Topics in Probability Theory **Fall 2023 (Georgia Institute of Technology)**

Lecture 6: Decoupling Inequalities for Generalized U-Statistics

Lecturer: ***Víctor H. de la Peña*** *Scribes: Heyuan Yao*

6.1 U-statistics

Let $X_1, ..., X_n$ be a random sample (i.i.d. observations) from an unknown distribution F in $\mathbb{R}$. Given a known function $h : \mathbb{R}^m \to \mathbb{R}$, consider the estimation of the "parameter"

$$\theta = \theta(F) = \mathbb{E}[h(X_1, ..., X_m)],$$

Of course, you may interested in more complex spaces, which the random variables live in or h maps to, but now let us think about the simpler case.

A natural unbiased estimator of θ you propose is $h(X_1, ..., X_m)$, and since n observations (with $n \geq m$) are available, this simple estimator can be improved. Now you decide to get the average of $h(X_{\alpha_1}, ..., X_{\alpha_m})$, where $(X_{\alpha_1}, ..., X_{\alpha_m}) \in \Pi$, the set of all permutations of m integers such that

$$1 \leq \alpha_i \leq n, \quad \alpha_i \neq \alpha_j \;\; if \;\; i \neq j, \;\; (i, j = 1, ...m).$$

Congratulations! You successfully construct a U-Statistic, which in this context is defined by

$$U_n = U(X_1, ..., X_n) = \frac{1}{n(n-1)...(n-m+1)} \sum_{(X_{\alpha_1}, ..., X_{\alpha_m}) \in \Pi} h(X_{\alpha_1}, ..., X_{\alpha_m}). \tag{6.1}$$

If h is permutation invariant[1], the definition (6.1) is equivalent to

$$U_n = \frac{1}{\binom{n}{m}} \sum_{1 \leq X_{\alpha_1} < ... < X_{\alpha_m} \leq n} h(X_{\alpha_1}, ..., X_{\alpha_m}) \tag{6.2}$$

Although it may be the first time you hear U-Statistics, you have played with it for a long time. Look at equation (6.2), then set $h : \mathbb{R}^2 \to \mathbb{R}$ be such that $h(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$, you can verify that U_n is exactly

[1] h is permutation invariant if the value of $h(\boldsymbol{x})$ does not change if we permute the components of $\boldsymbol{x}$, i.e., for instance, when $r = 3$: $h((x_1, x_2, x_3)) = h((x_2, x_1, x_3)) = h((x_3, x_1, x_2)) = h((x_1, x_3, x_1)) = h((x_2, x_3, x_1)) = h((x_3, x_2, x_1))$.

twice the sample variance, i.e.,

$$s_n^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{n-1} = \frac{\sum_{1\le i<j\le n} \frac{1}{2}(X_i - X_j)^2}{\binom{n}{2}} = \frac{\sum_{1\le i<j\le n}(X_i - X_j)^2}{n(n-1)}.$$

And by doing so, it is convenient to show that s_n^2 is an unbiasedness estimator. This is why we call such estimators U-statistics: the letter "U" stands for unbiased.

There are several examples of U-statistics. The sample mean is definitely a U-statistic. And when $X_1 \sim \mu \neq \delta_0$ is nonnegative a.s., the sample Gini mean difference (GMD), defined as

$$d = \frac{1}{n(n-1)} \sum_{i\neq j} |X_i - X_j|,$$

is also a U-statistic. You can find several examples of U-statistics, together with many brilliant limiting theorems, in the giant paper by W. Hoeffding (1948), who was also known for Hoeffding's inequality that we learned in Lecture 2.

6.2 The Generalized U-statistics with Applications

We now extend this notion of U-statistics. Let $\{X_i\}$ be a sequence of independent random variables in a measurable spqce $(S, \mathcal{S})$ and $\boldsymbol{f} = \{f_{ij},\ 1 \le i \neq j \le n\}$, a family of Bochner integrable functions of two variables taking $S \times S$ into $(D, \|\cdot\|)$ a Banach space[2]. Then we define the generalized U-statistic U_n as

$$U_n = \sum_{1\le i\neq j\le n} f_{ij}(X_i, X_j) \tag{6.3}$$

You can notice that the usual U-statistics can be obtained by letting $f_{ij} = f/\binom{n}{2}$. And such a generalized version may remind you more examples. For instance, the quadratic form $X^T AX = \sum_{1\le i\neq j\le n} a_{ij}X_iX_j$, where the diagonal elements of the symmetric matrix A are set to be zero.

We can also link the generalized U-Statistic to random colored graph theory. Let $\{X_i\}_{i=1}^n$ a independent sequence of i.i.d. random variables, i.e., $X_i \overset{\ulcorner}{=} X$ for some random variables X. Consider the complete graph $G = (V, E)$, where $|V|= n$ and X_i is the color of the vertex i. Now we let $f_{ij} = f$ for some f fixed, and if f

[2]If you are not familiar with Bochner integrability or Banach space, you may just think each f_{ij} is integrable and D is an Euclidian space.

is symmetric, then

$$S_n(f) = \sum_{1 \le i \ne j \le n} f(X_i, X_j)$$

is a U-statistic (not averaged) representing some color information of vertices. And if we let $X \sim Ber(p)$ and $f(x_1, x_2) = (1 - x_1)x_2$,which is not symmetric, then

$$S_n'(f) = \sum_{1 \le i < j \le n} f(X_i, X_j)$$

counts patterns beginning with 0 and ending with 1 in this random sequence. If the vertex $X_i = 1$ (resp, 0) indicates that this vertex is black (resp, white), then with $f(x_1, x_2) = \mathbb{I}_{\{x_1 \ne x_2\}}$, the statistic

$$S_n{}"(f) = \sum_{1 \le i < j \le n} f(X_i, X_j)$$

counts the edges with one black and one white end-point. You may see more complex random graph structures connected to U-statistics in the article by S, Janson and K. Nowicki (1991).

6.3 Decoupling Inequalities for U-statistics

You may notice that, although $X_1, ..., X_n$ are mutually independent, the random variables $f_{ij}(X_i, X_j)$'s are dependent, if i or j is fixed. This cause a difficulty in evaluating the expectation of $\|\sum_{1 \le i < j \le n} f(X_i, X_j)\|$ and $\Phi\|\sum_{1 \le i < j \le n} f(X_i, X_j)\|$ for some $\Phi : \mathbb{R}_0^+ \to \mathbb{R}$ convex increasing.

Let us make the problem more complex, but gives a formal statement: Let $X_1, ..., X_n$ be a sequence of independent random variables in a measurable space $(S, \mathcal{S})$ and let Π_{ij}^m be a family of Bochner integrable functions f_{ij}^m such that $f_{ij}^m : S \times S \mapsto D$ with $(D, \|\cdot\|)$ a Banach space. Let N_n be an arbitrary subset of $\{1, 2, ..., n\}$ and $\Phi : \mathbb{R}_0^+ \to \mathbb{R}$ be convex increasing such that

$$\max_{1 \le i \ne j \le n} \mathbb{E}\Phi\left(\max_{m \in N_n} \sup_{f_{ij}^m \in \Pi_{ij}^m} \|f_{ij}^m(X_i, X_j)\|\right) < \infty.$$

Then for $\boldsymbol{f}^m = \{f_{ij}^m\}_{1 \le i \ne j \le n}$, $\boldsymbol{\Pi}^m = \{\Pi_{ij}^m\}_{1 \le i \ne j \le n}$, how can we bound

$$\mathbb{E}\Phi\left(\max_{m \in N_n} \sup_{\boldsymbol{f}^m \in \boldsymbol{\Pi}^m} \|\sum_{1 \le i \ne j \le n} f_{ij}^m(X_i, X_j)\|\right)?$$

Remember that in lecture 1, I briefly introduced the tangent decoupling, which can be applied in this context. Think about the filtration $\mathcal{F}_i = \sigma(X_1, ..., X_i)$, and you can write

$$U_n = \sum_{1 \le i \ne j \le n} f_{ij}^m(X_i, X_j) = \sum_{j=2}^{n} \sum_{i=1}^{j-1} f_{ij}^m(X_i, X_j),$$

where $\sum_{i=1}^{j-1} f_{ij}^m(X_i, X_j)$ is adapted to $\mathcal{F}_j$. Suppose that we have $\{\tilde{X}_i\}_{i=1}^n$ an independent copy of $\{X_i\}_{i=1}^n$, then $\sum_{i=1}^{j-1} f_{ij}^m(X_i, \tilde{X}_j)$ is independent of $\sum_{i=1}^{j-1} f_{ij}^m(X_i, X_j)$ given $\mathcal{F}_j$. This provides a way to bound the above expectation, with the following theorem:

Theorem 6.1 *With the aforementioned setting,*

$$\mathbb{E}\Phi\left(\max_{m \in N_n} \sup_{\boldsymbol{f}^m \in \boldsymbol{\Pi}^m} \| \sum_{1 \le i \ne j \le n} f_{ij}^m(X_i, X_j) \|\right) \le \mathbb{E}\Phi\left(8 \max_{m \in N_n} \sup_{\boldsymbol{f}^m \in \boldsymbol{\Pi}^m} \| \sum_{1 \le i \ne j \le n} f_{ij}^m(X_i, \tilde{X}_j) \|\right). \tag{6.4}$$

And if $f_{ij}^m \in \Pi_{ij}^m$ satisfy the symmetry conditions

$$f_{ij}^m = f_{ji}^m \text{ and } f_{ij}^m(X_i, X_j) = f_{ij}^m(X_j, X_i),$$

then the reverse bound holds:

$$\mathbb{E}\Phi\left(\frac{1}{4} \max_{m \in N_n} \sup_{\boldsymbol{f}^m \in \boldsymbol{\Pi}^m} \| \sum_{1 \le i \ne j \le n} f_{ij}^m(X_i, \tilde{X}_j) \|\right) \le \mathbb{E}\Phi\left(\max_{m \in N_n} \sup_{\boldsymbol{f}^m \in \boldsymbol{\Pi}^m} \| \sum_{1 \le i \ne j \le n} f_{ij}^m(X_i, X_j) \|\right). \tag{6.5}$$

Remark 6.2 *The fact that the lower bound does not hold for general f_{ij} follows trivially by using*

$$f_{ij}(X_i, X_j) = X_j - X_i$$

because then $\sum_{i \ne j} f_{ij}(X_i, X_j) = 0$. But one may still obtain a lower bound by using the symmetrized kernels $\hat{f}_{ij} = [f_{ij}(X_i, X_j) + f_{ij}(X_j, X_i)]/2$ for $i \ne j$ and letting $\hat{f}_{ji} = \hat{f}_{ij}$

Proof: We demonstrate only the first equation (6.4) here, with a trivial lemma. But we first, for simplicity, denote by $\mathbb{E}_\sigma Y = \mathbb{E}[Y|\sigma]$, where Y is an r.v. and σ is a σ-field, and let sup denote $\sup_{\boldsymbol{f} \in \boldsymbol{\Pi}}$.

Lemma 6.3 *Let $\mathscr{X} = \sigma(X_i, i = 1, ..., n)$, where $\{Z_i\}$ is a sequence of independent random vectors with*

$Z_i = (X_i, \tilde{X}_i)$ w.p. $\frac{1}{2}$ and $Z_i = (\tilde{X}_i, X_i)$ w.p. $\frac{1}{2}$. Then,

$$\begin{aligned}\mathbb{E}_{\mathscr{Z}} f_{ij}(X_i, X_j) &= \mathbb{E}_{\mathscr{Z}} f_{ij}(X_i, X_j) \\ &= \frac{1}{4}[f_{ij}(X_i, X_j) + f_{ij}(X_i, \tilde{X}_j) + f_{ij}(\tilde{X}_i, X_j) + f_{ij}(\tilde{X}_i, \tilde{X}_j)]\end{aligned} \tag{6.6}$$

It is not hard to verify this lemma, by applying the same conditional law of $f_{ij}(X_i, \tilde{X}_j)$ and $f_{ij}(\tilde{X}_i, X_j)$ given $\mathscr{Z}$, and noticing that the sum of those four terms is measurable w.r.t. $\mathscr{Z}$.

Setting $\mathscr{X} = \sigma(X_1, ..., X_n)$, we use the following identity:

$$\begin{aligned}\sum_{1\le i\neq j\le n} f_{ij}(X_i, X_j) = &\sum_{1\le i\neq j\le n} \left[\mathbb{E}_{\mathscr{X}} f_{ij}(X_i, X_j) + \mathbb{E}_{\mathscr{X}} f_{ij}(X_i, \tilde{X}_j) + \mathbb{E}_{\mathscr{X}} f_{ij}(\tilde{X}_i, X_j) + \mathbb{E}_{\mathscr{X}} f_{ij}(\tilde{X}_i, \tilde{X}_j)\right] \\ &- \sum_{1\le i\neq j\le n} \left[\mathbb{E}_{\mathscr{X}} f_{ij}(X_i, \tilde{X}_j) + \mathbb{E}_{\mathscr{X}} f_{ij}(\tilde{X}_i, X_j) + \mathbb{E}_{\mathscr{X}} f_{ij}(\tilde{X}_i, \tilde{X}_j)\right].\end{aligned}$$

From the preceding and the complexity of Φ, it follows that

$$\begin{aligned}&\mathbb{E}\Phi\left(\sup\|\sum_{1\le i\neq j\le n} f_{ij}(X_i, X_j)\|\right) \\
&\le \mathbb{E}\Phi\left(\sup\|\sum_{1\le i\neq j\le n} \mathbb{E}_{\mathscr{X}}\left[f_{ij}(X_i, X_j) + f_{ij}(X_i, \tilde{X}_j) + f_{ij}(\tilde{X}_i, X_j) + f_{ij}(\tilde{X}_i, \tilde{X}_j)\right]\|\right. \\
&+ \mathbb{E}_{\mathscr{X}}\left[f_{ij}(X_i, \tilde{X}_j) + f_{ij}(\tilde{X}_i, X_j) + f_{ij}(\tilde{X}_i, \tilde{X}_j)\right]\| \qquad \text{[by \textsc{Minkowski} inequality]} \\
&\le \frac{1}{2}\mathbb{E}\Phi\left(2\sup\|\sum_{1\le i\neq j\le n} \mathbb{E}_{\mathscr{X}}\left[f_{ij}(X_i, X_j) + f_{ij}(X_i, \tilde{X}_j) + f_{ij}(\tilde{X}_i, X_j) + f_{ij}(\tilde{X}_i, \tilde{X}_j)\right]\|\right) \\
&+ \frac{1}{2}\mathbb{E}\Phi\left(2\sup\|\sum_{1\le i\neq j\le n} \mathbb{E}_{\mathscr{X}}\left[f_{ij}(X_i, \tilde{X}_j) + f_{ij}(\tilde{X}_i, X_j) + f_{ij}(\tilde{X}_i, \tilde{X}_j)\right]\|\right) \text{ [by the convexity of } \Phi] \\
&\le \frac{1}{2}\mathbb{E}\Phi\left(2\sup\|\sum_{1\le i\neq j\le n} \left[f_{ij}(X_i, X_j) + f_{ij}(X_i, \tilde{X}_j) + f_{ij}(\tilde{X}_i, X_j) + f_{ij}(\tilde{X}_i, \tilde{X}_j)\right]\|\right) \\
&+ \frac{1}{2}\mathbb{E}\Phi\left(2\sup\|\sum_{1\le i\neq j\le n} \mathbb{E}_{\mathscr{X}}\left[f_{ij}(X_i, \tilde{X}_j) + f_{ij}(\tilde{X}_i, X_j) + f_{ij}(\tilde{X}_i, \tilde{X}_j)\right]\|\right) \text{ [conditional \textsc{Jensen} inequality]} \\
&\le \frac{1}{2}\mathbb{E}\Phi\left(8\sup\|\sum_{1\le i\neq j\le n} \mathbb{E}_{\mathscr{Z}} f_{ij}(X_i, \tilde{X}_j)\|\right) + \frac{1}{6}\left[\mathbb{E}\Phi\left(6\sup\|\sum_{1\le i\neq j\le n} \mathbb{E}_{\mathscr{X}} f_{ij}(X_i, \tilde{X}_j)\|\right)\right. \\
&+ \mathbb{E}\Phi\left(6\sup\|\sum_{1\le i\neq j\le n} \mathbb{E}_{\mathscr{X}} f_{ij}(\tilde{X}_i, X_j)\|\right) + \mathbb{E}\Phi\left(6\sup\|\sum_{1\le i\neq j\le n} \mathbb{E} f_{ij}(\tilde{X}_i, \tilde{X}_j)\|\right) \text{[by (6.6) and } \Phi \text{ convex]}\end{aligned}$$

$$\leq \frac{1}{2}\mathbb{E}\Phi\left(8\sup\|\sum_{1\leq i\neq j\leq n}\mathbb{E}_{\mathscr{X}}f_{ij}(X_i,\tilde{X}_j)\|\right)+\frac{2}{6}\mathbb{E}\Phi\left(6\sup\|\sum_{1\leq i\neq j\leq n}\mathbb{E}f_{ij}(X_i,\tilde{X}_j)\|\right)$$

$$+\frac{1}{6}\mathbb{E}\Phi\left(6\sup\|\sum_{1\leq i\neq j\leq n}\mathbb{E}f_{ij}(\tilde{X}_i,\tilde{X}_j)\|\right)\qquad\text{[by conditional J}\textsc{ensen}\text{ and that }\mathbb{E}f_{ij}(\tilde{X}_i,\tilde{X}_j)=\mathbb{E}f_{ij}(X_i,\tilde{X}_j)\text{]}$$

$$\leq \frac{1}{2}\mathbb{E}\Phi\left(8\sup\|\sum_{1\leq i\neq j\leq n}\mathbb{E}_{\mathscr{X}}f_{ij}(X_i,\tilde{X}_j)\|\right)+\frac{1}{2}\mathbb{E}\Phi\left(6\sup\|\sum_{1\leq i\neq j\leq n}\mathbb{E}f_{ij}(X_i,\tilde{X}_j)\|\right)\qquad\text{[by J}\textsc{ensen}\text{ inequality]}$$

$$\leq \mathbb{E}\Phi\left(8\sup\|\sum_{1\leq i\neq j\leq n}\mathbb{E}_{\mathscr{X}}f_{ij}(X_i,\tilde{X}_j)\|\right)\qquad\text{[by }\Phi\text{ increasing]}$$

■