第五讲　U 统计量

张伟平

统计与金融系

from Peter Bartlett

第五讲　U 统计量,

1.1 U Statistcs

> **Definition** Let $T(F) = \mathrm{E}_F h(X_1, \ldots, X_r)$ be an expectation functional, where $h : \mathbb{R}^r \to \mathbb{R}$ is a function that is symmetric in its arguments. In other words, $h(x_1, \ldots, x_r) = h\left(x_{\pi(1)}, \ldots, x_{\pi(r)}\right)$ for any permutation h of the integers 1 through r. Then h is called the kernel function associated with $T(F)$.

- **V-Statistics**: (Von Mises) $V_n = T\left(\hat{F}_n\right) = \mathrm{E}_{\hat{F}_n} h(X_1, \ldots, X_r) = \frac{1}{n^r} \sum_{i_1=1}^n \cdots \sum_{i_r=1}^n h(X_{i_1}, \ldots, X_{i_r})$

- since the bias in V_n is due to the duplication among the subscripts, we might sum instead over all possible subscripts satisfying $i_1 < \cdots < i_r$

U-statistics

Definition: A U-statistic of order r with kernel h is

$$U = \frac{1}{\binom{n}{r}} \sum_{i_1 < \cdots < i_r} h\left(X_{i_1}, \ldots, X_{i_r}\right)$$

where h is symmetric in its arguments.

- "U" for "unbiased." Introduced by Wassily Hoeffding in the 1940s.

U-statistics: Examples

- s_n^2 is a U-statistic of order 2 with kernel $h(x, y) = (1/2)(x - y)^2$

  - because

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}_n \right)^2$$

$$= \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\left(X_i - \overline{X}_n \right)^2 + \left(X_j - \overline{X}_n \right)^2 \right)$$

$$= \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\left(X_i - \overline{X}_n \right) - \left(X_j - \overline{X}_n \right) \right)^2$$

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{2} \left(X_i - X_j \right)^2$$

$$= \frac{1}{\binom{n}{2}} \sum_{i<j} \frac{1}{2} \left(X_i - X_j \right)^2$$

- $\overline{X}_n$ is a U-statistic of order 1 with kernel $h(x) = x$

- The U-statistic with kernel $h(x, y) = |x - y|$ estimates the mean pairwise deviation or Gini mean difference. [The Gini coefficient, $G = \mathrm{E}|X - Y|/(2\mathrm{E}X)$, is commonly used as a measure of income inequality.]

- Third k-statistic, $k_3 = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(X_i - \overline{X}_n \right)^3$ is a U-statistic that estimates the 3rd cumulant $\kappa_3 = K^{(3)}(0)$, where $K(t) = \log \mathrm{E} \left[e^{tX} \right]$.

U-statistics: Examples

- The U-statistic with kernel $h(x,y) = (x-y)(x-y)^T$ estimates the variance-covariance matrix.

- **Kendall's τ:** For a random pair $P_1 = (X_1, Y_1)$, $P_2 = (X_2, Y_2)$ of points in the plane,

$$\tau = Pr\,(P_1 P_2 \text{ has positive slope}) \; - Pr\,(P_1 P_2 \text{ has negative slope })$$
$$= E1[(X_1 - X_2)(Y_1 - Y_2) > 0] - E1[(X_1 - X_2)(Y_1 - Y_2) < 0]$$
$$= 4P(X_1 < X_2, Y_1 < Y_2) - 1$$

where $P_1 P_2$ is the line from P_1 to P_2 . It is a measure of correlation: $\tau \in [-1, 1], \tau = 0$ for independent $X, Y, \tau = \pm 1$ for $Y = f(X)$ for monotone f. Clearly, τ can be estimated using a U-statistic of order 2.

U-statistics: Examples

■ The Wilcoxon one-sample rank statistic:

$$T^+ = \sum_{i=1}^{n} R_i 1\left[X_i > 0\right]$$

where R_i is the rank (position when $|X_1|, \ldots, |X_n|$ are arranged in ascending order). It's used to test if the distribution is symmetric about zero. Assuming the $|X_i|$ are all distinct, then we can write

$$R_i = \sum_{j=1}^{n} 1\left[|X_j| \le |X_i|\right]$$

Hence

$$T^+ = \sum_{i=1}^{n} \sum_{j=1}^{n} 1\left[|X_j| \le X_i\right]$$

$$= \sum_{i<j} 1\left[|X_j| < X_i\right] + \sum_{i<j} 1\left[|X_i| < X_j\right] + \sum_i 1\left[X_i > 0\right]$$

$$= \sum_{i<j} 1\left[X_i + X_j > 0\right] + \sum_i 1\left[X_i > 0\right]$$

$$= \frac{1}{\binom{n}{2}} \sum_{i<j} \binom{n}{2} 1\left[X_i + X_j > 0\right] + \frac{1}{n} \sum_i n 1\left[X_i > 0\right]$$

$$= \frac{1}{\binom{n}{2}} \sum_{i<j} h_2\left(X_i, X_j\right) + \frac{1}{n} \sum_i h_1\left(X_i\right)$$

where

$$h_2\left(X_i, X_j\right) = \binom{n}{2} 1\left[X_i + X_j > 0\right]$$

$$h_1\left(X_i\right) = n 1[X_i > 0]$$

So it's a sum of U-statistics. [Why is it not a U-statistic?]

1.2 Properties of U-statistics

- U for unbiased: U is an unbiased estimator for $\mathrm{E}h(X_1, \ldots, X_r)$,
  $\mathrm{E}U = \mathrm{E}h(X_1, \ldots, X_r)$

- U is a lower variance estimate than $h(X_1, \ldots, X_r)$, because U is an average over permutations. Indeed, since U is an average over permutations π of $h(X_{\pi(1)}, \ldots, X_{\pi(r)})$, we can write

$$U(X_1, \ldots, X_n) = \mathbf{E}\left[h(X_1, \ldots, X_r) | X_{(1)}, \ldots, X_{(n)}\right]$$

where $\left(X_{(1)}, \ldots, X_{(n)}\right)$ is the data in some sorted order. Thus, for $\mathrm{E}U = \theta$, we can write the variance as:

$$\mathbf{E}(U - \theta)^2 = \mathbf{E}\left(\mathbf{E}\left[h(X_1, \ldots, X_r) - \theta | X_{(1)}, \ldots, X_{(n)}\right]\right)^2$$
$$\leq \mathbf{E}\mathbf{E}\left[(h(X_1, \ldots, X_r) - \theta)^2 | X_{(1)}, \ldots, X_{(n)}\right]$$

$$= \mathbf{E}\left(h\left(X_1, \ldots, X_r\right) - \theta\right)^2$$

by Jensen's inequality (for a convex function ϕ, we have

$$\phi(\mathbf{E}X) \leq \mathbf{E}\phi(X))$$

This is the Rao-Blackwell theorem: the mean squared error of the estimator $h\left(X_1, \ldots, X_r\right)$ is reduced by replacing it by its conditional expectation, given the sufficient statistic $\left(X_{(1)}, \ldots, X_{(n)}\right)$

1.2.1 Variance of U-statistics

Now we'll compute the asymptotic variance of a U-statistic. Recall the definition:

$$U = \frac{1}{\binom{n}{r}} \sum_{i_1 < \cdots < i_r} h\left(X_{i_1}, \ldots, X_{i_r}\right)$$

So [letting S, S' range over subsets of $\{1, \ldots, n\}$ of size r]:

$$\text{Var}(U) = \frac{1}{\binom{n}{r}^2} \sum_S \sum_{S'} \text{Cov}\left(h\left(X_S\right), h\left(X_{S'}\right)\right)$$

$$= \frac{1}{\binom{n}{r}^2} \sum_{c=0}^r \binom{n}{r}\binom{r}{c}\binom{n-r}{r-c} \zeta_c$$

where $\binom{n}{r}\binom{r}{c}\binom{n-r}{r-c}$ is the number of ways of choosing S and S' with an intersection of size c (first choose S, then choose the intersection from S, then choose the non-intersection for the rest of S').

Also, $\zeta_c = \text{Cov}\left(h\left(X_S\right), h\left(X_{S'}\right)\right)$ depends only on $c = |S \cap S'|$. To see this, suppose that $S \cap S' = I$ with $|I| = c$,

$$
\begin{aligned}
\zeta_c &= \text{Cov}\left(h\left(X_S\right), h\left(X_{S'}\right)\right) \\
&= \text{Cov}\left(h\left(X_I, X_{S-I}\right), h\left(X_I, X_{S'-I}\right)\right) \\
&= \text{Cov}\left(h\left(X_1^c, X_{c+1}^r\right), h\left(X_1^c, X_{r+1}^{2r-c}\right)\right) \\
&= \text{Cov}\left(\text{E}\left[h\left(X_1^c, X_{c+1}^r\right) | X_1^c\right], \text{E}\left[h\left(X_1^c, X_{r+1}^{2r-c}\right) | X_1^c\right]\right) \\
&\quad + \text{ECov}\left[h\left(X_1^c, X_{c+1}^r\right), h\left(X_1^c, X_{r+1}^{2r-c}\right) | X_1^c\right] \\
&= \text{Var}\left(\text{E}\left[h\left(X_1^c, X_{c+1}^r\right) | X_1^c\right]\right)
\end{aligned}
$$

where $X_1^c = (X_1, \ldots, X_c)$. Clearly, $\zeta_0 = 0$.

Now,

$$
\text{Var}(U) = \frac{1}{\binom{n}{r}^2} \sum_{c=1}^r \binom{n}{r}\binom{r}{c}\binom{n-r}{r-c} \zeta_c
$$

$$= \frac{1}{\binom{n}{r}} \sum_{c=1}^{r} \binom{r}{c} \binom{n-r}{r-c} \zeta_c$$

$$= \theta\left(n^{-r}\right) \sum_{c=1}^{r} \theta\left(n^{r-c}\right) \zeta_c$$

$$= \sum_{c=1}^{r} \theta\left(n^{-c}\right) \zeta_c$$

So if $\zeta_1 \neq 0$, the first term dominates:

$$n \operatorname{Var}(U) \to \frac{nr!(n-r)!r(n-r)!}{n!(r-1)!(n-2r+1)!}\zeta_1 \to r^2\zeta_1$$

If $r^2\zeta_1 = 0$, we say that U is degenerate.

Variance of U-statistics: Example

■ Estimator of variance: $h(X_1, X_2) = (1/2)(X_1 - X_2)^2$

$$\zeta_1 = \text{Cov}(h(X_1, X_2), h(X_1, X_3))$$

$$= \text{Var}(\mathbf{E}[h(X_1, X_2)|X_1]) + \mathbf{E}[\text{Cov}(h(X_1, X_2), h(X_1, X_3)|X_1]$$

$$= \text{Var}(\mathbf{E}[h(X_1, X_2)|X_1]) = \text{Var}\left(\mathbf{E}\left[\frac{1}{2}(X_1 - X_2)^2|X_1\right]\right)$$

$$= \text{Var}\left(\mathbf{E}\left[\frac{1}{2}(X_1 - \mu + \mu - X_2)^2|X_1\right]\right)$$

$$= \text{Var}\left(\frac{1}{2}\left((X_1 - \mu)^2 + \sigma^2\right)\right) = \frac{1}{4}\left(\mu_4 - \sigma^4\right)$$

where $\mu_4 = \mathbf{E}\left((X_1 - \mu)^4\right)$ is the 4th central moment. So $n\,\text{Var}(U) \to \mu_4 - \sigma^4$. We'll see that $\sqrt{n}\left(U - \sigma^2\right) \rightsquigarrow N\left(0, \mu_4 - \sigma^4\right)$. (What if $\mu_4 - \sigma^4 = 0$?)

Variance of U-statistics: Example

Recall Kendall's τ: For a random pair $P_1 = (X_1, Y_1)$, $P_2 = (X_2, Y_2)$ of points in the plane, if X, Y are independent and continuous [and $P_1 P_2$ is the line from P_1 to P_2]

$$h(P_1, P_2) = (1 [P_1 P_2 \text{ has positive slope }] - 1 [P_1 P_2 \text{ has negative slope }])$$

$$\zeta_1 = \text{Cov}(h(P_1, P_2), h(P_1, P_3))$$

$$\ldots = 1/9$$

so $n \text{Var}(U) \to 4/9$. We'll see that $\sqrt{n} U \rightsquigarrow N(0, 4/9)$. And this gives a test for independence.

1.2.2   Asymptotic distribution of U-statistics

How do we find the asymptotic distribution of a U-statistic? We'll appeal to this theorem:

Theorem:

$$X_n \rightsquigarrow X \text{ and } d(X_n, Y_n) \xrightarrow{P} 0 \implies Y_n \rightsquigarrow X$$

In particular, we find another sequence $\hat{U}$ such that

- $\sqrt{n}(U - \theta - \hat{U}) \xrightarrow{P} 0$, and

- The asymptotics of $\hat{U}$ are easy to understand.

In this case, we find $\hat{U}$ of the form $\hat{U} = \sum_i f(X_i)$. Then the CLT gives the result.

Asymptotic distribution of U-statistics

1. **Why do functions of a single variable suffice?**

   Because the interactions are weak

2. **How do we find suitable functions?**

   By projecting: finding the element of the linear space of functions of single variables that captures most of the variance of U.

This leads us to the idea of **Hájek projections**.

Projection Theorem

Consider a random variable T and a linear space $\mathcal{S}$ of random variables, with $ES^2 < \infty$ for all $S \in \mathcal{S}$ and $ET^2 < \infty$. [Write $T \in L_2(P)$, $\mathcal{S} \subset L_2(P)$, the Hilbert space of finite variance random variables defined on a probability space.] A projection $\hat{S}$ of T on $\mathcal{S}$ is a minimizer over $\mathcal{S}$ of $\mathbf{E}(T - S)^2$.]

> **Theorem**: $\hat{S}$ is a projection of T on $\mathcal{S}$ iff $\hat{S} \in \mathcal{S}$ and, for all $S \in \mathcal{S}$, the error $T - \hat{S}$ is orthogonal to $\mathcal{S}$, that is,
>
> $$\mathbf{E}(T - \hat{S})S = 0$$
>
> If $\hat{S}_1$ and $\hat{S}_2$ are projections of T onto $\mathcal{S}$, then $\hat{S}_1 = \hat{S}_2$ a.s.

Projection Theorem

Notice that if $\mathcal{S}$ contains constants, then $S = 1 \in \mathcal{S}$ shows that

$$\mathbf{E}(T - \hat{S}) = 0, \quad \text{i.e., } \mathbf{E}T = \mathbf{E}\hat{S}$$

Also, for all $S \in \mathcal{S}, S - \mathbf{E}S \in \mathcal{S}$, so

$$\text{Cov}(T - \hat{S}, S) = \mathbf{E}((T - \hat{S})(S - \mathbf{E}S)) = 0$$

Projection Theorem Proof

Theorem: 1. $\hat{S} \in \mathcal{S}$ is a projection of T on $\mathcal{S}$ (minimizes $\mathbf{E}(T-S)^2$ iff , for all $S \in \mathcal{S}, \mathbf{E}(T-\hat{S})S = 0$

2. If $\hat{S}_1$ and $\hat{S}_2$ are projections of T onto $\mathcal{S}$, then $\hat{S}_1 = \hat{S}_2$ as.

We can write the criterion, for any $S \in \mathcal{S}$ as

$$\mathbf{E}\left(T-S\right)^2 = \mathbf{E}(T-\hat{S}+\hat{S}-S)^2$$
$$= \mathbf{E}(T-\hat{S})^2 + 2\mathbf{E}((T-\hat{S})(\hat{S}-S)) + (\hat{S}-S)^2$$

If $\mathrm{E}(T-\hat{S})S = 0$, then this is $\mathrm{E}(T-\hat{S})^2 + (\hat{S}-S)^2$, which is minimized for $S = \hat{S}$, and strictly minimized unless $\mathrm{E}(\hat{S}-S)^2 = 0$, so $\hat{S}$ is unique.

Projection Theorem Proof

If $\hat{S}$ is a projection, then

$$\mathbf{E}(T - \hat{S} - \alpha S)^2 = \mathbf{E}(T - \hat{S})^2 - 2\alpha \mathbf{E}(T - \hat{S})S + \alpha^2 \mathbf{E}S^2$$

is at least $\mathbf{E}(T - \hat{S})^2$ for any $S \in \mathcal{S}$ and any α. And this implies that

$$\mathbf{E}(T - \hat{S})S = 0$$

Projection Theorem

- Pythagoras theorem: $\mathbf{E}(T)^2 = \mathbf{E}(T - \hat{S} + \hat{S})^2 = \mathbf{E}(T - \hat{S})^2 + \mathbf{E}(\hat{S})^2$

- If $\mathcal{S}$ contains constants, $\mathbf{E}(T) = \mathbf{E}(\hat{S})$ and $\mathrm{Var}(T) = \mathrm{Var}(T - \hat{S}) + \mathrm{Var}(\hat{S})$

- So if $\mathcal{S}$ contains constants and $\hat{S}$ and T have the same variance, then $\hat{S} = T$ a.s.

- A similar property holds asymptotically...

Projections and Asymptotics

Consider $\mathcal{S}_n$ a sequence of linear spaces of random variables that contain the constants and that have finite second moments.

> **Theorem**: For T_n with projections $\hat{S}_n$ on $\mathcal{S}_n$,
>
> $$\frac{\text{Var}(T_n)}{\text{Var}(\hat{S}_n)} \to 1 \quad \implies \quad \frac{T_n - \text{E}T_n}{\sqrt{\text{Var}(T_n)}} - \frac{\hat{S}_n - \text{E}\hat{S}_n}{\sqrt{\text{Var}(\hat{S}_n)}} \xrightarrow{P} 0$$

Projections and Asymptotics: Proof

Define

$$Z_n = \frac{T_n - \mathbf{E}T_n}{\sqrt{\mathrm{Var}(T_n)}} - \frac{\hat{S}_n - \mathbf{E}\hat{S}_n}{\sqrt{\mathrm{Var}(\hat{S}_n)}}$$

Clearly, $\mathbf{E}Z_n = 0$, and

$$\mathrm{Var}\left(Z_n\right) = 2 - 2\frac{\mathrm{Cov}\left(T_n, \hat{S}_n\right)}{\sqrt{\mathrm{Var}\left(T_n\right)}\sqrt{\mathrm{Var}\left(\hat{S}_n\right)}}$$

$$= 2 - 2\frac{\sqrt{\mathrm{Var}(\hat{S}_n)}}{\sqrt{\mathrm{Var}\left(T_n\right)}} \to 0$$

where the second equality is because $\mathcal{S}$ contains constants, so $\mathrm{Cov}\left(T_n - \hat{S}_n, \hat{S}_n\right) = 0$, hence $\mathrm{Cov}\left(T_n, \hat{S}_n\right) = \mathrm{Var}\left(\hat{S}_n\right)$.

Linear Spaces

■ What linear spaces should we project onto? We need a rich space, since we have to lose nothing asymptotically when we project.

■ We'll consider **the space of functions of a single random variable**. Then projection corresponds to computing conditional expectations.

Just as $\mathrm{E}X = \arg\min_{a \in \mathbb{R}} \mathrm{E}(X-a)^2$

$$\mathrm{E}[X|Y] = \arg\min_{g:\mathbb{R}\to\mathbb{R}} \mathrm{E}(X-g(Y))^2$$

This is the projection of X onto the linear space $\mathcal{S}$ of measurable functions of Y.

Conditional Expectations as Projections

The projection theorem says: for all measurable g,

$$\mathbf{E}(X - \mathbf{E}[X|Y])g(Y) = 0$$

Properties of $\mathrm{E}[X|Y]$:

- $\mathrm{E}X = \mathrm{E}\mathrm{E}[X|Y]$ (consider $g = 1$)

- For a joint density $f(x, y)$

$$\mathbf{E}[X|Y] = \int x \frac{f(x, Y)}{f(Y)} dx$$

- For independent X, Y, $\mathbf{E}(X - \mathrm{E}X)g(Y) = 0$, so $\mathrm{E}[X|Y] = \mathrm{E}X$

Conditional Expectations as Projections

Properties of $E[X|Y]$:

- $E[f(Y)X|Y] = f(Y)E[X|Y]$ (Because $\mathrm{E}[f(Y)X - f(Y)\mathrm{E}[X|Y]g(Y) = \mathrm{E}[X - \mathrm{E}[X|Y]f(Y)g(Y)] = 0.$)

- $\mathrm{E}[\mathrm{E}[X|Y,Z]|Y] = \mathrm{E}[X|Y]$ (Because $\mathrm{E}(\mathrm{E}[X|Y,Z] - \mathrm{E}[X|Y])g(Y) = \mathrm{E}(\mathrm{E}[g(Y)X|Y,Z] - \mathrm{E}[g(Y)X|Y]) = 0.$)

Projection on Sums

Definition: For independent random vectors $X_1, \ldots, X_n$, the Hájek projection of a random variable is its projection onto the set of sums

$$\sum_{i=1}^{n} g_i(X_i)$$

$\sum_{i=1}^{n} g_i(X_i)$ of measurable functions satisfying $\mathrm{E} g_i(X_i)^2 < \infty$

Hájek Projections

Theorem: [Hájek projection principle:] The Hájek projection of $T \in L_2(P)$ is

$$\hat{S} = \sum_{i=1}^{n} \mathrm{E}\left[T | X_i\right] - (n-1)\mathbf{E}T$$

Hájek Projections Principle: Proof

From the projection theorem, we need to check that $T - \hat{S}$ is orthogonal to each $g_i(X_i)$. It suffices if $\mathbf{E}[T|X_i] = \mathbf{E}\left[\hat{S}|X_i\right]$:

$$\mathrm{E}(T - \hat{S})g_i(X_i) = \mathrm{E}\left(\mathrm{E}\left[T - \hat{S}|X_i\right] g_i(X_i)\right)$$

But

$$\begin{aligned}
\mathbf{E}\left[\hat{S}|X_i\right] &= \mathbf{E}\left[\sum_{j=1}^n \mathbf{E}[T|X_j] - (n-1)\mathbf{E}T|X_i\right] \\
&= \mathbf{E}[T|X_i] + \sum_{j \neq i} \mathbf{E}[\mathbf{E}[T|X_j]|X_i] - (n-1)\mathbf{E}T \\
&= \mathbf{E}[T|X_i]
\end{aligned}$$

because the X_i are independent, so $T - \hat{S}$ is orthogonal to $\mathcal{S}$.

Asymptotic Normality of U-Statistics

Theorem: If $Eh^2 < \infty$, define $\hat{U}$ as the Hájek projection of $U - \theta$. Then

$$\hat{U} = \frac{r}{n} \sum_{i=1}^{n} h_1 (X_i),$$

with

$$h_1(x) = Eh\ (x, X_2, \ldots, X_r) - \theta$$
$$\sqrt{n}(U - \theta - \hat{U}) \xrightarrow{P} 0,$$
$$\sqrt{n}(U - \theta) \rightsquigarrow N\left(0, r^2\zeta_1\right), \quad \text{where}$$
$$\zeta_1 = Eh_1^2 (X_1)$$

$$\boxed{\textbf{Proof}}$$

Recall:
$$U = \frac{1}{\binom{n}{r}} \sum_{j \subseteq [n]} h\left(X_{j_1}, \ldots, X_{j_r}\right)$$

By the Hájek projection principle, the projection of $U - \theta$ is

$$\hat{U} = \sum_{i=1}^{n} \mathrm{E}\left[U - \theta | X_i\right]$$

$$= \sum_{i=1}^{n} \frac{1}{\binom{n}{r}} \sum_{j \subseteq [n]} \mathbf{E}\left[h\left(X_{j_1}, \ldots, X_{j_r}\right) - \theta | X_i\right]$$

But

$$\mathbf{E}\left[h\left(X_{j_1}, \ldots, X_{j_r}\right) - \theta | X_i\right] = \begin{cases} h_1\left(X_i\right) - \theta & \text{if } i \in j = (j_1, \ldots, j_r) \\ 0 & \text{otherwise} \end{cases}$$

where $h_1(x_1) = Eh(x_1, X_2, \ldots, X_r)$.

For each X_i, there are $\begin{pmatrix} n-1 \\ r-1 \end{pmatrix}$ of the $\begin{pmatrix} n \\ r \end{pmatrix}$ subsets that contain i. Thus,

$$\hat{U} = \sum_{i=1}^{n} \frac{r!(n-r)!(n-1)!}{n!(r-1)!(n-r)!}(h_1(X_i) - \theta) = \frac{r}{n}\sum_{i=1}^{n}(h_1(X_i) - \theta)$$

To see that $\hat{U}$ has the same asymptotics as U, notice that $E\hat{U} = 0$ and so its variance is asymptotically the same as that of U:

$$\operatorname{var}\hat{U} = \frac{r^2}{n}\operatorname{E}(h_1(X_1) - \theta)^2 = \frac{r^2}{n}\operatorname{E}(\operatorname{E}[h(X_1^r)|X_1] - \theta)^2$$

$= \frac{r^2}{n}\operatorname{Var}(\operatorname{E}[h(X_1,\dots,X_r) - \theta|X_1]) = \frac{r^2}{n}\zeta_1$

CLT (and finiteness of $\operatorname{Var}(\hat{U})$) implies $\sqrt{n}\hat{U} \rightsquigarrow N(0, r^2\zeta_1)$

Also [recall that $n\operatorname{Var}U \to r^2\zeta_1$, $\operatorname{Var}\hat{U}/\operatorname{Var}U \to 1$, so

$$\frac{U-\theta}{\sqrt{\operatorname{Var}(U)}} - \frac{\hat{U}}{\sqrt{\operatorname{Var}(\hat{U})}} \xrightarrow{P} 0$$

which implies $\sqrt{n}(U - \theta - \hat{U}) \xrightarrow{P} 0$, and hence

$$\sqrt{n}(U - \theta) \rightsquigarrow N\left(0, r^2\zeta_1\right)$$

Examples

Estimator of variance: $h(X_1, X_2) = (1/2)(X_1 - X_2)^2$:

$$\zeta_1 = \frac{1}{4}\left(\mu_4 - \sigma^4\right)$$

where $\mu_4 = \mathrm{E}\left((X_1 - \mu)^4\right)$ is the 4 th central moment. So $n\,\mathrm{Var}(U) \to \mu_4 - \sigma^4$, hence $\sqrt{n}\left(U - \sigma^2\right) \rightsquigarrow N\left(0, \mu_4 - \sigma^4\right)$

Examples

Recall Kendall's τ : For a random pair $P_1 = (X_1, Y_1)$, $P_2 = (X_2, Y_2)$ of points in the plane, if X, Y are independent and continuous [recall: $P_1 P_2$ is the line from P_1 to P_2]

$$h(P_1, P_2) = (1\left[P_1 P_2 \text{ has positive slope }\right] - 1\left[P_1 P_2 \text{ has negative slope }\right])$$

$$\mathrm{E}\tau = 0$$

$$\zeta_1 = \mathrm{Cov}\left(h(P_1, P_2), h(P_1, P_3)\right) = \frac{1}{9}$$

Thus $\sqrt{n}U \rightsquigarrow N(0, 4/9)$. And this gives a test for independence of X and Y :

$$\mathrm{Pr}\left(\sqrt{9n/4}|\tau| > z_{\alpha/2}\right) \to \alpha$$

Recall Wilcoxon's one sample rank statistic:

$$T^+ = \sum_{i=1}^n R_i 1\left[X_i > 0\right]$$
$$= \frac{1}{\binom{n}{2}} \sum_{i<j} h_2\left(X_i, X_j\right) + \frac{1}{n} \sum_i h_1\left(X_i\right)$$

$$h_2\left(X_i, X_j\right) = \binom{n}{2} 1\left[X_i + X_j > 0\right]$$

$$h_1\left(X_i\right) = n1\left[X_i > 0\right]$$

where R_i is the rank (position when $|X_1|, \ldots, |X_n|$ are arranged in ascending order). It's used to test if the distribution is symmetric about zero.

It's a sum of U-statistics. The first sum dominates the asymptotics. So consider

$$U = \frac{1}{\binom{n}{2}} \sum_{i<j} \binom{n}{2} 1\left[X_i + X_j > 0\right]$$

The Hájek projection of $U - \theta$ is

$$\hat{U} = \frac{2}{n} \sum_{i=1}^{n} h_1\left(X_i\right)$$

and

$$\begin{aligned}
h_1(x) &= \mathrm{E}h\left(x, X_2\right) - \mathrm{E}h\left(X_1, X_2\right) \\
&= \binom{n}{2} \left(P\left(x + X_2 > 0\right) - P\left(X_1 + X_2 > 0\right)\right) \\
&= -\binom{n}{2} \left(F(-x) - \mathrm{E}F\left(-X_1\right)\right)
\end{aligned}$$

For F symmetric about $0, (F(x) = 1 - F(-x))$, we have

$$\hat{U} = -\frac{2\binom{n}{2}}{n} \sum_{i=1}^{n} \left(F(-X_i) - \mathrm{E}F(-X_i) \right)$$

$$= \frac{2\binom{n}{2}}{n} \sum_{i=1}^{n} \left(F(X_i) - \mathrm{E}F(X_i) \right)$$

But $F(X_i)$ is always uniform on $[0, 1]$, and so $\mathrm{E}F(X_i) = 1/2$ and $\mathrm{Var}\, F(X_i) = 1/12$. Thus,

$$\mathrm{Var}(\hat{U}) = \frac{4\binom{n}{2}^2}{n}; \quad \mathrm{Var}(F(X_i)) = \frac{n(n-1)^2}{12}$$

Thus, for symmetric distributions,

$$n^{-3/2} \left(T^+ - \frac{\binom{n}{2}}{2} \right) \rightsquigarrow N(0, 1/12)$$

So we have a test for symmetry:

$$Pr\left(\sqrt{12}n^{-3/2}\left|T^+ - \frac{\binom{n}{2}}{2}\right| > z_{\alpha/2}\right) \to \alpha$$