Some notes on U-Statistics

Ahnaf Rafi

August 30, 2025

Contents

1 Definitions of U-Statistics and V-Statistics

Throughout, all randomness is defined on a probability space $(\Omega, \mathscr{F}, \Pr)$ with $\mathrm{E}[\cdot]$ and $\mathrm{Var}[\cdot]$ denoting expectation and variance respectively against Pr. Furthermore, $\{Z_i\}_{i=1}^n$ is a sample of i.i.d. random vectors, with F denoting their common distribution. Finally $\|\cdot\|$ denotes the Euclidean norm for vectors.

For $n, m \in \mathbb{N}$ with $n \geq m$, define

$$\begin{aligned}
\mathbb{N}_n &= \{1, \ldots, n\} = \{j \in \mathbb{N} : j \leq n\}, \\
\mathrm{Inj}_{n,m} &= \{i \in \mathbb{N}_n^m : [j, k \in \mathbb{N}_m, j \neq k] \implies i(j) \neq i(k)\}, \\
\mathrm{NonInj}_{n,m} &= \mathbb{N}_n^m \setminus \mathrm{Inj}_{n,m} = \{i \in \mathbb{N}_n^m : \exists j, k \in \mathbb{N}_m, j \neq k \text{ such that } i(j) = i(k)\}.
\end{aligned} \tag{1.1}$$

In (1.1), Inj stands for "injective" (i.e. one-to-one) and NonInj stands for "non-injective". The cardinality of $\mathrm{Inj}_{n,m}$ is

$$\gamma_{n,m} := \left|\mathrm{Inj}_{n,m}\right| = \prod_{k=0}^{m-1} (n-k). \tag{1.2}$$

For a function $\eta(x_1, \ldots, x_m)$, the associated U- and V-statistics are respectively:

$$\begin{aligned}
U_n(\eta) &= \frac{1}{\gamma_{n,m}} \sum_{i \in \mathrm{Inj}_{n,m}} \eta\left(Z_{i(1)}, \ldots, Z_{i(m)}\right), \\
V_n(\eta) &= \frac{1}{n^m} \sum_{i \in \mathbb{N}_n^m} \eta\left(Z_{i(1)}, \ldots, Z_{i(m)}\right).
\end{aligned} \tag{1.3}$$

The function η is called a *kernel* and the number of arguments $m := m_\eta$, is called the *order* of the kernel. Though the order m is dependent on η, we suppress this dependence for brevity. In addition, we hold m fixed throughout and impose the restriction $n \geq m$.

The kernel η is permutation symmetric iff

$$\text{for any permutation } \{i(1), \ldots, i(m)\} \text{ of } \{1, \ldots, m\},$$
$$\eta\left(z_{i(1)}, \ldots, z_{i(m)}\right) = \eta\left(z_1, \ldots, z_m\right).$$

When η is permutation symmetric, the corresponding U-statistic in (1.3) becomes

$$U_n(\eta) = \binom{n}{m}^{-1} \sum_{i \in \mathrm{Inc}_{n,m}} \eta\left(Z_{i(1)}, \ldots, Z_{i(m)}\right),$$

where $\mathrm{Inc}_{n,m} \subseteq \mathrm{Inj}_{n,m}$ is the collection of strictly increasing index vectors:

$$\mathrm{Inc}_{n,m} = \left\{(i(1), \ldots, i(m)) \in \mathbb{N}_n^m : [j, l \in \mathbb{N}_m, j < l] \implies i(j) < i(l)\right\}. \tag{1.4}$$

2 First order sum-space projection of U-statistics

We now consider the projection of the U-statistic $U_n(\eta, F)$ onto the sum-space for the n i.i.d. random variables $Z_1, \ldots, Z_n$ drawn from F. Denote this sum space by

$$\mathcal{S}_n := \mathcal{S}_n(F) := \left\{\sum_{j=1}^{n} g_j\left(Z_j\right) : \forall j \in \mathbb{N}_n, Z_j \sim F, g_j \text{ measurable}, \int g_j^2 \mathrm{d}F < \infty\right\}. \tag{2.1}$$

Henceforth, let $m \geq 2$. Denote

$$\Pi_0(\eta) := \Pi_0(\eta, F) := \mathrm{E}_{F^m}\left[\eta\left(Z_1, \ldots, Z_m\right)\right], \tag{2.2}$$

$$\begin{aligned}\forall c \in \mathbb{N}_{m-1}, \quad \Pi_c\left(z_1, \ldots, z_c; \eta\right) &:= \Pi_c\left(z_1, \ldots, z_c; \eta, F\right) \\ &:= \mathrm{E}_{F^m}\left[\eta\left(Z_1, \ldots, Z_m\right) | Z_1 = z_1, \ldots, Z_c = z_c\right] \\ &= \int \eta\left(z_1, \ldots, z_c, z_{c+1}, \ldots, z_m\right) F^{m-c}\left(\mathrm{d}z_{c+1}, \ldots, \mathrm{d}z_m\right),\end{aligned} \tag{2.3}$$

$$
\begin{aligned}
\text{and} \quad \widehat{U}_n(\eta) :=& \widehat{U}_n(\eta, F) \\
:=& \Pi_0(\eta, F) + \frac{m}{n}\sum_{j=1}^{n}\left\{\Pi_1\left(Z_j; \eta, F\right) - \Pi_0(\eta, F)\right\} \\
=& \Pi_0(\eta) + \frac{m}{n}\sum_{j=1}^{n}\left\{\Pi_1\left(Z_j; \eta\right) - \Pi_0(\eta)\right\}.
\end{aligned}
\tag{2.4}
$$

In all of the above expressions, the notation expresses the fact that though the functions Π_c and $\widehat{U}_n$ all depend on both η and F, we shall suppress dependence on F. We follow this scheme throughout, so that $\mathrm{E} \equiv \mathrm{E}_{F^n}$ throughout.

Lemma 2.1. *$\widehat{U}_n(\eta)$ in (2.4) is the projection of $U_n(\eta)$ onto the sum space $\mathcal{S}_n$ in (2.1).*

Proof of Lemma 2.1. We apply Theorem A.4. Take any $i \in \mathrm{Inj}_{n,m}$. Then, for any $j = 1, \ldots, n$,

$$
\mathrm{E}\left[\eta\left(Z_{i(1)}, \ldots, Z_{i(m)}\right) \middle| Z_j = z\right] = \begin{cases} \Pi_1(z; \eta) & \text{if } j \in \{i(1), \ldots, i(m)\}, \\ \Pi_0(\eta) & \text{otherwise.} \end{cases}
$$

For fixed $j \in \{1, \ldots, n\}$, consider averaging over $i \in \mathrm{Inj}_{n,m}$. The first case happens $m \cdot \gamma_{n-1,m-1}$ times. To see this, given an 'empty' m-vector i, there are m possible places to place j and $\gamma_{n-1,m-1}$ ways to select the remaining elements of i. Hence

$$
\begin{aligned}
\mathrm{E}\left[U_n(\eta) | Z_j\right] &= \frac{1}{\gamma_{n,m}} \sum_{i \in \mathrm{Inj}_{n,m}} \mathrm{E}\left[\eta\left(Z_{i(1)}, \ldots, Z_{i(m)}\right) \middle| Z_j\right] \\
&= m\frac{\gamma_{n-1,m-1}}{\gamma_{n,m}}\Pi_1\left(Z_j; \eta\right) + \left(1 - m\frac{\gamma_{n-1,m-1}}{\gamma_{n,m}}\right)\Pi_0(\eta).
\end{aligned}
$$

Since $\gamma_{n-1,m-1}/\gamma_{n,m} = 1/n$,

$$
\mathrm{E}\left[U_n(\eta) | Z_j\right] = \frac{m}{n}\Pi_1\left(Z_j; \eta\right) + \left(1 - \frac{m}{n}\right)\Pi_0(\eta).
$$

Apply Theorem A.4; the projection of $U_n(\eta)$ is therefore

$$
\begin{aligned}
&\mathrm{E}\left[U_n(\eta)\right] + \sum_{j=1}^{n} \left(\mathrm{E}\left[U_n(\eta)|Z_j\right] - \mathrm{E}\left[U_n(\eta)\right]\right) \\
=& \Pi_0(\eta) + \sum_{j=1}^{n} \left(\frac{m}{n}\Pi_1\left(Z_j;\eta\right) + \left(1-\frac{m}{n}\right)\Pi_0(\eta) - \Pi_0(\eta)\right) \\
=& \Pi_0(\eta) + \frac{m}{n}\sum_{j=1}^{n}\left\{\Pi_1\left(Z_j;\eta\right) - \Pi_0(\eta)\right\} \\
=& \widehat{U}_n(\eta),
\end{aligned}
$$

as desired. □

3 $\mathscr{L}_2$ distance between a U-statistic and its first order sum-space projection under permutation symmetry

In this section, we assume that the kernel η is permutation symmetric. Hence, the relevant set of indices to work with is $\mathrm{Inc}_{n,m}$ in (1.4). Consider the (squared) $\mathscr{L}_2$ distance between a U-statistic and its projection, $\widehat{U}_n(\eta, F)$, onto the first order sum space $\mathcal{S}_n$ defined in (2.1):

$$
\rho_n(\eta) := \rho_n(\eta, F) := \mathrm{E}_{F^n}\left[\left(U_n(\eta) - \widehat{U}_n(\eta, F)\right)^2\right]. \tag{3.1}
$$

We can readily apply Theorem A.3 since $\mathcal{S}_n$ contains the constant (non-stochastic) random variables. Hence, (3.1) becomes

$$
\rho_n(\eta) = \mathrm{Var}\left[U_n(\eta)\right] - \mathrm{Var}\left[\widehat{U}_n(\eta)\right]. \tag{3.2}
$$

The variance of $\widehat{U}_n(\eta)$ is easily characterized:

$$
\mathrm{Var}\left[\widehat{U}_n(\eta)\right] = \frac{m^2}{n}\mathrm{Var}\left[\Pi_1(Z;\eta)\right]. \tag{3.3}
$$

Furthermore, we can always write

$$\operatorname{Var}\left[U_n(\eta)\right]=\binom{n}{m}^{-2}\sum_{i\in\mathrm{Inc}_{n,m}}\sum_{i'\in\mathrm{Inc}_{n,m}}\operatorname{Cov}\left(\eta\left(Z_{i(1)},\ldots,Z_{i(m)},\right),\eta\left(Z_{i'(1)},\ldots,Z_{i'(m)}\right)\right). \quad (3.4)$$

We now introduce a notation scheme that will help us to simplify (3.4) and will allow us to provide a unified treatment of (3.3) and (3.4) under i.i.d. $Z_1,\ldots,Z_n$. For $0\le c\le m$, denote

$$\begin{aligned}\zeta_c(\eta):=\zeta_c(\eta,F)&:=\operatorname{Var}_{F^c}\left[\Pi_c\left(Z_1,\ldots,Z_c;\eta,F\right)\right]\\&=\operatorname{Var}_{F^c}\left[\mathrm{E}_{F^m}\left[\eta\left(Z_1,\ldots,Z_m\right)|Z_1,\ldots,Z_c\right]\right].\end{aligned} \quad (3.5)$$

It can be shown that $\zeta_c(\eta)$ has the following alternative representation:

$$\zeta_c(\eta)=\operatorname{Cov}\left(\eta\left(Z_1,\ldots,Z_c,Z_{1,c+1},\ldots,Z_{1,m}\right),\eta\left(Z_1,\ldots,Z_c,Z_{2,c+1},\ldots,Z_{2,m}\right)\right), \quad (3.6)$$

where $\{Z_j\}_{j=1}^n$, $\{Z_{1,j}\}_{j=1}^n$ and $\{Z_{2,j}\}_{j=1}^n$ are all mutually independent random vectors all with distribution F. That is, $\zeta_c(\eta)$ is the covariance of $\eta\left(Z_{i(1)},\ldots,Z_{i(m)}\right)$ and $\eta\left(Z_{i'(1)},\ldots,Z_{i'(m)}\right)$ when i and i' are strictly increasing and have exactly c elements in common. Of course for $c=0$, $\zeta_0(\eta)=0$ by independence of the Z_j's.

Revisiting (3.3) using (3.5),

$$\operatorname{Var}\left[\widehat{U}_n(\eta)\right]=\frac{m^2}{n}\zeta_1(\eta). \quad (3.7)$$

For (3.4), we can instead use (3.6). Since there are $\binom{n}{m}\binom{m}{c}\binom{n-m}{m-c}$ pairs $i,i'\in\mathrm{Inc}_{n,m}$ with exactly c elements in common (see Theorem B.1 for a formal statement and proof),

$$\begin{aligned}\operatorname{Var}\left[U_n(\eta)\right]&=\binom{n}{m}^{-2}\sum_{c=0}^{m}\binom{n}{m}\binom{m}{c}\binom{n-m}{m-c}\zeta_c(\eta)\\&=\binom{n}{m}^{-2}\sum_{c=1}^{m}\binom{n}{m}\binom{m}{c}\binom{n-m}{m-c}\zeta_c(\eta)\qquad[\text{by }\zeta_0(\eta)=0],\end{aligned}$$

and so,

$$\mathrm{Var}\left[U_n(\eta)\right] = \sum_{c=1}^{m} \binom{m}{c} \binom{n}{m}^{-1} \binom{n-m}{m-c} \zeta_c(\eta). \tag{3.8}$$

Combining (3.2), (3.7) and (3.8),

$$\rho_n(\eta) = \sum_{c=1}^{m} \binom{m}{c} \binom{n}{m}^{-1} \binom{n-m}{m-c} \zeta_c(\eta) - \frac{m^2}{n} \zeta_1(\eta). \tag{3.9}$$

Theorem 3.1. *For $m \in \mathbb{N} \setminus \{1\}$, let η be a m^{th} order kernel that is permutation symmetric. Furthermore, let F be a probability measure such that $\eta \in \mathscr{L}_2(F^m)$. Let $\rho_n(\eta, F)$ be as defined in (3.1). In the case of $m = 2$, we have the equality*

$$\rho_n(\eta, F) = \frac{2}{n(n-1)} \left(\mathrm{E}_F \left[\mathrm{Var}_{F^2} \left[\eta\left(Z_1, Z_2\right) | Z_1 \right] \right] - \mathrm{Var}_F \left[\mathrm{E}_{F^2} \left[\eta\left(Z_1, Z_2\right) | Z_1 \right] \right] \right). \tag{3.10}$$

More generally, for any $m \geq 2$, we can bound $\rho_n(\eta, F)$ by

$$\begin{aligned} \rho_n(\eta, F) \leq & \frac{m^2(m-1)^2 \left(1 + \frac{m-1}{n-m+1}\right)}{n^2} \zeta_1(\eta, F) \\ & + \sum_{c=2}^{m} \frac{1}{\gamma_{n,c}} \cdot \frac{(m!)^2}{((m-c)!)^2 c!} \left(1 + \delta_{n,m,c}\right) \zeta_c(\eta, F), \end{aligned} \tag{3.11}$$

where for each $c \in \{2, \ldots, m\}$, the quantities $\delta_{n,m,c}$ are do not depend on η, F or the dimension of Z_i (except possibly through n and m) and satisfy $\lim_{n \to \infty} \delta_{n,m,c} = 0$.

3.1 Proof of Theorem 3.1 in the case of $m = 2$

For $m = 2$, (3.9) becomes

$$\begin{aligned} \rho_n(\eta) &= \frac{4}{n} \left(\frac{(n-2)!}{(n-1)!} \cdot \frac{(n-2)!}{(n-3)!} - 1 \right) \zeta_1(\eta) + \binom{n}{2}^{-1} \zeta_2(\eta) \\ &= \frac{4}{n} \left(\frac{n-2}{n-1} - 1 \right) \zeta_1(\eta) + \frac{2}{n(n-1)} \zeta_2(\eta), \end{aligned}$$

and so

$$\rho_n(\eta) = \frac{2}{n(n-1)}\left(\zeta_2(\eta) - 2\zeta_1(\eta)\right). \tag{3.12}$$

Since $m = 2$, $\Pi_2(\cdot;\eta) = \eta(\cdot)$. Using (3.5),

$$\begin{aligned}
\zeta_2(\eta) &= \mathrm{Var}\left[\eta\left(Z_1, Z_2\right)\right] \\
\text{[by Law of Total Variance]} \quad &= \mathrm{Var}\left[\mathrm{E}\left[\eta\left(Z_1, Z_2\right)|Z_1\right]\right] + \mathrm{E}\left[\mathrm{Var}\left[\eta\left(Z_1, Z_2\right)|Z_1\right]\right] \\
&= \mathrm{Var}\left[\Pi_1\left(Z_1;\eta\right)\right] + \mathrm{E}\left[\mathrm{Var}\left[\eta\left(Z_1, Z_2\right)|Z_1\right]\right] \\
&= \zeta_1(\eta) + \mathrm{E}\left[\mathrm{Var}\left[\eta\left(Z_1, Z_2\right)|Z_1\right]\right].
\end{aligned}$$

Then, from (3.12), using the fact that $\zeta_1(\eta) = \mathrm{Var}\left[\mathrm{E}\left[\eta\left(Z_1, Z_2\right)|Z_1\right]\right]$ (from (3.5)),

$$\rho_n(\eta) = \frac{2}{n(n-1)}\left(\mathrm{E}\left[\mathrm{Var}\left[\eta\left(Z_1, Z_2\right)|Z_1\right]\right] - \mathrm{Var}\left[\mathrm{E}\left[\eta\left(Z_1, Z_2\right)|Z_1\right]\right]\right),$$

which is exactly (3.10). $\square$

3.2 Proof of Theorem 3.1 in the case of $m > 2$

For $m > 2$, recall that in (3.9),

$$\rho_n(\eta) = \sum_{c=1}^{m}\binom{m}{c}\binom{n}{m}^{-1}\binom{n-m}{m-c}\zeta_c(\eta) - \frac{m^2}{n}\zeta_1(\eta).$$

For $c = m$,

$$\binom{n}{m}^{-1}\binom{n-m}{0} = \frac{(n-m)!m!}{n!} = \frac{m!}{\gamma_{n,m}}$$

Thus,

$$\text{for } c = m, \quad \binom{n}{m}^{-1}\binom{n-m}{m-c} = \frac{m!}{(m-c)!}\frac{1}{\gamma_{n,c}}\left(1+\delta_{n,m,c}\right),$$

$$\text{where } \delta_{n,m,c} = 0, \text{ for each } n.$$

Now let $c \in \{2, \ldots, m-1\}$. Then,

$$\begin{aligned}\binom{n}{m}^{-1}\binom{n-m}{m-c} &= \frac{(n-m)!m!}{n!} \cdot \frac{(n-m)!}{(n-2m+c)!(m-c)!} \\ &= \frac{m!}{(m-c)!} \cdot \frac{(n-m)!}{n!}\frac{(n-m)!}{(n-2m+c)!} \\ &= \frac{m!}{(m-c)!} \cdot \frac{(n-m)\cdots(n-2m+c+1)}{n\cdots(n-m+1)}.\end{aligned}$$

The numerator has $m-c$ terms, whereas the denominator has m terms. Hence for fixed m we can factor out $1/\gamma_{n,c}$ to get

$$\begin{aligned}\binom{n}{m}^{-1}\binom{n-m}{m-c} &= \frac{m!}{(m-c)!} \cdot \frac{(n-m)\cdots(n-2m+c+1)}{n(n-1)\cdots(n-m+1)} \\ &= \frac{m!}{(m-c)!} \cdot \frac{1}{n\cdots(n-c+1)} \cdot \frac{(n-m)\cdots(n-2m+c+1)}{(n-c)\cdots(n-m+1)} \\ &= \frac{m!}{(m-c)!} \cdot \frac{1}{\gamma_{n,c}} \cdot \frac{(n-m)\cdots(n-2m+c+1)}{(n-c)\cdots(n-m+1)}.\end{aligned}$$

Factorising the ratio in the end further,

$$\binom{n}{m}^{-1}\binom{n-m}{m-c} = \frac{m!}{(m-c)!} \cdot \frac{1}{\gamma_{n,c}} \cdot \prod_{l=0}^{m-c-1}\left(1 - \frac{m-c}{n-c-l}\right).$$

Therefore, set

$$\delta_{n,m,c} = \left[\prod_{l=0}^{m-c-1}\left(1-\frac{m-c}{n-c-l}\right)\right] - 1. \tag{3.13}$$

Then,

$$\binom{n}{m}^{-1}\binom{n-m}{m-c} = \frac{m!}{(m-c)!}\frac{1}{\gamma_{n,c}}\left(1+\delta_{n,m,c}\right),$$
$$\text{where } \lim_{n\to\infty}\delta_{n,m,c} = 0, \text{ for each } c \le m.$$

Therefore, applying the above expression to (3.9)

$$\begin{aligned}\rho_n(\eta) &= \sum_{c=1}^{m}\frac{1}{\gamma_{n,c}}\cdot\frac{(m!)^2}{((m-c)!)^2 c!}\left(1+\delta_{n,m,c}\right)\zeta_c(\eta) - \frac{m^2}{n}\zeta_1(\eta)\\ &= \frac{m^2}{n}\delta_{n,m,1}\zeta_1(\eta) + \sum_{c=2}^{m}\frac{1}{\gamma_{n,c}}\cdot\frac{(m!)^2}{((m-c)!)^2 c!}\left(1+\delta_{n,m,c}\right)\zeta_c(\eta). \end{aligned} \tag{3.14}$$

For fixed η, visual inspection of (3.14), first suggests that the term corresponding to $c = 1$ will exhibit the slowest tendency towards zero, unless $\delta_{n,m,1} = O(1/n)$. We can show that $\delta_{n,m,1} = O(1/n)$ is in fact true. To that end, from (3.13),

$$\delta_{n,m,1} = \left(1 - \frac{m-1}{n-1}\right)\cdots\left(1 - \frac{m-1}{n-m+1}\right) - 1.$$

By Billingsley (1995, Lemma 1, p. 358), given any $l \in \mathbb{N}$ and $z_{1,1}, z_{2,1}, \ldots, z_{1,l}, z_{2,l} \in \mathbb{C}$ with $|z_{j,k}| \le 1$,

$$\left|z_{1,1}\cdots z_{1,l} - z_{2,1}\cdots z_{2,l}\right| \le \sum_{k=1}^{l}\left|z_{1,k} - z_{2,k}\right|.$$

Applying this with $l = m-1$, $z_{1,j} = 1-((m-1)/(n-j))$ and $z_{2,j} = 1$ for every $j = 1, \ldots, m-1$,

$$|\delta_{n,m,1}| \le \sum_{k=1}^{m-1}\frac{m-1}{n-k} \le (m-1)^2\left(1 + \frac{m-1}{n-m+1}\right)\cdot\frac{1}{n}. \tag{3.15}$$

Revisiting (3.14),

$$\rho_n(\eta) = \frac{m^2}{n}\delta_{n,m,1}\zeta_1(\eta) + \sum_{c=2}^{m}\frac{1}{\gamma_{n,c}}\cdot\frac{(m!)^2}{((m-c)!)^2 c!}\left(1+\delta_{n,m,c}\right)\zeta_c(\eta)$$

$$\leq \frac{m^2}{n} |\delta_{n,m,1}| \, \zeta_1(\eta) + \sum_{c=2}^{m} \frac{1}{\gamma_{n,c}} \cdot \frac{(m!)^2}{((m-c)!)^2 c!} \left(1 + \delta_{n,m,c}\right) \zeta_c(\eta),$$

so that

$$\begin{aligned}\rho_n(\eta) \leq &\frac{m^2(m-1)^2 \cdot \left(1 + \frac{m-1}{n-m+1}\right)}{n^2} \zeta_1(\eta) \\ &+ \sum_{c=2}^{m} \frac{1}{\gamma_{n,c}} \cdot \frac{(m!)^2}{((m-c)!)^2 c!} \left(1 + \delta_{n,m,c}\right) \zeta_c(\eta),\end{aligned}$$

which is exactly (3.11). □

4 Asymptotic comparison of U- and V-statistics

Theorem 4.1. *If $q \in [1, \infty)$ and $\max_{i \in \mathbb{N}_n^m} \mathrm{E}\left[\left\|\eta\left(Z_{i(1)}, \ldots, Z_{i(m)}\right)\right\|^q\right] < \infty$,*

$$\mathrm{E}\left[\|U_n(\eta) - V_n(\eta)\|^q\right]^{\frac{1}{q}} \leq \frac{m(m-1)}{n} \max_{i \in \mathbb{N}_n^m} \mathrm{E}\left[\left\|\eta\left(Z_{i(1)}, \ldots, Z_{i(m)}\right)\right\|^q\right]^{\frac{1}{q}}. \tag{4.1}$$

Proof of Theorem 4.1. Apply Lemma 4.1:

$$\mathrm{E}\left[\|U_n(\eta) - V_n(\eta)\|^q\right]^{\frac{1}{q}} \leq \frac{2\left(n^m - \gamma_{n,m}\right)}{n^m} \max_{i \in \mathbb{N}_n^m} \mathrm{E}\left[\left\|\eta\left(Z_{i(1)}, \ldots, Z_{i(m)}\right)\right\|^q\right]^{\frac{1}{q}}.$$

By Lemma 4.2, $n^m - \gamma_{n,m} = (m(m-1)/2)n^{m-1}$ so that $\left(2\left(n^m - \gamma_{n,m}\right)\right)/n^m = (m(m-1))/n$. Hence (4.1) follows. □

Lemma 4.1. *Let $\mathcal{I}$ be a non-empty finite set and $\{a_\iota : \iota \in \mathcal{I}\}$ be a subset of a normed space with norm $\|\cdot\|$. Let $\mathcal{I}_0 \subseteq \mathcal{I}$ be a strict subset of $\mathcal{I}$, i.e. $|\mathcal{I} \setminus \mathcal{I}_0| \geq 1$. Define*

$$V = \frac{1}{|\mathcal{I}|} \sum_{\iota \in \mathcal{I}} a_\iota, \qquad U = \frac{1}{|\mathcal{I}_0|} \sum_{\iota \in \mathcal{I}_0} a_\iota, \qquad W = \frac{1}{|\mathcal{I}| - |\mathcal{I}_0|} \sum_{\iota \in \mathcal{I} \setminus \mathcal{I}_0} a_\iota. \tag{4.2}$$

Then, the following hold

$$V - U = \frac{|\mathcal{I}| - |\mathcal{I}_0|}{|\mathcal{I}|}(W - U), \tag{4.3}$$

$$\|V - U\| \leq 2\frac{|\mathcal{I}| - |\mathcal{I}_0|}{|\mathcal{I}|} \max_{\iota \in \mathcal{I}} \|a_\iota\| . \tag{4.4}$$

Proof of Lemma 4.1. From (4.2), $|\mathcal{I}|V = \sum_{\iota \in \mathcal{I}_0} a_\iota + \sum_{\iota \in \mathcal{I} \setminus \mathcal{I}_0} a_\iota = |\mathcal{I}_0| U + (|\mathcal{I}| - |\mathcal{I}_0|) W$. Subtract $|\mathcal{I}|U$ from both sides, so that $|\mathcal{I}|(V - U) = (|\mathcal{I}| - |\mathcal{I}_0|)(W - U)$. Division by $|\mathcal{I}|$ establishes (4.3). For (4.4), by (4.3) and the triangle inequality,

$$\begin{aligned} \|U - V\| &= \frac{|\mathcal{I}| - |\mathcal{I}_0|}{|\mathcal{I}|}\|U - W\| \leq \frac{|\mathcal{I}| - |\mathcal{I}_0|}{|\mathcal{I}|}(\|U\| + \|W\|) \\ &\leq 2\frac{|\mathcal{I}| - |\mathcal{I}_0|}{|\mathcal{I}|} \max\{\|U\|, \|W\|\}. \end{aligned}$$

Since U and W are arithmetic averages over subsets of $\mathcal{I}$, we get $\max\{\|U\|, \|W\|\} \leq \max_{\iota \in \mathcal{I}} \|a_\iota\|$ from repeated applications of the triangle inequality. This establishes (4.4). $\square$

Lemma 4.2. *For $n, m \in \mathbb{N}$, with $n \geq m$, with $\gamma_{n,m}$ as in (1.2),*

$$n^m - \gamma_{n,m} = \frac{m(m-1)}{2} n^{m-1}. \tag{4.5}$$

Proof of Lemma 4.2. For $m = 1$, (4.5) follows from $\gamma_{n,1} = n$ by definition in (1.2). For $m = 2$, by definition in (1.2), $\gamma_{n,2} = n(n-1)$. Thus

$$n^2 - \gamma_{n,2} = n^2 - n(n-1) = n,$$

so that (4.5) follows for $m = 2$.

Now we verify (4.5) for $m > 2$. Note that

$$n^m - \gamma_{n,m} = \left|\mathrm{NonInj}_{n,m}\right| ,$$

where as in (1.1),

$$\mathrm{NonInj}_{n,m} = \{i \in \mathbb{N}_n^m : \exists j,k \in \mathbb{N}_m, j \neq k \text{ such that } i(j) = i(k)\}.$$

Now consider the task of picking an element from $\mathrm{NonInj}_{n,m}$. Recall that we are considering the cases $m > 2$. We can exhaust $\mathrm{NonInj}_{n,m}$ by using the following steps:

1. Pick two indices $j, k \in \mathbb{N}_m$ such that $j < k$;

2. Select $i(j) \in \mathbb{N}_n$ and set $i(k) = i(j)$;

3. Pick the restriction $\{i(l) : l \in \mathbb{N}_m \setminus \{j,k\}\} \in \mathbb{N}_n^{m-2}$ arbitrarily.

There are $m(m-1)/2$ ways to complete step 1. For each instance of step 1, there are n ways to complete step 2. Finally, for each instance of steps 1 and 2, there are n^{m-2} ways to complete step 3. Therefore,

$$n - \gamma_{n,m} = \left|\mathrm{NonInj}_{n,m}\right| = \frac{m(m-1)}{2} \cdot n^{m-1}.$$

□

References

Billingsley, Patrick. 1995. *Probability and Measure.* Wiley Series in Probability and Statistics. Wiley.

A Preliminaries on projections

Denote the set of random variables that are square-integrable against Pr by

$$\mathscr{L}_2 \equiv \mathscr{L}_2(\Pr) = \left\{ X : \mathrm{E}\left[|X|^2\right] = \int |X|^2 \mathrm{d}\Pr < \infty \right\}. \tag{A.1}$$

It is well known that $\mathscr{L}_2$ is a Hilbert space under the inner product $\langle X, Y \rangle = \mathrm{E}[X \cdot Y]$. That is, the inner product space $(\mathscr{L}_2, \langle \cdot, \cdot \rangle)$ is a complete metric space under the induced normed metric:

$$\rho_2 (X, Y)^2 := \|X - Y\|_2^2 := \langle X - Y, X - Y \rangle = \mathrm{E}\left[|X - Y|^2\right] \quad \forall\, X, Y \in \mathscr{L}_2. \tag{A.2}$$

Let $\mathcal{S} \subseteq \mathscr{L}_2(\Pr)$ and $T \in \mathscr{L}_2$. $T_* \in \mathcal{S}$ is a projection of T onto $\mathcal{S}$ if and only if

$$T_* \in \mathcal{S} \quad \text{and} \quad \mathrm{E}\left[(T - T_*)^2\right] \leq \mathrm{E}\left[(T - S)^2\right] \quad \forall S \in \mathcal{S}. \tag{A.3}$$

A.1 Existence of projections onto closed and convex subsets

Theorem A.1 shows that a projection on $\mathcal{S} \subset \mathscr{L}_2$ exists whenever $\mathcal{S}$ is closed and convex.

Theorem A.1. *Let $T \in \mathscr{L}_2$ and let $\mathcal{S} \subseteq \mathscr{L}_2$ be closed and convex. Then there exists T_* satisfying (A.3). Furthermore, T_* is almost surely unique, i.e. if T_{**} also satisfies (A.3), then* $\Pr\{T_* \neq T_{**}\} = 0$.

Proof of Theorem A.1. Let

$$d_* := \inf_{S \in \mathcal{S}} \mathrm{E}\left[(S - T)^2\right].$$

Then for each $n \in \mathbb{N}$, there necessarily exists $S_n \in \mathcal{S}$ such that

$$d_* \leq \mathrm{E}\left[(S_n - T)^2\right] \leq d_* + \frac{1}{n}.$$

Furthermore for any $n, m \in \mathbb{N}$,

$$\begin{aligned}
\mathrm{E}\left[(S_n - S_m)^2\right] &= \mathrm{E}\left[(S_n - T)^2\right] + \mathrm{E}\left[(S_m - T)^2\right] - 2\mathrm{E}\left[(S_n - T)(S_m - T)\right], \\
4\mathrm{E}\left[\left(\frac{S_n + S_m}{2} - T\right)^2\right] &= \mathrm{E}\left[(S_n - T)^2\right] + \mathrm{E}\left[(S_m - T)^2\right] + 2\mathrm{E}\left[(S_n - T)(S_m - T)\right].
\end{aligned}$$

This implies

$$\mathrm{E}\left[(S_n - S_m)^2\right] = 2\mathrm{E}\left[(S_n - T)^2\right] + 2\mathrm{E}\left[(S_m - T)^2\right] - 4\mathrm{E}\left[\left(\frac{S_n + S_m}{2} - T\right)^2\right].$$

Since $\mathcal{S}$ is convex, $\frac{1}{2}(S_n + S_m) \in \mathcal{S}$ and so, $\mathrm{E}\left[\left\{\frac{1}{2}(S_n + S_m) - T\right\}^2\right] \geq d_*$. Using the upper bounds in the definition of S_n and S_m, and the lower bound now established,

$$\mathrm{E}\left[(S_n - S_m)^2\right] \leq 2\left(d_* + \frac{1}{n}\right) + 2\left(d_* + \frac{1}{m}\right) - 4d_* = 2\left(\frac{1}{n} + \frac{1}{m}\right).$$

Hence, $\{S_n\}$ is Cauchy in $\mathcal{S}$. Since $\mathscr{L}_2$ is complete and $\mathcal{S}$ is a closed subset of $\mathscr{L}_2$, $\mathcal{S}$ is also necessarily complete. Thus, there is a $T_* \in \mathcal{S}$ such that $\lim_{n\to\infty} \mathrm{E}\left[(S_n - T_*)^2\right] = 0$. Then,

$$\begin{aligned}
d_* &\leq \mathrm{E}\left[(T_* - T)^2\right] = \mathrm{E}\left[(S_n - T)^2\right] + \mathrm{E}\left[(T_* - S_n)^2\right] + 2\mathrm{E}\left[(S_n - T)\cdot(T_* - S_n)\right] \\
&\leq \mathrm{E}\left[(S_n - T)^2\right] + \mathrm{E}\left[(T_* - S_n)^2\right] + 2\mathrm{E}\left[(S_n - T)^2\right]^{\frac{1}{2}} \mathrm{E}\left[(T_* - S_n)^2\right]^{\frac{1}{2}}.
\end{aligned}$$

Hence,

$$d_* \leq \mathrm{E}\left[(T_* - T)^2\right] \leq d_* + \frac{1}{n} + \mathrm{E}\left[(T_* - S_n)^2\right] + 2\left(d_* + \frac{1}{n}\right)^{\frac{1}{2}} \mathrm{E}\left[(T_* - S_n)^2\right]^{\frac{1}{2}}.$$

Since $\mathrm{E}\left[(T_* - S_n)^2\right] \to 0$, taking limits, $\mathrm{E}\left[(T_* - T)^2\right] = d_* = \inf_{S\in\mathcal{S}} \mathrm{E}\left[(S - T)^2\right]$.

Next suppose T_{**} also satisfies (A.3). Then as before,

$$\mathrm{E}\left[(T_* - T_{**})^2\right] = 2\mathrm{E}\left[(T_* - T)^2\right] + 2\mathrm{E}\left[(T_{**} - T)^2\right] - 4\mathrm{E}\left[\left(\frac{T_* + T_{**}}{2} - T\right)^2\right].$$

Since T_* and T_{**} both satisfy (A.3),

$$\mathrm{E}\left[(T_* - T_{**})^2\right] = 4d_* - 4\mathrm{E}\left[\left(\frac{T_* + T_{**}}{2} - T\right)^2\right].$$

Since $\mathcal{S}$ is convex, $\frac{T_*+T_{**}}{2} \in \mathcal{S}$ and so,

$$\mathrm{E}\left[\left(\frac{T_* + T_{**}}{2} - T\right)^2\right] \geq d_*.$$

Hence,

$$0 \leq \mathrm{E}\left[(T_* - T_{**})^2\right] \leq 4d_* - 4d_* = 0.$$

The above implies that $\Pr\{T_* \neq T_{**}\} = 0$. □

A.2 Projection onto linear subspaces and orthogonality

When the subset for projection is a linear subspace of $\mathscr{L}_2$, a useful equivalent characterization of a projection is through an orthogonality condition.

Theorem A.2. *Let $T \in \mathscr{L}_2$ and $\mathcal{S}$ be a linear subspace of $\mathscr{L}_2$. Then T_* satisfies (A.3) if and only if*

$$T_* \in \mathcal{S} \quad \textit{and} \quad \mathrm{E}\left[S \cdot (T - T_*)\right] = 0 \quad \forall S \in \mathcal{S}. \tag{A.4}$$

Furthermore, T_* satisfying (A.4) is almost surely unique. That is, if T_{**} also satisfies (A.4), then $\Pr\{T_* \neq T_{**}\} = 0$.

Proof of Theorem A.2. Suppose T_* satisfies (A.4). Then, for any $S \in \mathcal{S}$,

$$\begin{aligned}\mathrm{E}\left[(T - S)^2\right] &= \mathrm{E}\left[(T - T_*)^2\right] + 2\mathrm{E}\left[(T - T_*) \cdot (T_* - S)\right] + \mathrm{E}\left[(T_* - S)^2\right] \\ &= \mathrm{E}\left[(T - T_*)^2\right] + \mathrm{E}\left[(T_* - S)^2\right],\end{aligned}$$

where the final equality follows from (A.4) since $T_* - S \in \mathcal{S}$ by linearity of $\mathcal{S}$. Therefore, since $\mathrm{E}\left[(T_* - S)^2\right] \geq 0$, T_* satisfies (A.3).

Next, we prove the converse by proving its contrapositive. To that end, suppose that (A.4) fails. That is, $T_* \notin \mathcal{S}$ or there exists $S_* \in \mathcal{S}$ such that

$$\mathrm{E}\left[S_* \cdot (T - T_*)\right] \neq 0. \tag{A.5}$$

If $T_* \notin \mathcal{S}$, then it is immediate that (A.3) fails. Thus, suppose that $T_*, S_* \in \mathcal{S}$ and (A.5) holds. Note that (A.5) necessitates $\Pr\left\{S_* \neq 0\right\} > 0$, i.e. S_* cannot be almost surely zero. Since $\mathcal{S} \subseteq \mathscr{L}_2$, this also implies that $0 < \mathrm{E}\left[S_*^2\right] < \infty$. By linearity of $\mathcal{S}$, $T_* + \tau S_* \in \mathcal{S}$ for every $\tau \in \mathbb{R}$, and

$$\mathrm{E}\left[(T - T_* - \tau S_*)^2\right] = \mathrm{E}\left[(T - T_*)^2\right] - 2\tau \mathrm{E}\left[S_* \cdot (T - T_*)\right] + \tau^2 \mathrm{E}\left[S_*^2\right].$$

As a function of τ, the above is minimized at

$$\tau_* = \frac{\mathrm{E}\left[S_* \cdot (T - T_*)\right]}{\mathrm{E}\left[S_*^2\right]} \neq 0.$$

The associated minimum value is

$$\mathrm{E}\left[(T - T_* - \tau_* S_*)^2\right] = \mathrm{E}\left[(T - T_*)^2\right] - \frac{\mathrm{E}\left[S_* \cdot (T - T_*)\right]^2}{\mathrm{E}\left[S_*^2\right]} < \mathrm{E}\left[(T - T_*)^2\right].$$

This means that (A.3) fails. By contrapositive, we have that if (A.3) holds, then (A.4) must also hold.

Finally, to see that T_* is almost surely unique, let T_{**} also satisfy (A.4). Then,

$$\mathrm{E}\left[(T_* - T_{**})^2\right] = \mathrm{E}\left[(T_* - T_{**}) \cdot (T_* - T)\right] + \mathrm{E}\left[(T_* - T_{**}) \cdot (T - T_{**})\right] = 0,$$

since by linearity of $\mathcal{S}$, $T_* - T_{**} \in \mathcal{S}$. The above implies that $\Pr\left\{T_* \neq T_{**}\right\} = 0$. $\square$

A.3 Second-moment distance between a random variable and its projection

The following result provides a useful characterization of the second moment distance between a random variable and its projection onto a space that contains the constants.

Theorem A.3. *Let $T \in \mathscr{L}_2$ and $\mathcal{S}$ be a linear subspace of $\mathscr{L}_2$ containing the constants (i.e. non-stochastic random variables). Suppose the projection of T onto $\mathcal{S}$ exists and denote this projection by T_*. Then,*

$$\mathrm{E}\left[(T - T_*)^2\right] = \mathrm{Var}[T] - \mathrm{Var}\left[T_*\right]. \tag{A.6}$$

Proof of Theorem A.3. By $T_* \in \mathcal{S}$ and (A.4),

$$\mathrm{E}\left[T_* \cdot (T - T_*)\right] = 0 \iff \mathrm{E}\left[T \cdot T_*\right] = \mathrm{E}\left[T_*^2\right]. \tag{A.7}$$

Similarly, since $\mathcal{S}$ contains the constants,

$$\mathrm{E}\left[T - T_*\right] = 0 \iff \mathrm{E}[T] = \mathrm{E}\left[T_*\right]. \tag{A.8}$$

Therefore, by (A.7) and (A.8),

$$\mathrm{Cov}\left(T, T_*\right) = \mathrm{E}\left[T T_*\right] - \mathrm{E}[T]\mathrm{E}\left[T_*\right] = \mathrm{E}\left[T_*^2\right] - \mathrm{E}\left[T_*\right]^2.$$

Hence,

$$\mathrm{Cov}\left(T, T_*\right) = \mathrm{Var}\left[T_*\right]. \tag{A.9}$$

Next, by (A.8), $\mathrm{E}\left[T - T_*\right] = 0$ and so,

$$\mathrm{E}\left[(T - T_*)^2\right] = \mathrm{Var}\left[T - T_*\right] = \mathrm{Var}[T] + \mathrm{Var}\left[T_*\right] - 2\mathrm{Cov}\left(T, T_*\right).$$

Therefore by (A.9),

$$\mathrm{E}\left[(T-T_*)^2\right]=\mathrm{Var}[T]-\mathrm{Var}\left[T_*\right], \tag{A.10}$$

which is exactly (A.6). $\square$

A.4 Projection onto sum spaces

Suppose now that $\{Z_j\}_{j=1}^n$ are independent random variables/vectors and let

$$\mathcal{S}=\left\{\sum_{j=1}^n g_j\left(Z_j\right): g_j \text{ measurable and } \mathrm{E}\left[g_j\left(Z_j\right)^2\right]<\infty \text{ for all } j=1,\ldots,n\right\}. \tag{A.11}$$

Theorem A.4. *Let $\{Z_j\}_{j=1}^n$ be independent random variables. The projection of an arbitrary random variable $T\in\mathscr{L}_2$ onto the class $\mathcal{S}$ in (A.11) is given by*

$$T_*=\sum_{j=1}^n \mathrm{E}\left[T|Z_j\right]-(n-1)\mathrm{E}[T]=\mathrm{E}[T]+\sum_{j=1}^n\left\{\mathrm{E}\left[T|Z_j\right]-\mathrm{E}[T]\right\}.$$

Proof of Theorem A.4. We proceed by verifying the orthogonality condition (A.4). Let $j\in\{1,\ldots,n\}$ be given. If $g_j(\cdot)$ is a measurable function with $\mathrm{E}\left[g_j\left(Z_j\right)^2\right]<\infty$, then certainly $g_j\left(Z_j\right)\in\mathcal{S}$. Furthermore,

$$\begin{aligned}
\mathrm{E}\left[g_j\left(Z_j\right)\cdot(T-T_*)\right] &= \mathrm{E}\left[g_j\left(Z_j\right)\cdot\left\{T-\mathrm{E}[T]-\mathrm{E}\left[T|Z_j\right]+\mathrm{E}[T]\right\}\right] \\
&\quad -\sum_{\substack{l=1\\ l\neq j}}^n \mathrm{E}\left[g_j\left(Z_j\right)\cdot\left\{\mathrm{E}\left[T|Z_l\right]-\mathrm{E}[T]\right\}\right] \\
&= \mathrm{E}\left[g_j\left(Z_j\right)\cdot\left\{T-\mathrm{E}\left[T|Z_j\right]\right\}\right] \\
[\text{by } Z_j \perp\!\!\!\perp Z_l] \qquad &\quad -\sum_{\substack{l=1\\ i\neq j}}^n \mathrm{E}\left[g_j\left(Z_j\right)\right]\cdot\mathrm{E}\left[\left\{\mathrm{E}\left[T|Z_l\right]-\mathrm{E}[T]\right\}\right] \\
&= 0.
\end{aligned}$$

In the last line above, the second term is zero since $\mathrm{E}\left[\mathrm{E}\left[T|Z_l\right]\right]=\mathrm{E}[T]$ for every l and the first term is zero by definition of $\mathrm{E}[T|Z_j]$. Now, take a sum of such functions $g_j(Z_j)$ over $j = 1, \ldots, n$:

$$\mathrm{E}\left[\left(\sum_{j=1}^{n} g_j(Z_j)\right) \cdot (T - T_*)\right] = \sum_{j=1}^{n} \mathrm{E}\left[g_j(Z_j) \cdot (T - T_*)\right] = \sum_{j=1}^{n} 0 = 0.$$

$\square$

B Exercises in counting

Theorem B.1. *Let m, p, n be natural numbers and c be a non-negative integer such that $c \le \min\{m, p\}$ and $\max\{m, p\} \le n$. Let $a = \{a_i\}_{i=1}^{m}, b = \{b_i\}_{i=1}^{p} \subseteq \{1, \ldots, n\}$ where $i \mapsto a_i$ and $i \mapsto b_i$ are both strictly increasing. There are $\binom{n}{m} \cdot \binom{m}{c} \cdot \binom{n-m}{p-c}$ ways to select a and b such that they have exactly c elements in common.*

Proof of Theorem B.1. Let C denote the number of ways to pick a and b with exactly c common elements. We can select a and b with exactly c common elements by first selecting a, then selecting c elements of a that will be common with b, and finally selecting b to ensure only those c elements are common to a and b. Therefore,

$$\begin{aligned} C = {} & \#\text{ of ways to select } a \\ & \times \#\text{ of ways to select } c \text{ elements of } a \text{ in common with } b \\ & \times \#\text{ of ways to select remaining elements of } b \end{aligned}$$

Clearly,

$$\begin{aligned} \#\text{ of ways to select } a &= \binom{n}{m}, \\ \#\text{ of ways to select } c \text{ elements of } a &= \binom{m}{c}. \end{aligned}$$

Then, consider the last problem of selecting the remaining elements of b that are not common with a. Since the c common elements have already been selected, there are $p - c$ remaining elements of b requiring selection. Since we have fulfilled the "exactly c elements in common" constraint, this selection must be from $\{1, \ldots, n\} \setminus a$, which has cardinality $n - m$. Therefore, upon fixing a and the c common elements,

$$\# \text{ of ways to select remaining elements of } b = \binom{n-m}{p-c}.$$

$\square$