U Statistics

John Duchi

Stats 300b – Winter Quarter 2021

Outline

- Definitions and motivation
- Examples
- Two-sample U-statistics
- Variance decompositions
- Analysis: projections in Hilbert spaces and conditional expectations
- Hájek projections and asymptotic normality

Reading:

- van der Vaart, *Asymptotic Statistics*, Chapters 11–12, 14.1 (again)

The problem

- have a (permutation) symmetric function $h : \mathcal{X}^k \to \mathbb{R}$
- wish to estimate

$$\theta := \mathbb{E}[h(X_1, \ldots, X_k)]$$

Example (Some measures of dispersion)

- variance with $h(y, z) = \frac{1}{2}(y - z)^2$

$$\theta = \mathsf{Var}(X) = \frac{1}{2}\mathbb{E}[(X_1 - X_2)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2,$$

- dispersion probabilities

$$\theta = \mathbb{P}(X_1 \geq X_2 + t), \quad h(y, z) = \frac{1}{2}1\{y \geq z + t\} + \frac{1}{2}1\{z \geq y + t\}$$

U-statistics

Definition (U-statistic)

Let $h : \mathcal{X}^r \to \mathbb{R}$ be a symmetric function (called the *kernel*) and $X_i \overset{\text{iid}}{\sim} P$. Define $\theta(P) := \mathbb{E}_P[h(X_1, \ldots, X_r)]$. Then

$$U_n := \binom{n}{r}^{-1} \sum_{|\beta|=r, \beta \subset [n]} h(X_\beta)$$

is the associated *U-statistic of order r*, β ranging over (ordered) subsets of $[n] = \{1, \ldots, n\}$.

- unbiased estimator as $\mathbb{E}[U_n] = \mathbb{E}[h(X_1, \ldots, X_r) = \theta(P)$.

Why U-statistics?

- we could just use

$$\frac{1}{n/r}\sum_{l=1}^{n/r} h(X_{l(r-1)+1}, \ldots, X_{lr})$$

- consider instead un-ordered sample $\{X_{(1)}, \ldots, X_{(n)}\}$ (order statistics, histogram, or type)

Rao-Blackwellization: statistic $\{X_{(1)}, \ldots, X_{(n)}\}$ is sufficient

Observation (Rao-Blackwell)

If L is convex and $X_i \stackrel{\text{iid}}{\sim} P$, then

$$U_n = \mathbb{E}[h(X_1, \ldots, X_r) \mid X_{(1)}, \ldots, X_{(n)}]$$

and for any $T_n = \text{Conv}\{h(X_\beta) \mid |\beta| = r, \beta \subset [n]\}$,

$$\mathbb{E}[L(U_n - \theta(P))] \le \mathbb{E}[L(T_n - \theta(P))]$$

Examples of U-statistics

- Variance: for $h(x, y) = \frac{1}{2}(x - y)^2$,

$$U_n = \binom{n}{2}^{-1} \sum_{i<j} h(X_i, X_j) = \frac{1}{2n(n-1)} \sum_{i,j=1}^{n} (X_i - X_j)^2$$
$$= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

- Gini's mean difference: $h(x, y) = |x - y|$, and

$$\theta = \mathbb{E}[|X_1 - X_2|]$$

- Quantiles: $r = 1$, and $h(x) = 1\{x \leq t\}$, $\theta = P(X \leq t)$
- Signed rank statistic (gives location information): $h(x, y) = 1\{x + y > 0\}$, so $r = 2$ and

$$\theta(P) = P(X_1 + X_2 > 0)$$

Two-sample U-statistics

- samples $\{X_1, \ldots, X_m\}$ and $\{Y_1, \ldots, Y_m\}$
- function h symmetric in first r and second s inputs (not across all)

$$U := \binom{m}{r}^{-1} \binom{n}{s}^{-1} \sum_{|\alpha|=r, |\beta|=s} h(X_\alpha, Y_\beta)$$

Example (Mann-Whitney statistic)

To test a difference in location, use $h(x, y) = 1\{x \le y\}$,

$$U := \frac{1}{nm} \sum_{i,j} 1\{X_i \le Y_j\}$$

with null $H_0 : P(X \le Y) = \frac{1}{2}$

Variance of U-statistics (Hoeffding)

idea: project out low order terms and use what's left for asymptotics

- for h an order r kernel, for $c < r$ define

$$\begin{aligned} h_c(x_1, \dots, x_c) &:= \mathbb{E}[h(x_1, \dots, x_c, X_{c+1}, \dots, X_r)] \\ &= \mathbb{E}[h(X_1^c, X_{c+1}^r) \mid X_1^c = x_1^c] \end{aligned}$$

- notice $h_0 = \mathbb{E}[h(X_1^r)] = \theta(P)$ and $\mathbb{E}[h_c(X_1^c)] = \theta(P)$ for all c
- centered statistics $\widehat{h}_c = h_c - \theta(P)$

Variance of U-statistics

Define "intersection" covariances

$$\zeta_c := \mathbb{E}[\widehat{h}(X_A)\widehat{h}(X_B)]$$

for $A, B \subset [n]$ with $|A| = |B| = r$ and $|A \cap B| = c \leq r$

Observation

These covariances satisfy $\zeta_c = \mathbb{E}[\widehat{h}_c(X_1^c)^2]$

The variance of a U-statistic (finally)

Proposition

Let h be a kernel of order r and $U_n = \binom{n}{r}^{-1} \sum_{|\beta|=r} h(X_\beta)$. Then

$$\mathrm{Var}(U_n) = \frac{r^2}{n}\zeta_1 + O(n^{-2}).$$

Hilbert spaces

Definition

A vector space $\mathcal{H}$ is a *Hilbert space* if is a complete normed space with inner product $\langle \cdot, \cdot \rangle$, where

$$\|u\|^2 = \langle u, u \rangle \text{ and } \langle x + y, u + v \rangle = \langle x, u \rangle + \langle x, v \rangle + \langle y, u \rangle + \langle y, v \rangle$$

Example (Standard Hilbert spaces)

- $\mathbb{R}^n$ with standard inner product $\langle x, y \rangle = x^T y = \sum_{i=1}^n x_i y_i$
- $L^2(P) = \{f : \mathcal{X} \to \mathbb{R} \text{ with } \int f(x)^2 dP(x) < \infty\}$, with inner product $\langle f, g \rangle = \int fgdP$

Projections

- Let $\mathcal{S}$ be a closed linear subspace of $\mathcal{H}$
- For $v \in \mathcal{H}$, define the projection

$$\pi_{\mathcal{S}}(v) := \underset{s \in \mathcal{S}}{\operatorname{argmin}} \|v - s\|^2$$

Theorem

The projection $\pi_{\mathcal{S}}(v)$ exists and is uniquely defined by

$$\langle v - \pi_{\mathcal{S}}(v), s \rangle = 0 \quad \textit{for all } s \in \mathcal{S}.$$

Projection examples

Example (Collections of random variables)

In $L^2(P)$, let $\mathcal{S}$ be a collection of random variables (functions) with $\mathbb{E}[S^2] < \infty$, closed under linear combinations. Then $\widehat{S}$ is the projection of a random variable T onto $\mathcal{S}$ if and only if

$$\mathbb{E}[(T - \widehat{S})S] = 0 \quad \text{for all } S \in \mathcal{S}.$$

Proposition (Moreau decomposition)

For any $v \in \mathcal{H}$, $\|v\|^2 = \|\pi(v)\|^2 + \|v - \pi(v)\|^2$.

Conditional expectations as projections in $L^2(P)$

Let Y be a random variable and

$$\mathcal{S} = \{\text{Linear span of } \; g(Y) \text{ s.t. } g \text{ measurable}, \; Pg(Y)^2 < \infty\}$$

Definition

The conditional expectation of $X \in L^2(P)$ given Y is

$$\mathbb{E}[X \mid Y] := \text{Projection of } X \text{ onto } \mathcal{S}.$$

- $\mathbb{E}[X \mid Y]$ is random variable (function of Y) with

$$\mathbb{E}[(X - \mathbb{E}[X \mid Y])g(Y)] = 0 \quad \text{all } g \text{ integrable}$$

Consequences of projections

- iterating expectations: $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]$
- product properties: for all f,

$$\mathbb{E}[f(Y)X \mid Y] = f(Y)\mathbb{E}[X \mid Y]$$

- tower property: $\mathbb{E}[\mathbb{E}[X \mid Y, Z] \mid Y] = \mathbb{E}[X \mid Y]$

Why projections?

- ignore smaller-order terms!

Proposition

Let $\mathcal{S}_n$ be a sequences of subspaces of $L^2(P)$ and T_n be random variables. Let $S_n = \pi_{\mathcal{S}_n}(T_n)$. If $\mathrm{Var}(T_n)/\mathrm{Var}(S_n) \to 1$, *then*

$$\frac{T_n - \mathbb{E}[T_n]}{\sqrt{\mathrm{Var}(T_n)}} - \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}} \xrightarrow{p} 0.$$

Hájek Projections

idea: project U-statistic U_n onto independent sums

Let $X_1, \dots, X_n$ be independent and

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

Lemma

Let $\mathbb{E}[T^2] < \infty$. *Then*

$$\widehat{S} = \pi_{\mathcal{S}}(T) = \sum_{i=1}^{n} \mathbb{E}[T \mid X_i] - (n-1)\mathbb{E}[T]$$

Projecting U_n onto i.i.d. sums

- recall $h_1(x) = \mathbb{E}[h(x, X_2^r)] - \theta$

Lemma (Hájek projection)

If

$$\widehat{U}_n = \mathit{Projection}\left(U_n - \theta \text{ onto } \mathcal{S} = \left\{ \sum_{i=1}^{n} g_i(X_i) \right\} \right)$$

then

$$\widehat{U}_n = \sum_{i=1}^{n} \mathbb{E}[U_n - \theta \mid X_i] = \frac{r}{n} \sum_{i=1}^{n} h_1(X_i)$$

Asymptotic normality of U-statistics

Theorem

Let h be an order r symmetric kernel with $\zeta_1 = \mathbb{E}[h_1(X)^2]$. Then

$$\sqrt{n}(U_n - \theta - \widehat{U}_n) \xrightarrow{p} 0 \quad \textit{and} \quad \sqrt{n}(U_n - \theta) \xrightarrow{d} \mathcal{N}(0, r^2\zeta_1).$$

Projection of two-sample U-statistics

$$U_N = \binom{m}{r}^{-1}\binom{n}{s}^{-1} \sum_{|\alpha|=r} \sum_{|\beta=s|} h(X_\alpha, Y_\beta)$$

where for $N = m + n$, have $\frac{m}{N} \to \lambda \in (0, 1)$

Lemma

Define

$$h_{1,0}(x) = \mathbb{E}[h(x, X_2^r, Y_1^s)] - \theta$$
$$h_{0,1}(x) = \mathbb{E}[h(X_1^r, y, Y_2^s)] - \theta.$$

Then the projection $\widehat{U}_N$ *onto* $\sum_{i=1}^m f_i(X_i) + \sum_{j=1}^n g_j(Y_j)$ *is*

$$\widehat{U}_N = \frac{r}{m} \sum_{i=1}^m h_{1,0}(X_i) + \frac{s}{n} \sum_{i=1}^n h_{0,1}(Y_i)$$

Asymptotic normality of two-sample U-statistics

Theorem

Let $\mathbb{E}[h^2(X_1^r, Y_1^s)] < \infty$. *Then*

$$\sqrt{N}(U_N - \theta - \widehat{U}_N) \xrightarrow{p} 0 \text{ and } \sqrt{N}(U_N - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{r^2\zeta_{1,0}}{\lambda} + \frac{s^2\zeta_{0,1}}{1-\lambda}\right)$$

where

$$\zeta_{c,d} = \mathrm{Cov}\left(h(X_1^r, Y_1^s), h(X_1^c, X_{c+1}', \ldots, X_r', Y_1^d, Y_{d+1}', \ldots, Y_s')\right).$$

Pivotal quantities (robustness of U-statistics)

- often U-statistic has asymptotically pivotal distribution under null H_0

Example (Signed rank)

$U_n = \binom{n}{2}^{-1} \sum_{i<j} 1\{X_i + X_j > 0\}$ is pivotal under null H_0 that X has continuous cdf F symmetric about 0.

Example: the Mann-Whitney statistic

- test "stochastically larger" statistic $\theta = P(X \leq Y)$ (or for a difference in location)
- kernel $h(x, y) = 1\{x \leq y\}$

Example (Mann-Whitney)

Under $H_0 : X \stackrel{\text{dist}}{=} Y$ with continuous distribution,

$$U_N := \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} 1\{X_i \leq Y_j\}$$

satisfies $\sqrt{12mn/N}(U_N - \frac{1}{2}) \xrightarrow{d} \mathcal{N}(0, 1)$