STAT 620: Asymptotic Statistics **Spring 2022**

Lecture: Mar 3

Lecturer: Xianyang Zhang

1 Hajek projection

The main idea is to use Hajek projections onto sets of the form:

$$\mathcal{S}_n = \left\{ \sum_{i=1}^{n} g_i(X_i) : g_i(X_i) \in L^2(P) \right\}$$

to approximate U_n by a sum of independent random variables.

1.1 Theorem

Let h be a symmetric kernel of order r with $E[h^2(X_1, \ldots, X_r)] < \infty$. Let U_n be the associated U-statistic and $\theta = E[U_n]$. If $\hat{U}_n$ is the projection of $U_n - \theta$ onto $\mathcal{S}_n$, then

$$\hat{U}_n = \sum_{i=1}^{n} E[U_n - \theta | X_i] = \frac{r}{n} \sum_{i=1}^{n} \hat{h}_1(X_i),$$

where $\hat{h}_1(x) = E[h(x, X_2, \ldots, X_r)] - \theta$.

Proof: The first equality is just a direct application of the last result from the previous lecture, noting that $E[U_n - \theta] = 0$. To show the second equality, we note that

$$E[h(X_\beta) - \theta | X_i] = \begin{cases} 0, & \text{if } i \notin \beta, \\ \hat{h}_1(X_i), & \text{if } i \in \beta. \end{cases}$$

Thus we get

$$\begin{aligned} E[U_n - \theta | x_i] =& \binom{n}{r}^{-1} \sum_{\beta \subset [n], |\beta| = r, i \in \beta} \hat{h}_1(X_i) \\ =& \binom{n}{r}^{-1} \binom{n-1}{r-1} h_1(X_i) = \frac{r}{n} \hat{h}_1(X_i), \end{aligned}$$

which implies the result.

1.2 Theorem

We have

1. $\sqrt{n}\hat{U}_n \longrightarrow^d N(0, r^2\xi_1)$.
2. $\sqrt{n}(U_n - \theta - \hat{U}_n) \longrightarrow^p 0$.
3. $\sqrt{n}(U_n - \theta) = \sqrt{n}\hat{U}_n + o_p(1) \longrightarrow^d N(0, r^2\xi_1)$.

Proof: The first result follows from CLT. As

$$\text{var}(U_n) = \frac{r^2}{n}\xi_1 + o(n^{-2})$$

and

$$\operatorname{var}(\hat{U}_n) = \frac{r^2}{n}\xi_1,$$

we have

$$\frac{\operatorname{var}(\sqrt{n}(U_n - \theta))}{\operatorname{var}(\sqrt{n}\hat{U}_n)} \longrightarrow 1.$$

Using a result before, we have

$$\frac{\sqrt{n}(U_n - \theta)}{r\xi_1} - \frac{\sqrt{n}\hat{U}_n}{r\xi_1} \longrightarrow^p 0.$$

By application of Slutsky's theorem, we can conclude the desired results.