Peter Bartlett

1. Projections.
2. Conditional expectations as projections.
3. Hájek projections.
Review. \textit{U}-statistics

**Definition:** A \textit{U}-statistic of order \( r \) with kernel \( h \) is

\[
U = \frac{1}{\binom{n}{r}} \sum_{i \subseteq [n]} h(X_{i_1}, \ldots, X_{i_r}),
\]

where \( h \) is symmetric in its arguments.
Review. Variance of U-statistics

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

\[ = \sum_{c=1}^{r} \theta(n^{-c}) \zeta_c, \]

\[ \zeta_c = \text{Cov}(h(X_S), h(X_{S'})) \quad \text{where } |S \cap S'| = c \]

\[ = \text{Var}(\mathbb{E}[h(X_1) | X_1^c]). \]

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

\[ n \text{Var}(U) \to \frac{nr!(n-r)!r(n-r)!}{n!(r-1)!(n-2r+1)!} \zeta_1 \to r^2 \zeta_1. \]
Review. Asymptotic distribution of U-statistics

Theorem:

\[ X_n \rightsquigarrow X \text{ and } d(X_n, Y_n) \xrightarrow{P} 0 \implies Y_n \rightsquigarrow X. \]
Consider a random variable $T$ and a linear space $S$ of random variables, with $\mathbb{E}S^2 < \infty$ for all $S \in S$ and $\mathbb{E}T^2 < \infty$. A projection $\hat{S}$ of $T$ on $S$ is a minimizer over $S$ of $\mathbb{E}(T - S)^2$.

**Theorem:** $\hat{S}$ is a projection of $T$ on $S$ iff $\hat{S} \in S$ and, for all $S \in S$, the error $T - \hat{S}$ is orthogonal to $S$, that is,

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

If $\hat{S}_1$ and $\hat{S}_2$ are projections of $T$ onto $S$, then $\hat{S}_1 = \hat{S}_2$ a.s.
Consider $S_n$ 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 $S_n$,

$$\frac{\operatorname{Var}(T_n)}{\operatorname{Var}(\hat{S}_n)} \to 1 \implies \frac{T_n - \mathbb{E}T_n}{\sqrt{\operatorname{Var}(T_n)}} - \frac{\hat{S}_n - \mathbb{E}\hat{S}_n}{\sqrt{\operatorname{Var}(\hat{S}_n)}} \xrightarrow{P} 0.$$
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 $\mathbb{E}X = \arg \min_{a \in \mathbb{R}} \mathbb{E}(X - a)^2$,

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

This is the projection of $X$ onto the linear space $S$ of measurable functions of $Y$. 


Conditional Expectations as Projections

The projection theorem says: for all measurable $g$,

$$E(X - E[X|Y])g(Y) = 0.$$ 

Properties of $E[X|Y]$:

- $EX = E[E[X|Y]]$ (consider $g = 1$).
- For a joint density $f(x, y)$,

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

- For independent $X, Y$, $E(X - EX)g(Y) = 0$, so $E[X|Y] = EX$. 

Conditional Expectations as Projections

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

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

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

**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

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

of measurable functions satisfying $Eg_i(X_i)^2 < \infty$. 
**Hájek Projection**

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

$$\hat{S} = \sum_{i=1}^{n} \mathbb{E}[T|X_i] - (n - 1)\mathbb{E}T.$$
Hájek Projection 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 $\mathbb{E} [T|X_i] = \mathbb{E} [\hat{S}|X_i]$: 

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

But

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

$$
= \mathbb{E}[T|X_i] + \sum_{j \neq i} \mathbb{E}[\mathbb{E}[T|X_j]|X_i] - (n-1)\mathbb{E}T
$$

$$
= \mathbb{E}[T|X_i],
$$

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

Asymptotic Normality of U-Statistics

**Theorem:** If $\mathbb{E}h^2 < \infty$, 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

$$h_1(x) = \mathbb{E}h(x, X_2, \ldots, X_r) - \theta,$$

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

hence,

$$\sqrt{n}(U - \theta) \xrightarrow{d} N(0, r^2 \zeta_1),$$

where

$$\zeta_1 = \mathbb{E}h_1^2(X_1).$$
Asymptotic Normality of U-Statistics: Proof

Recall:

\[ U = \frac{1}{\binom{n}{r}} \sum_{j \subseteq [n]} h(X_{j1}, \ldots, X_{jr}). \]

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

\[ \hat{U} = \sum_{i=1}^{n} \mathbb{E}[U - \theta | X_i] \]

\[ = \sum_{i=1}^{n} \frac{1}{\binom{n}{r}} \sum_{j \subseteq [n]} \mathbb{E}[h(X_{j1}, \ldots, X_{jr}) - \theta | X_i]. \]

But

\[ \mathbb{E}[h(X_{j1}, \ldots, X_{jr}) - \theta | X_i] = \begin{cases} h_1(X_i) & \text{if } i \in j, \\ 0 & \text{otherwise.} \end{cases} \]
Asymptotic Normality of U-Statistics: Proof

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

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

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 $U$:

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

$$= \frac{r^2}{n} \text{Var}(E[h(X_1^r)|X_1]) = \frac{r^2}{n} \zeta_1.$$
Asymptotic Normality of U-Statistics: Proof

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

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

$$\frac{U - \theta}{\sqrt{\text{Var}(U)}} - \frac{\hat{U}}{\sqrt{\text{Var}(\hat{U})}} \overset{P}{\to} 0,$$

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

$$\sqrt{n}(U - \theta) \rightsquigarrow N(0, r^2 \zeta_1).$$
Asymptotic Normality of U-Statistics: Examples

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

\[ \zeta_1 = \frac{1}{4}(\mu_4 - \sigma^4), \]

where \( \mu_4 = \mathbb{E}((X_1 - \mu)^4) \) is the 4th central moment. So
\[ n \operatorname{Var}(U) \to \mu_4 - \sigma^4, \] hence
\[ \sqrt{n}(U - \sigma^2) \rightsquigarrow N(0, \mu_4 - \sigma^4). \]
Recall Kendall’s $\tau$: 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: $P_1P_2$ is the line from $P_1$ to $P_2$]

$$h(P_1, P_2) = (1[P_1P_2 \text{ has positive slope}] - 1[P_1P_2 \text{ has negative slope}]),$$

$E\tau = 0,$

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

$$= \frac{1}{9},$$

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

$$Pr(\sqrt{9n/4}|\tau| > z_{\alpha/2}) \rightarrow \alpha.$$
Recall Wilcoxon’s one sample rank statistic:

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

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

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

\[ h_1(X_i) = n 1[X_i > 0]. \]

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.
Asymptotic Normality of U-Statistics: Examples

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

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

The Hájek projection of \( U - \theta \) is

\[ \hat{U} = \frac{2}{n} \sum_{i=1}^{n} h_1(X_i), \]
and

\[ h_1(x) = \mathbb{E} h(x, X_2) - \mathbb{E} h(X_1, X_2) \]

\[ = \binom{n}{2} (P(x + X_2 > 0) - P(X_1 + X_2 > 0)) \]

\[ = -\binom{n}{2} (F(-x) - \mathbb{E} F(-X_1)). \]
Asymptotic Normality of U-Statistics: Examples

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

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

$$= \frac{2 \binom{n}{2}}{n} \sum_{i=1}^{n} (F(X_i) - EF(X_i)).$$

But $F(X_i)$ is always uniform on $[0, 1]$, and so $EF(X_i) = 1/2$ and $VarF(X_i) = 1/12$. Thus,

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

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