Theoretical Statistics. Lecture 7.

Peter Bartlett

1. Projections.
2. Conditional expectations as projections.
3. Hájek projections.
4. Asymptotic normality of U-statistics. Examples.

Review. U-statistics

Definition: A 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

$$\begin{aligned}
\mathrm{Var}(U) &= \frac{1}{\binom{n}{r}} \sum_{c=1}^{r} \binom{r}{c} \binom{n-r}{r-c} \zeta_c \\
&= \sum_{c=1}^{r} \theta(n^{-c}) \zeta_c, \\
\zeta_c &= \mathrm{Cov}(h(X_S), h(X_{S'})) \qquad \text{where } |S \cap S'| = c \\
&= \mathrm{Var}\left(\mathbf{E}\left[\, h(X_1^r) \middle|\, X_1^c\right]\right).
\end{aligned}$$

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

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

Review. Asymptotic distribution of U-statistics

Theorem:

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

Review. Projection Theorem

Consider a random variable T and a linear space $\mathcal{S}$ of random variables, with $\mathbf{E}S^2 < \infty$ for all $S \in \mathcal{S}$ and $\mathbf{E}T^2 < \infty$. 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.

Review. 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 $\mathcal{S}_n$,

$$\frac{\mathrm{Var}(T_n)}{\mathrm{Var}(\hat{S}_n)} \to 1 \qquad \Longrightarrow \qquad \frac{T_n - \mathbf{E}T_n}{\sqrt{\mathrm{Var}(T_n)}} - \frac{\hat{S}_n - \mathbf{E}\hat{S}_n}{\sqrt{\mathrm{Var}(\hat{S}_n)}} \xrightarrow{P} 0.$$

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 $\mathbf{E}X = \arg\min_{a\in\mathbb{R}} \mathbf{E}(X-a)^2$,

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

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 $\mathbf{E}[X|Y]$:

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

Conditional Expectations as Projections

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

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

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 $\mathbf{E} g_i(X_i)^2 < \infty$.

Hájek Projection

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

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

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 $\mathbf{E}\left[T|X_i\right] = \mathbf{E}\left[\hat{S}|X_i\right]$:

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

But

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

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

Asymptotic Normality of U-Statistics

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

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

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

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

$$\sqrt{n}(U - \theta) \rightsquigarrow N(0, r^2\zeta_1), \qquad \text{where}$$

$$\zeta_1 = \mathbf{E}h_1^2(X_1).$$

Asymptotic Normality of U-Statistics: Proof

Recall:

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

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

$$\begin{aligned} \hat{U} &= \sum_{i=1}^{n} \mathbf{E}[U - \theta | X_i] \\ &= \sum_{i=1}^{n} \frac{1}{\binom{n}{r}} \sum_{j \subseteq [n]} \mathbf{E}[h(X_{j_1}, \ldots, X_{j_r}) - \theta | X_i]. \end{aligned}$$

But

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

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 $\mathbf{E}\hat{U} = 0$ and so its variance is asymptotically the same as that of U:

$$\begin{aligned} \operatorname{var} \hat{U} &= \frac{r^2}{n} \mathbf{E} h_1^2(X_1) = \frac{r^2}{n} \mathbf{E}(\mathbf{E}[h(X_1^r)|X_1] - \theta)^2 \\ &= \frac{r^2}{n} \mathrm{Var}(\mathbf{E}[h(X_1^r)|X_1]) = \frac{r^2}{n} \zeta_1. \end{aligned}$$

Asymptotic Normality of U-Statistics: Proof

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

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

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

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

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

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

Asymptotic Normality of U-Statistics: 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_1P_2 is the line from P_1 to P_2]

$$
\begin{aligned}
h(P_1, P_2) &= (1[P_1P_2 \text{ has positive slope}] - 1[P_1P_2 \text{ has negative slope}])\,, \\
\mathbf{E}\tau &= 0, \\
\zeta_1 &= \text{Cov}(h(P_1, P_2), h(P_1, P_3)) \\
&= \frac{1}{9},
\end{aligned}
$$

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

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

Asymptotic Normality of U-Statistics: Examples

Recall Wilcoxon's one sample rank statistic:

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

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

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

The Hájek projection of $U - \theta$ is

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

and

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

Asymptotic Normality of U-Statistics: Examples

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

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

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

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

Asymptotic Normality of U-Statistics: Examples

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:

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