Reading:

Outline of the Lecture 7:

- U-Statistics
  - Variance computations
  - Projections of random variables and vectors
  - Asymptotic normality of U-statistics

1 Variance of U-Statistics

Recall these definitions that we set up last lecture:

**Definition 1.1.** Given a symmetric kernel function $h : X^r \to \mathbb{R}$, define the associated **U-statistic** as

$$ U_n := \frac{1}{\binom{n}{r}} \sum_{\beta \subseteq [n], |\beta| = r} h(X_\beta). $$

**Definition 1.2.** For each $c \in \{0, \ldots, r\}$, define

$$ h_c(x_1, \ldots, x_c) := \mathbb{E}[h(x_1, \ldots, x_c, X_{c+1}, \ldots, X_r)]. $$

Define $\hat{h}_c$ to be the centered version of $h_c$, i.e.

$$ \hat{h}_c := h_c - \mathbb{E}[h_c] = h_c - \theta, $$

where $\theta = \mathbb{E}U_n$.

**Definition 1.3.** For each $c \in \{0, \ldots, r\}$, define

$$ \zeta_c := \text{Var}[h_c(X_1, \ldots, X_c)] = \mathbb{E}[h_c(X_1, \ldots, X_c)^2]. $$

(Note that $\zeta_0 = 0$.)

**Goal:** Write $\text{Var}U_n$ as a sum of the $\zeta_c$’s.

**Lemma 1.** If $\alpha, \beta \subseteq [n], S = \alpha \cap \beta$, $c = |S|$, then

$$ \mathbb{E}\left[\hat{h}(X_\alpha)\hat{h}(X_\beta)\right] = \zeta_c. $$
Proof Using the symmetry of $h$,

\[
\mathbb{E} \left[ \hat{h}(X_\alpha) \hat{h}(X_\beta) \right] = \mathbb{E} \left[ \hat{h}(X_\alpha \setminus S, X_S) \hat{h}(X_\beta \setminus S, X_S) \right] \\
= \mathbb{E} \left[ \mathbb{E}[\hat{h}(X_\alpha \setminus S, X_S) | X_S] \cdot \mathbb{E}[\hat{h}(X_\beta \setminus S, X_S) | X_S] \right] \quad \text{(since $X_\alpha \setminus S, X_\beta \setminus S$ indep.)} \\
= \mathbb{E} \left[ \hat{h}_c(X_S) \cdot \hat{h}_c(X_S) \right] \\
= \zeta_c.
\]

\[\square\]

**Theorem 2.** Let $U_n$ be an $r^{th}$ order U-statistic. Then

\[
\text{Var}U_n = \frac{r^2}{n} \zeta_1 + O(n^{-2}).
\]

Proof There are $\binom{n}{r} \binom{n-r}{r-c}$ ways to select a pair of subsets of $[n]$, each of size $r$, with $c$ common elements. Hence,

\[
U_n - \theta = \left( \binom{n}{r} \right)^{-1} \sum_{|\beta|=r} \hat{h}(X_\beta),
\]

\[
\text{Var}U_n = \left( \binom{n}{r} \right)^{-2} \sum_{|\alpha|=r} \sum_{|\beta|=r} \mathbb{E} \left[ \hat{h}(X_\alpha) \hat{h}(X_\beta) \right] \\
= \left( \binom{n}{r} \right)^{-2} \sum_{c=1}^{r} \binom{n}{r} \binom{r}{c} \binom{n-r}{r-c} \zeta_c \\
= \sum_{c=1}^{r} \frac{r!^2}{c!(r-c)!^2} \frac{(n-r)(n-r-1) \ldots (n-2r+c+1)}{n(n-1) \ldots (n-r+1)} \zeta_c.
\]

For fixed $c$, $\frac{(n-r)(n-r-1) \ldots (n-2r+c+1)}{n(n-1) \ldots (n-r+1)}$ has $r - c$ terms in the numerator and $r$ terms in the denominator. Hence,

\[
\text{Var}U_n = r^2 \left( \frac{n-r)(n-r-1) \ldots (n-2r+2)}{n(n-1) \ldots (n-r+1)} \zeta_1 + \sum_{c=2}^{r} O \left( \frac{n^{r-c}}{n^r} \right) \zeta_c
\]

\[
= r^2 \left[ \frac{1}{n} + O(n^{-2}) \right] \zeta_1 + O(n^{-2})
\]

\[
= \frac{r^2}{n} \zeta_1 + O(n^{-2}).
\]

\[\square\]

With this theorem, we know that the variance of U-statistics behaves like the variance of a sample mean plus high-order errors.
New Goal: Show that \( U_n \) is asymptotically normal by projecting out all high-order interactions. To do this, we need some theory on projections.

2 Projections

Let \( V \) be a Hilbert space, i.e. there is an inner product \( \langle \cdot, \cdot \rangle \) on \( V \), an associated norm \( \|v\|_2^2 = \langle v, v \rangle \), and \( V \) is complete w.r.t. this norm. (Note that \( \|v\|_2 = 0 \) iff \( v = 0 \).)

Definition 2.1. Let \( C \subseteq V \) be a convex and closed set. Define the projection of \( w \) onto \( C \) as

\[
\pi_C(w) := \text{argmin}_{v \in C} \{\|w - v\|_2^2\}.
\]

Theorem 3. \( \pi_C(w) \) exists, is unique, and is characterized by the inequality

\[
\langle w - \pi_C(w), v - \pi_C(w) \rangle \leq 0.
\]

Loosely speaking, the inequality means that the “angle” between \( w - \pi_C(w) \) and \( v - \pi_C(w) \) is obtuse.

Corollary 4. Suppose \( C \) is a linear subspace of \( V \). Then \( \pi_C(w) \) is the projection of \( w \) onto \( C \) iff for all \( v \in C \),

\[
\langle w - \pi_C(w), v \rangle = 0.
\]

Proof. If \( C \) is linear, then \( v \in C \iff -v \in C \). Hence, by Equation 1,

\[
\langle w - \pi_C(w), v \rangle \leq \langle w - \pi_C(w), \pi_C(w) \rangle \quad \text{and} \quad \langle w - \pi_C(w), v \rangle \leq -\langle w - \pi_C(w), \pi_C(w) \rangle,
\]

\[
\Rightarrow \langle w - \pi_C(w), v \rangle = 0.
\]

Let’s now put these ideas in the random variable setting.

Fact 5. Random variables with 2 moments form a Hilbert space with inner product \( \langle X, Y \rangle = \mathbb{E}[XY] \). We will call this space \( L_2(P) \).

Corollary 6. If \( S \) is a linear subspace of \( L_2(P) \), then \( \hat S \in S \) is the projection of \( T \in L_2(P) \) onto \( S \) iff for all \( S \in S \),

\[
\mathbb{E}[(T - \hat S)S] = 0.
\]

If this is the case, then

\[
\mathbb{E}[T^2] = \mathbb{E}[(T - \hat S)^2] + \mathbb{E}[\hat S^2].
\]

Proof. The characterization of \( \hat S \) follows directly from Corollary 4.

\[
\mathbb{E}[T^2] = \mathbb{E}[(T - \hat S + \hat S)^2]
\]

\[
= \mathbb{E}[(T - \hat S)^2] + \mathbb{E}[\hat S^2] + 2 \text{Cov}(T - \hat S, S)
\]

\[
= \mathbb{E}[(T - \hat S)^2] + \mathbb{E}[\hat S^2].
\]
Idea: Try to understand when $T_n$ and its projections have the same asymptotic behavior.

**Theorem 7.** Let $T_n$ be statistics, and let $\hat{S}_n$ be the projections of $T_n$ onto subspaces $S_n$ which contain constant random variables.

If $\frac{\text{Var} T_n}{\text{Var} \hat{S}_n} \to 1$, then $\frac{T_n - E T_n}{\sqrt{\text{Var} T_n}} - \frac{\hat{S}_n - E \hat{S}_n}{\sqrt{\text{Var} \hat{S}_n}} \to 0$.

**Proof** Let $A_n = \frac{T_n - E T_n}{\sqrt{\text{Var} T_n}} - \frac{\hat{S}_n - E \hat{S}_n}{\sqrt{\text{Var} \hat{S}_n}}$. Note that $E A_n = 0$. Thus, if we can show that $\text{Var} A_n \to 0$, we are done.

Note that

\[
\text{Cov}(T_n, \hat{S}_n) = E[T_n \hat{S}_n] - E[T_n]E[\hat{S}_n] = E[(T_n - \hat{S}_n + \hat{S}_n) \hat{S}_n] - E[\hat{S}_n]^2 = E[\hat{S}_n^2] - E[\hat{S}_n]^2 = \text{Var} \hat{S}_n.
\]

Hence,

\[
\text{Var} A_n = \text{Var} \frac{T_n - E T_n}{\sqrt{\text{Var} T_n}} + \text{Var} \frac{\hat{S}_n - E \hat{S}_n}{\sqrt{\text{Var} \hat{S}_n}} - \frac{2 \text{Cov}(T_n, \hat{S}_n)}{\sqrt{\text{Var} T_n \text{Var} \hat{S}_n}}
\]

\[
= 2 - 2 \sqrt{\frac{\text{Var} \hat{S}_n}{\text{Var} T_n}} \to 0.
\]

\[\square\]

### 2.1 Conditional Expectations

Conditional expectations are simply projections.

**Definition 2.2.** If $X \in L_2(P)$, $Y$ is a random variable, $S = \{\text{all measurable functions } g(Y) \text{ with } E[g^2(Y)] < \infty\}$, we define the **conditional expectation of $X$ given $Y$**, $E[X \mid Y]$, as the projection of $X$ onto $S$, i.e.

\[
E[(X - E[X \mid Y]) g(Y)] = 0
\]

for all $g \in S$.

By choosing $g$ appropriately, some nice properties of conditional expectation are immediate:

- $E[X - E[X \mid Y]] = 0$, and
- $E[f(Y)X \mid Y] = f(Y)E[X \mid Y]$.
2.2 Hájek Projections

Idea: Apply these ideas to U-statistics, i.e. project them onto spaces of the form \( \sum_{i=1}^{n} g_i(X_i) \).

Lemma 8 (11.10 in VDV). Let \( X_1, \ldots, X_n \) be independent. Let \( S = \left\{ \sum_{i=1}^{n} g_i(X_i) : g_i \in L_2(P) \right\} \).

If \( \mathbb{E}T^2 < \infty \), then the projection \( \hat{S} \) of \( T \) onto \( S \) is given by

\[
\hat{S} = \sum_{i=1}^{n} \mathbb{E}[T \mid X_i] - (n-1)\mathbb{E}T.
\]  

(2)

Proof Note that

\[
\mathbb{E}[\mathbb{E}[T \mid X_i] \mid X_j] = \begin{cases} \mathbb{E}[T \mid X_i] & \text{if } i = j, \\ \mathbb{E}T & \text{if } i \neq j. \end{cases}
\]

If \( \hat{S} \) is as stated in Equation 2, then

\[
\mathbb{E}[\hat{S} \mid X_j] = (n-1)\mathbb{E}T + \mathbb{E}[T \mid X_j] - (n-1)\mathbb{E}T = \mathbb{E}[T \mid X_j],
\]

\[
\mathbb{E}[(T - \hat{S})g_j(X_j)] = \mathbb{E}[^{\mathbb{E}[T - \hat{S} \mid X_j]}g_j(X_j)]
\]

\[
= 0,
\]

\[
\mathbb{E} \left[ (T - \hat{S}) \sum_{j=1}^{n} g_j(X_j) \right] = 0.
\]

Thus, \( \hat{S} \) must be the projection of \( T \) onto \( S \).

Next move: Project the U-statistic \( U_n \) onto the space \( \left\{ \sum_{i=1}^{n} g_i(X_i) : g_i \in L_2(P) \right\} \). We will show that \( \text{Var}U_n = \text{Var}U_n + O(n^{-2}) \) so that \( \frac{\text{Var}U_n}{\text{Var}U_n} \to 1 \), and then use it to show that \( \hat{U}_n \overset{d}{\to} \text{Normal} \).