estimator.

(e) Suppose $t(P) = E|X_1 - X_2| = \int\int |x_1 - x_2| dF(x_1)dF(x_2)$ (measure of dispersion). Then

$$u = \frac{2}{n(n-1)} \sum\sum_{i<j} |X_i - X_j|.$$

$u$ is called Gini's Mean Difference.

(f) Suppose $t(P) = P(X_1 + X_2 \le 0) = \int\int 1\{x_1 + x_2 \le 0\} dF(x_1)dF(x_2)$. Then, $h(x_1, x_2) = 1\{x_1 + x_2 \le 0\}$ and

$$u = \frac{2}{n(n-1)} \sum\sum_{i<j} 1\{x_i + x_j \le 0\}.$$

Remark 9.3 (Preliminary Remark). Write $\underline{X}_{(n)} = (X_{(1)}, \dots, X_{(n)})$ as the order statistic. U-statistic can be regarded as conditional expectation given $\underline{X}_{(n)}$. For $m = 1$,

$$u = n^{-1} \sum_i h(X_i) = n^{-1} \sum_i h(X_{(i)}) = E_{\hat{F}}[h(X_1) | \underline{X}_{(n)}].$$

For m=2,

$$u = \frac{1}{\binom{n}{2}} \sum\sum_{i<j} h(X_i, X_j) = \frac{1}{\binom{n}{2}} \sum\sum_{i<j} h(X_{(i)}, X_{(j)}) = E_{\hat{F}}[h(X_1, X_2) | \underline{X}_{(n)}].$$

For arbitrary $m$,

$$u = E_{\hat{F}}[h(X_1, X_2, \dots, X_m) | \underline{X}_{(n)}].$$

Now: any unbiased estimator $S = S(X_1, \dots, X_n)$ can be improved by its U-statistic version (or $S^*$ if $S$ is not symmetric)

$$u = E_{\hat{F}}[S | \underline{X}_{(n)}].$$

For example, $X_1$ is unbiased for $EX$. Now, $u = E_{\hat{F}}(X_1 | \underline{X}_{(n)}) = n^{-1} \sum_{i=1}^n X_{(i)} = \bar{X} = n^{-1} \sum_{i=1}^n X_i = E_{\hat{F}}(X_1)$.

Theorem 9.4. Let $S = S(X_1, \dots, X_n)$ be an unbiased estimator of $t(P)$ with corresponding U-statistic $u$. Then, $u$ is unbiased as well and $\mathrm{Var} u \le \mathrm{Var} S$ with the equality holding if $P(u = S) = 1$.

Proof. u is unbiased:

$$E(u) = \underbrace{E[E_{\hat{F}}(S | \underline{X}_{(n)})]}_{E(S)} = t(P).$$

Since both $u$ and $S$ are unbiased we show $Eu^2 \le ES^2$,

$$Eu^2 = E[E_{\hat{F}}^2(S | \underline{X}_{(n)})] \overset{\text{Jensen inequ.}}{\le} E[E_{\hat{F}}(S^2 | \underline{X}_{(n)})] = E(S^2)$$

with "=" if the distribution of $E_{\hat{F}}(S | \underline{X}_{(n)})$ is degenerate with $E_{\hat{F}}(S | \underline{X}_{(n)}) = S$ almost surely. □

Note: This also follows from the Rao-Blackwell Theorem: Taking conditional expectation of an unbiased statistic conditional on a sufficient statistic (eg. $\underline{X}_{(n)}$ here) will give us an estimator which is as least as good in the sense of lower risk/variance.

Remark 9.5 (The Variance for $m \le 2$ (heuristics)).

$m = 1$.

$$\mathrm{Var} u = \mathrm{Var}\left( \frac{1}{n} \sum_{i=1}^n h(X_i) \right) = \frac{1}{n} \mathrm{Var} h(X_1) = O(1/n).$$

$m = 2$. Since

$$u = \frac{1}{\binom{n}{2}} \sum_{i<j} h(X_i, X_j) = \frac{2}{n(n-2)} \sum_{i<j} h(X_i, X_j)$$
$$= \frac{1}{\binom{n}{2}} [\underbrace{h(X_1, X_2) + h(X_1, X_3) + \cdots + h(X_{n-1}, X_n)}_{\text{not independent}}],$$

then

$$\mathrm{Var}\, u = \mathrm{Var}\left(\frac{2}{n(n-1)} \sum_i \sum_{j>i} h(X_i, X_j)\right)$$
$$= \left(\frac{2}{n(n-1)}\right)^2 \sum_i \sum_{j>i} \sum_k \sum_{l>k} \underbrace{\mathrm{Cov}[h(X_i, X_j), h(X_k, X_l)]}_{=0 \text{ if } i,j,k,l \text{ different}}.$$

The second largest term (three sums): e.g. $i = k$ but $k, j, l$ are different ($\simeq u^3$)

$$\mathrm{Var}\, u \sim \left(\frac{1}{n(n-1)}\right)^2 n^3 \sim \frac{n^3}{n^4} = \frac{1}{n} \text{ same order as } m = 1.$$

Thus, we expect in general, $\mathrm{Var}\, u = O(1/n)$.

Notation. $h_i(x_1, \ldots, x_i) = E[h(X_1, \ldots, X_m) | X_1 = x_1, \ldots, X_i = x_i]$ and $\sigma_i^2 = \mathrm{Var} h_i(X_1, \ldots, X_i)$

Lemma 9.6.

(a) $Eh_i(X_1, \ldots, X_i) = t(P)(= Eh(X_1, \ldots, X_m))$ for all $1 \le i \le m$.
(b) $\mathrm{Cov}[h(X_1, \ldots, X_i, X_{i+1}, \ldots, X_m), h(X_1, \ldots, X_i, X'_{i+1}, \ldots, X'_m)] = \sigma_i^2$

Proof. We only consider $m = 2$ here.

(a)

$$Eh_2(X_1, X_2) = E\{E[h(X_1, X_2) | X_1, X_2]\} = E[h(X_1, X_2)] = t(P)$$
$$Eh_1(X_1, X_2) = E\{E[h(X_1, X_2) | X_1]\} = E[h(X_1, X_2)] = t(P)$$

(b) $i = 2$.

$$\mathrm{Cov}[h(X_1, X_2), h(X_1, X_2)] = \mathrm{Var}[h(X_1, X_2)] = \mathrm{Var} h_2(X_1, X_2) = \sigma_2^2.$$

The second equality is because $h_2(X_1, X_2) = E[h(X_1, X_2) | X_1, X_2] = h(X_1, X_2)$.

$i = 1$.

$$\mathrm{Cov}[h(X_1, X_2), h(X_1, X'_2)] = E[h(X_1, X_2) h(X_1, X'_2)] - \underbrace{E[h(X_1, X_2)] E[h(X_1, X'_2)]}_{=\{E[h(X_1, X_2)]\}^2 = [t(P)]^2}.$$

Note that first term on the right hand side can be computed by

$$\begin{aligned} E[h(X_1,X_2)h(X_1,X_2')] &= E\{E[h(X_1,X_2)h(X_1,X_2')|X_1]\} \\ &= E\{E[h(X_1,X_2)|X_1]E[h(X_1,X_2')|X_1]\} \\ &= Eh_1^2(X_1). \end{aligned}$$

□

Theorem 9.7 (Hoeffding).

(a) The variance of the U-statistic is

$$\operatorname{Var} u = \frac{1}{\binom{n}{m}} \sum_{i=1}^{m} \binom{m}{i} \binom{n-m}{m-i} \sigma_i^2.$$

Note that one can compute $\sigma_i^2$ from Lemma 9.6.

(b) If $\sigma_i^2 > 0$ and $\sigma_i^2 < \infty$ for $i = 1, 2, \dots, m$, then

$$\operatorname{Var}(\sqrt{n}\, u) \to m^2 \sigma_i^2.$$

Proof. (a) For general proof, see Lee "U-statistics" (1990). We only prove the case of $m = 2$. We want to show that

$$\operatorname{Var} u = \frac{1}{\binom{n}{2}} \left( \binom{2}{1}\binom{n-2}{2-1}\sigma_1^2 + \binom{2}{2}\binom{n-2}{2-2}\sigma_2^2 \right) = \frac{1}{\binom{n}{2}} [2(n-2)\sigma_1^2) + \sigma_2^2]$$

with $\sigma_1^2 = \operatorname{Cov}[h(X_1,X_2), h(X_1,X_2')]$ and $\sigma_2^2 = \operatorname{Cov}[h(X_1,X_2), h(X_1,X_2)] = \operatorname{Var}[h(X_1,X_2)]$.
From Remark 9.5 we have

$$\begin{aligned} \operatorname{Var} u &= \left(\frac{1}{\binom{n}{2}}\right)^2 \sum_i \sum_{j>i} \sum_k \sum_{l>k} \operatorname{Cov}[h(X_i,X_j), h(X_k,X_l)] \\ &= \frac{1}{\binom{n}{2}} \underbrace{\sum_i \sum_{j>i} \sum_k \sum_{l>k} \frac{\operatorname{Cov}[h(X_i,X_j), h(X_k,X_l)]}{\binom{n}{2}}}_{=2(n-2)\sigma_1^2+\sigma_2^2}. \end{aligned}$$

Case 1 $i, j, k, l$ are all different. Then $\operatorname{Cov} = 0$.
Case 2 $i = k$ and $j = l$:

$$\operatorname{Cov}[h(X_i,X_j), h(X_k,X_l)] = \operatorname{Cov}[h(X_i,X_j), h(X_i,X_j)] = \operatorname{Var}[h(X_i,X_j)] = \sigma_2^2.$$

Note that the number of ways to choose $i, j$ out of $\{1, 2, \dots, n\}$ is $\binom{n}{2}$. This gives us $\binom{n}{2}\sigma_2^2 / \binom{n}{2} = \sigma_2^2$.

Case 3 ($i=k$ and $j\neq l$) or ($i\neq k$ and $j=l$).

$$\operatorname{Cov}[h(X_i,X_j),h(X_k,X_l)]=\sigma_1^2.$$

Note that the number of ways to choose $i,j,k,l$ is

$$\underbrace{n}_{i}\cdot\underbrace{(n-1)}_{j\neq i}\cdot\frac{1}{2}\cdot\underbrace{n-2}_{l}\cdot\underbrace{2}_{\text{or}}=\binom{n}{2}2(n-2).$$

This gives

$$\frac{\binom{n}{2}2(n-2)\sigma_1^2}{\binom{n}{2}}=2(n-2)\sigma_1^2.$$

(b) Consider the variance formula from (a)

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

is large if $k$ is large. This implies that $\binom{n-m}{m-i}$ is large if $m-i$ is large, i.e. $i=1$. Thus the terms of the sum are dominated by the $i=1$ term, i.e., by

$$\binom{m}{1}\binom{n-m}{m-1}\sigma_1^2\frac{1}{\binom{n}{m}}\sim m\frac{n^{m-1}}{(m-1)!}\sigma_1^2\frac{1}{n^m/m!}=m^2\frac{\sigma_2^2}{n}.$$

Hence, $\operatorname{Var}(\sqrt{n}u)\to m^2\sigma_1^2$.

□

**Theorem 9.8.**

$$\sqrt{n}(u-t(P))\xrightarrow{\mathscr{D}}N(0,m^2\sigma_1^2).$$

Proof. For general proof, see p. 178 of Serfling.

$$u_n=\sum_{c=1}^{m}\binom{m}{c}\binom{n}{c}^{-1}\sum_{1\leq i_1<\cdots<i_c\leq n}g_c(X_{i_1},\ldots,X_{i_c})+o_p(n^{-1/2}).$$

For $m=2$,

$$T_n=n^{1/2}(u_n-t(P))=n^{-1/2}\sum 2[h_1(X_i)-t(P)]+o_p(1):=T_n^*.$$

We want to show that $E(T_n-T_n^*)\to 0$ ($\Rightarrow T_n-T_n^*=o_p(1)$).

$$\begin{aligned}
E(T_n-T_n^*)^2&=\operatorname{Var}(T_n-T_n^*)\\
&=\operatorname{Var}(T_n)+\operatorname{Var}(T_n^*)-2\operatorname{Cov}(T_n,T_n^*)\\
&=\operatorname{Var}[\sqrt{n}(u-t(P))]+\operatorname{Var}\left(\frac{2}{\sqrt{n}}\sum[h_1(X_i)-t(P)]\right)\\
&\quad-2\operatorname{Cov}\left(\sqrt{n}(u-t(P)),\frac{2}{\sqrt{n}}\sum[h_1(X_i)-t(P)\right)\\
&=\underbrace{\operatorname{Var}(\sqrt{n}u)}_{\to m^2\sigma_1^2}+\underbrace{\frac{4}{n}\sum\operatorname{Var}[h_1(X_i)]}_{4\sigma_1^2}-\underbrace{4\sum\operatorname{Cov}(u,h_1(X_i))}_{\overset{(*)}{=}2\sigma_1^2}\overset{n\to\infty}{\to}0.
\end{aligned}$$

For $(*)$,

$$\sum \mathrm{Cov}(u, h_1(X_i)) = \frac{1}{n(n-1)} \sum_i \sum_k \sum_{l \neq k} \underbrace{\mathrm{Cov}[h(X_l, X_k), h_1(X_i)]}_{\stackrel{(\dagger)}{=} \sigma_1^2, \text{ if } k=i \text{ or } l=i} = 2\sigma_1^2.$$

The number of non-zero terms is $2n(n-1)$.

For $(\dagger)$,

$$\begin{aligned}
\sigma_1^2 &\stackrel{(9.6)}{=} \mathrm{Cov}[h(X_1, X_2), h(X_1, X_2')] \\
&= E[h(X_1, X_2) h(X_1, X_2')] - [t(P)]^2 \\
&= E\{E[h(X_1, X_2) h(X_1, X_2') | X_1, X_2]\} - [t(P)]^2 \\
&= E\{h(X_1, X_2) \underbrace{E[h(X_1, X_2') | X_1, X_2]}_{=E(h(X_1, X_2)|X_1) = h_1(X_1)}\} - [t(P)]^2 \\
&= \mathrm{Cov}[h(X_1, X_2), h_1(X_1)].
\end{aligned}$$

□

Example 9.9. Suppose $t(P) = P(X_1 + X_2 > 0)$, $m = 2$, $h(X_1, X_2) = 1(X_1 + X_2 > 0)$. Then

$$u = \frac{1}{n(n-1)} \sum_{i<j} 1(X_i + X_j > 0)$$

and, by Theorem 9.8, $\sqrt{n}(u - P(X_1 + X_2 > 0)) \xrightarrow{\mathscr{D}} N(0, 4\sigma_1^2)$. The next thing is to calculate $\sigma_1^2$.

$$\begin{aligned}
\sigma_1^2 &\stackrel{(9.6)}{=} \mathrm{Cov}[h(X_1, X_2), h(X_1, X_2')] \\
&= E[h(X_1, X_2) h(X_1, X_2')] - [P(X_1 + X_2 > 0)]^2 \\
&= E[1(X_1 + X_2 > 0) 1(X_1 + X_2' > 0)] - [P(X_1 + X_2 > 0)]^2 \\
&= E[1(X_1 + X_2 > 0,\ X_1 + X_2' > 0)] - [P(X_1 + X_2 > 0)]^2.
\end{aligned}$$

To obtain an explicit form we need some assumptions. Suppose, for example, F is symmetric around zero, i.e. $P(X_1 < a) = P(X_1 > -a) = P(-X_1 < a)$ which implies $X_1$ and $-X_1$ have the same distribution. Moreover, suppose F is continuous. Then $P(X_1 + X_2 > 0) = P(-X_1 - X_2 > 0) = P(X_1 + X_2 < 0)$ with $P(X_1 + X_2 = 0 = 0)$, and therefore

$$1 = P(X_1 + X_2 < 0) + P(X_1 + X_2 = 0) + P(X_1 + X_2 > 0) = 2P(X_1 + X_2 > 0) \Rightarrow P(X_1 + X_2 > 0) = \frac{1}{2}.$$

On the other hand, since $P(X_1 + X_2 > 0) = P(X_1 > -X_2) = P(X_1 > X_2)$,

$$P(X_1 + X_2 > 0, X_1 + X_2' > 0) = P(X_1 = \max\{X_1, X_2, X_2'\}) = P(X_i = \max\{X_1, X_2, X_3\}, i = 1, 2, 3) = 1/3.$$

In sum, $\sigma_1^2 = 1/3 - (1/2)^2 = 1/12$.

Remark 9.10 (Generalization: Two-Sample Problems). Suppose $X_1,\ldots,X_n$ are i.i.d. with cdf $F$, $Y_1,\ldots,Y_n$ are i.i.d. with cdf $G$, and $F$ and $G$ are unknown, $X_i$ and $Y_i$ are independent. Let

$$h(\underbrace{X_1,\ldots,X_{m_X}}_{\text{symmetric}},\underbrace{Y_1,\ldots,Y_{m_Y}}_{\text{symmetric}})$$

be a function of $m_X+m_Y$ arguments, with $m_X \le n_X$ and $m_Y \le n_Y$. We want to estimate $t(P) = Eh(X_1,\ldots,X_{m_X},Y_1,\ldots,Y_{m_Y})$. For example, $t(P) = P(X_1 < Y_1) = Eh(X_1,Y_1)$ where $h(X,Y) = 1(X < Y)$; or $t(P) = E|Y - X|$. Note also that $h(X_1,\ldots,X_{m_X},Y_1,\ldots,Y_{m_Y})$ is trivally unbiased for $t(P) = Eh$. Thus $h(X_{i_1},\ldots,X_{i_{m_X}},Y_{j_1},\ldots,Y_{j_{m_Y}})$ with $1 \le i_1 \le \ldots \le i_{m_X} \le n_X$ and $1 \le j_1 \le \ldots \le j_{m_Y} \le n_Y$ is unbiased as well. The number of combinations is $\binom{n_X}{m_X}\binom{n_Y}{m_Y}$. Our unbiased estiamtor is

$$u = \frac{1}{\binom{n_X}{m_X}\binom{n_Y}{m_Y}} \sum \cdots \sum h\left(X_{i_1},\ldots,X_{i_{m_X}},Y_{j_1},\ldots,Y_{j_{m_Y}}\right)$$

Example ($m_X = m_Y = 2$).

$$u = \frac{1}{\binom{n_X}{2}\binom{n_Y}{2}} \sum_{i<k}\sum_{j<l} h(X_i,X_k,Y_j,Y_l)$$

where $h(X_1,X_2,Y_1,Y_2) = 1(|X_2 - X_1| < |Y_2 - Y_1|)$. Then

$$u = \frac{1}{\binom{n_X}{2}\binom{n_Y}{2}} \underbrace{\sum_{i<k}\sum_{j<l} 1(|X_i - X_k| < |Y_j - Y_l|)}_{=\text{the number of } (i,j,k,l) \text{ where } y-\text{distance is larger}}$$

and $t(P) = P(|Y_2 - Y_1| > |X_2 - X_1|)$. Note that this U-statistic can be used to test $Y$ is more dispersed than $X$.

Remark 9.11 (Properties of the Two-Sample U-statistic).

(a) Formula for $\operatorname{Var} u$, see Lee or Serfling.
(b) Asymptotic variance and normality. Let $n = n_X + n_Y$ and $n_X/n \to c$ where $c \in (0,1)$. Suppose $\operatorname{Var}[h(X_1,X_2,\ldots,X_{m_X},Y_1,Y_2,\ldots,Y_{m_Y})] > 0$. Then

$$\operatorname{Var}(\sqrt{n}u) \to \sigma^2 = \frac{m_X^2}{c}\sigma_{10}^2 + \frac{m_Y^2}{c}\sigma_{01}^2$$

and

$$\sqrt{n}(u - t(P)) \xrightarrow{\mathscr{D}} N(0,\sigma^2)$$

where

$$\sigma_{10}^2 = \operatorname{Cov}[h(X_1,X_2,\ldots,X_{m_X},Y_1,Y_2,\ldots,Y_{m_Y}),h(X_1,X_2',\ldots,X_{m_X}',Y_1',Y_2',\ldots,Y_{m_Y}')$$
$$\sigma_{01}^2 = \operatorname{Cov}[h(X_1,X_2,\ldots,X_{m_X},Y_1,Y_2,\ldots,Y_{m_Y}),h(X_1',X_2',\ldots,X_{m_X}',Y_1,Y_2',\ldots,Y_{m_Y}')]$$

Example 9.12. Suppose $t(P) = P(X < Y) = E[1(X < Y)]$ (thus $m_X = m_Y = 1$). Then

$$u = \frac{1}{n_X n_Y} \sum_i \sum_j 1(X_i < Y_j)$$

and

$$\begin{aligned}
\sigma_{10}^2 &= \mathrm{Cov}[1(X_1 < Y_1), 1(X_1 < Y_1')] = E[(1(X_1 < Y_1)1(X_1 < Y_1')] - \{E[1(X_1 < Y_1)]\}^2 \\
&= P(X_1 < Y_1, X_1 < Y_1') - [P(X_1 < Y_1)]^2 \\
\sigma_{01}^2 &= \mathrm{Cov}[1(X_1 < Y_1), 1(X_1' < Y_1)] = E[(1(X_1 < Y_1)1(X_i' < Y_j)] - \{E[1(X_1 < Y_1)]\}^2 \\
&= P(X_1 < Y_1, X_1' < Y_1) - [P(X_1 < Y_1)]^2.
\end{aligned}$$

Now suppose $F = G$ is continuous. Then $P(X_1 < Y_1) = 1 - P(X_1 \geq Y_1) = 1 - P(X_1 > Y_1) = 1 - P(Y_1 > X_1)$ and hence $P(X_1 < Y_1) = 1/2$. Furthermore, $P(X_1 < Y_1, X_1 < Y_1') = P(X_1 = \min\{X_1, Y_1, Y_1'\}) = 1/3$ and $P(X_1 < Y_1, X_1' < Y_1) = P(Y_1 = \max\{X_1, X_1', Y_1\}) = 1/3$. Then $\sigma_{10}^2 = \sigma_{01}^2 = 1/12$ and the asymptotic variance is

$$\sigma^2 = \frac{m_X^2}{c}\sigma_{10}^2 + \frac{m_Y^2}{c}\sigma_{01}^2 = \frac{1}{12c(1-c)}.$$

In sum,

$$\sqrt{n}(u - P(X < Y)) \xrightarrow{\mathscr{D}} N\left(0, \frac{1}{12c(1-c)}\right)$$

Note: $n_X n_Y u = \sum_i \sum_j 1(X_i < Y_j)$ is the Mann-Whitney test statistic with $H_0 : F = G$ (and equivalent to a Wilcoxon rank sum statistic).

Definition 9.13. Consider a symmetric function $h : \mathbb{R}^m \mapsto \mathbb{R}$ with $m \leq n$. The V-statistic for estimating $t(P) = Eh(X_1, \ldots, X_m)$ is

$$V = V(X_1, \ldots, X_m) = \frac{1}{n^m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} h(X_{i_1}, \ldots, X_{i_m})$$

Remark 9.14 (Comparing U- and V-Statistics).

$m = 1$. $u = n^{-1} \sum h(X_i) = v$
$m = 2$. First note that

$$u = \frac{2}{n(n-1)} \sum_{i<j} h(X_i, X_j) = \frac{1}{n(n-1)} \sum_{i \neq j} h(X_i, X_j).$$

On the other hand,

$$\begin{aligned}
v &= \frac{1}{n^2}\sum_i\sum_j h(X_i,X_j) = \frac{1}{n^2}\sum_i\left[\sum_{j\neq i} h(X_i,X_j) + h(X_i,X_i)\right] \\
&= \frac{1}{n^2}\sum_i\sum_{j\neq i} h(X_i,X_j) + \frac{1}{n^2}\sum_i h(X_i,X_i) \\
&= \underbrace{\frac{n(n-1)}{n^2}}_{\to 1} u + \underbrace{\frac{1}{n^2}\sum_i h(X_i,X_i)}_{\xrightarrow{P} 0}.
\end{aligned}$$

Moreover, $Eh(X_1,X_2) = t(P)$.

$$\begin{aligned}
Ev &= \frac{n-1}{n} Eu + \frac{1}{n} Eh(X_1,X_1) \\
&= t(P) - \frac{1}{n} t(P) + \frac{1}{n} Eh(X_1,X_1) \\
&= t(P) + \underbrace{\frac{1}{n}\left[\underbrace{Eh(X_1,X_1) - t(P)}_{=\text{constant}}\right]}_{=\text{bias}\to 0}
\end{aligned}$$

Theorem 9.15. Let $m = 2$, $\sigma_i^2 = \operatorname{Var} h_i(X_1,\dots,X_i)$ (see 9.6), and suppose $0 < \sigma_1^2 < \infty$, $\sigma_2^2 < \infty$. Then U- and V-statistics have the same asymptotic distribution,

$$\sqrt{n}(V - t(P)) \xrightarrow{\mathscr{D}} N(0, 4\sigma_1^2).$$

Proof. From Remark 9.14,

$$\begin{aligned}
\sqrt{n}(V - t(P)) &= \sqrt{n}\left(\frac{n-1}{n} u + \frac{1}{n^2}\sum_i h(X_i,X_i) - t(P)\frac{n-1+1}{n}\right) \\
&= \sqrt{n}\left(\frac{n-1}{n}(u - t(P)) + \frac{1}{n^2}\sum h(X_i,X_i) - \frac{1}{n} t(P)\right) \\
&= \frac{n-1}{n}\sqrt{n}(u - t(P)) + \frac{1}{n^2}\sqrt{n}\left[\sum (h(X_i,X_i) - t(P))\right] \\
&= \underbrace{\frac{n-1}{n}}_{\to 1}\underbrace{\sqrt{n}(u - t(P))}_{\xrightarrow{\mathscr{D}} N(0,4\sigma_1^2)} + \frac{1}{\sqrt{n}}\underbrace{\frac{1}{n}\sum (h(X_i,X_i) - t(P))}_{\xrightarrow{P} E[h(X_i,X_i) - t(P)] = \text{constant}} \xrightarrow{\mathscr{D}} N(0, 4\sigma_1^2).
\end{aligned}$$

□

Conclusion: U- and V-statistics are asymptotically equivalent. The V-statistic is a more intuitive estimator, the U-statistic is more convenient for proofs (and unbiased).