Notes on the Hájek projection and Hoeffding Decomposition

Jon A. Wellner

May 3, 2011

1. The Hájek projection

The following treatment is largely from Van der Vaart (1998).

Suppose that $X_1, \ldots, X_n$ are independent random vectors, and let $\mathcal{S}$ denote the set of all variables of the form

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

for arbitrary measurable functions $g_i : \mathbb{R}^d \to \mathbb{R}$ with $E g_i^2(X_i) < \infty$ for $i \in \{1, \ldots, n\}$.

Lemma 1. The projection of an arbitrary random variable $T = T(X_1, \ldots, X_n)$ with finite second moment onto $\mathcal{S}$ is given by

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

$\hat{S}$ is called the *Hájek projection* of T onto $\mathcal{S}$.

Proof. Note that $\hat{S} \in \mathcal{S}$. Thus it suffices to show that

$$E\{(T - \hat{S})S\} = 0 \quad \text{for all} \quad S \in \mathcal{S}.$$

But

$$\begin{aligned} E\{(T - \hat{S})S\} &= E\{(T - \hat{S}) \sum_{i=1}^{n} g_i(X_i)\} \\ &= \sum_{i=1}^{n} E\{(T - \hat{S}) g_i(X_i)\} \\ &= \sum_{i=1}^{n} E E\{T - \hat{S}) g_i(X_i) | X_i\} \end{aligned}$$

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

where

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

for each $i \in \{1, \ldots, n\}$. Thus $E[(T-\hat{S})S] = 0$ for each $S \in \mathcal{S}$. □

Note that if $X_1, \ldots, X_n$ are also identically distributed and $T(\underline{X}) \equiv T(X_1, \ldots, X_n)$ is permutation symmetric (so that $T(\pi\underline{X}) = T(\underline{X})$ for all $\pi \in \Pi_n$), then

$$
\begin{aligned}
E(T|X_i = x) &= E(T|X_1 = x) \quad \text{for all} \quad i \in \{1, \ldots, n\}, \\
&= ET(x, X_2, \ldots, X_n)
\end{aligned}
$$

which does not depend on i. It follows that in this case $\hat{S}$ is also the projection of T onto the smaller set $\mathcal{S}_0$ consisting of all variables of the form $\sum_{i=1}^{n} g(X_i)$ for an arbitrary measurable function g.

Now suppose that $\{T_n\}$ is a sequence of statistics T_n ((think of $T_n = T(\mathbb{P}_n)$ or $T_n = T_n(X_1, \ldots, X_n)$ based on independent X_i's), and suppose that $\mathcal{S}_n$ is a sequence of linear spaces. Let $\hat{S}_n$ denote the (L_2-) projection of T_n onto $\mathcal{S}_n$.

Theorem 1. If the linear spaces $\mathcal{S}_n$ contain constants and $Var(T_n)/Var(\hat{S}_n) \to 1$, then

$$
R_n \equiv \frac{T_n - E(T_n)}{\sqrt{Var(T_n)}} - \frac{\hat{S}_n - E(\hat{S}_n)}{\sqrt{Var(\hat{S}_n)}} \ \to_p \ 0.
$$

Proof. Now $E(R_n) = 0$ and

$$
\begin{aligned}
Cov(T_n, \hat{S}_n) &= E(T_n\hat{S}_n) - E(T_n)E(\hat{S}_n) \\
&= E\hat{S}_n^2 - \{E(\hat{S}_n)\}^2 = Var(\hat{S}_n)
\end{aligned}
$$

since both

$$
E[(T_n - \hat{S}_n)\hat{S}_n] = 0 \quad \text{and} \quad E[(T - \hat{S}_n)1] = 0.
$$

Therefore

$$\begin{aligned} Var(R_n) &= 2 - 2\frac{Cov(T_n, \hat{S}_n)}{\sqrt{Var(T_n)\cdot Var(\hat{S}_n)}} \\ &= 2\left(1 - \frac{Var(\hat{S}_n)}{\sqrt{Var(T_n)\cdot Var(\hat{S}_n)}}\right) \\ &= 2\left(1 - \sqrt{\frac{Var(\hat{S}_n)}{Var(T_n)}}\right) \\ &\to 0. \end{aligned}$$

The claimed result follows by Chebychev's (or Markov's) inequality. □

2. $U-$ statistics and Hájek's projection

Now suppose that $X_1, \ldots, X_n$ are i.i.d. random vectors with distribution P on $(\mathbb{R}^d, \mathcal{B}_d) = (\mathcal{X}, \mathcal{A})$ with $n \geq r$. Suppose that $h : \mathcal{X}^r \to \mathbb{R}$ is measurable and permutation symmetric: $h(\pi \underline{x}) = h(\underline{x})$ for all $\pi \in \Pi_r$. Consider the

$$\theta \equiv T(P) = E_P h(X_1, \ldots, X_r).$$

Thus $h(X_1, \ldots, X_r)$ is an unbiased estimator of θ. Since n observations (with $n \geq r$ are available, this simple estimator can be improved: by Rao-Blackwell, the new unbiased estimator formed by computing the conditional expectation given a sufficient statistic has smaller variance. Here, for X_i's with values in $\mathbb{R}$, the vector of order statistics $X_{(1)}, \ldots, X_{(n)}$ is sufficient; and for i.i.d. X_i's more generally, the empirical measure $\mathbb{P}_n$ is sufficient (see e.g. Dudley (2002), Theorem 1.5.9, page 177):

$$\begin{aligned} U_n &= \begin{cases} E\left(h(X_1, \ldots, X_r) | X_{(1)}, \ldots, X_{(n)}\right), & \text{if } \mathcal{X} = \mathbb{R}, \\ E\left(h(X_1, \ldots, X_r) | \mathbb{P}_n\right), & \text{in general.} \end{cases} \\ &= \frac{1}{\binom{n}{r}} \sum_c h(X_{i_1}, \ldots, X_{i_r}) \end{aligned}$$

where $(i_1, \ldots, i_r)$ denotes one of the $\binom{n}{r}$ collections of r distinct integers between 1 and n.

Now the Hájek projection of U_n is $\hat{S}_n = \hat{U}_n$ given by

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

where

$$h_1(x) = E h(x, X_2, \ldots, X_r) - \theta.$$

Theorem 2. If $Eh^2(X_1, \ldots, X_r) < \infty$ and $\zeta_1 \equiv Var(h_1(X_1)) > 0$, then

$$\sqrt{n}(U_n - \theta - \hat{U}_n) \to_p 0.$$

Hence

$$\begin{aligned} \sqrt{n}(U_n - \theta) &= \sqrt{n}\hat{U}_n + o_p(1) = \frac{r}{\sqrt{n}} \sum_{i=1}^n h_1(X_i) + o_p(1) \\ &\to_d N(0, r^2 Var(h_1(X_1))) \end{aligned}$$

where

$$\zeta_1 = Var(h_1(X_1)) = Cov(h(X_1, X_2, \ldots, X_r), h(X_1, X_2', \ldots, X_r')).$$

Proof. The first job is to prove that the form of the Hájek projection $\hat{U}_n$ is as claimed. Since the X_i's are independent and h is permutation symmetric

$$E\left(h(X_{i_1}, \ldots, X_{i_r}) - \theta \middle| X_i = x\right) = \begin{cases} h_1(x), & \text{if } i \in \{i_1, \ldots, i_r\}, \\ 0, & \text{if } i \notin \{i_1, \ldots, i_r\}. \end{cases}$$

Then we compute

$$\begin{aligned} E(U_n - \theta | X_i) &= \frac{1}{\binom{n}{r}} \sum_c h_1(X_i) 1_{\{i_1, \ldots, i_r\}}(i) \\ &= \frac{\binom{n-1}{r-1}}{\binom{n}{r}} h_1(X_i) = \frac{r}{n} h_1(X_i). \end{aligned}$$

Summing these over i gives the claimed result.

To apply Theorem 1 we need to show that

$$\frac{Var(\hat{U}_n)}{Var(U_n)} \to 1. \tag{2.1}$$

Calculation of $Var(\hat{U}_n)$ is easy:

$$Var(\hat{U}_n) = \frac{r^2}{n^2} \sum_{i=1}^n Var(h_1(X_i) = \frac{r^2}{n} \zeta_1. \tag{2.2}$$

Since the random variables $Y_i \equiv h_1(X_i)$ are i.i.d. with mean zero and finite variance, $\sqrt{n}\hat{U}_n \to_d N(0, r^2\zeta_1)$ by the (ordinary) CLT.

Calculation of $Var(U_n)$ is a bit more involved, but not too bad. Let k be the number of indices in common between $\{i_1, \ldots, i_r\}$ and $\{i_1', \ldots, i_r'\}$. Then, since there are $\binom{n}{r}$ ways of choosing

$\{i_1, \ldots, i_r\}$, $\binom{r}{k}$ ways of choosing k common indices from the choice of $\{i_1, \ldots, i_r\}$, and then $\binom{n-r}{r-k}$ ways of choosing the rest of the second block $\{i_1', \ldots, i_r'\}$, we find that

$$
\begin{aligned}
Var(U_n) &= \frac{1}{\binom{n}{r}^2} \sum_c \sum_{c'} Cov(h(X_{i_1}, \ldots, X_{i_r}), h(X_{i_1'}, \ldots, X_{i_r'})) \\
&= \frac{1}{\binom{n}{r}^2} \sum_{k=1}^{r} \binom{n}{r} \binom{r}{k} \binom{n-r}{r-k} \zeta_k \\
&= \sum_{k=1}^{r} \frac{\binom{r}{k}\binom{n-r}{r-k}}{\binom{n}{r}} \cdot \zeta_k \qquad (2.3) \\
&= E\zeta_K
\end{aligned}
$$

where $\zeta_0 \equiv 0$ and for $k \in \{1, \ldots, r\}$,

$$
\zeta_k = Cov(h(X_1, \ldots, X_k, X_{k+1}, \ldots, X_r), h(X_1, \ldots, X_k, X_{k+1}', \ldots, X_r')),
$$

and where $K \sim$ Hypergeometric(n, r) (sampling without replacement from an urn containing n balls, r of which are red and $n-r$ of which are blue): i.e.

$$
P(K = k) = \frac{\binom{r}{k}\binom{n-r}{r-k}}{\binom{n}{r}} \quad \text{for} \quad k \in \{1, \ldots, r\}.
$$

The expression in (2.3) can be rewritten as

$$
\begin{aligned}
\sum_{k=1}^{r} \frac{\binom{r}{k}\binom{n-r}{r-k}}{\binom{n}{r}} \cdot \zeta_k &= \sum_{k=1}^{r} \frac{(r!)^2}{k!(r-k)!^2} \frac{(n-r)\cdots(n-2r+k+1)}{n(n-1)\cdots(n-r+1)} \cdot \zeta_k \\
&\sim r^2 \zeta_1 \cdot \frac{1}{n} \quad \text{if} \quad \zeta_1 > 0.
\end{aligned}
$$

Putting this together with (2.2) yields (2.1) and completes the proof. □

REFERENCES

DUDLEY, R. M. (1999). *Uniform Central Limit Theorems.* Cambridge University Press, Cambridge.

SERFLING, ROBERT J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

SPROULE, R. N. (1974). Asymptotic properties of $U-$statistics. *Trans. Amer. Math. Soc.* **199**, 55 - 64.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

Department of Statistics
University of Washington
P.O. Box 354322
Seattle, Washington 98195-4322
U.S.A.
e-mail: jaw@stat.washington.edu