THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

SDS 384 11: Theoretical Statistics

Lecture 8: U Statistics

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

U Statistics

- We will see many interesting examples of U statistics.
- Interesting properties
  - Unbiased
  - Reduces variance
  - Concentration (via McDiarmid)
  - Aymptotic variance
  - Asymptotic distribution

An estimable parameter

- Let $\mathcal{P}$ be a family of probability measures on some arbitrary measurable space.
- We will now define a notion of an an estimable parameter. (coined "regular parameters" by Hoeffding.)
- An estimable parameter $\theta(P)$ satisfies the following.

Theorem (Halmos)

θ admits an unbiased estimator iff for some integer m there exists an unbiased estimator of $\theta(P)$ based on $X_1, \ldots, X_m \overset{iid}{\sim} P$ that is, if there exists a real-valued measurable function $h(X_1, ..., X_m)$ such that

$$\theta = Eh(X_1, \ldots, X_m).$$

The smallest integer m for which the above is true is called the degree of $\theta(P)$.

U statistics

- The function h may be taken to be a symmetric function of its arguments.
- This is because if $f(X_1, \ldots, X_m)$ is an unbiased estimator of $\theta(P)$, so is
$$h(X_1, \ldots, X_m) := \frac{\sum_{\pi \in \Pi_m} f(X_{\pi_1}, \ldots, X_{\pi_m})}{m!}$$
- For simplicity, we will assume h is symmetric for our notes.

U Statistics (Due to Wassily Hoeffding in 1948)

Definition

Let $X_i \overset{iid}{\sim} f$, let $h(x_1, \ldots, x_r)$ be a symmetric kernel function and $\Theta(F) = E[h(x_1, \ldots, x_r)]$. A U-statistic U_n of order r is defined as

$$U_n = \frac{\sum_{\{i_1,\ldots,i_r\} \in \mathcal{I}_r} h(X_{i_1}, X_{i_2}, \ldots, X_{i_r})}{\binom{n}{r}},$$

where $\mathcal{I}_r$ is the set of subsets of size r from $[n]$.

Sample variance as an U-Statistic

Example

The sample variance is an U-statistic of order 2.

Proof.

Let $\theta(F) = \sigma^2$.

$$\begin{aligned}\sum_{i \neq j}^{n} (X_i - X_j)^2 &= 2n \sum_i X_i^2 - 2 \sum_{i,j} X_i X_j \\ &= 2n \sum_i X_i^2 - 2n^2 \bar{X}^2 \\ &= 2n(n-1) \frac{\sum_i X_i^2 - n\bar{X}^2}{n-1}\end{aligned}$$

$$U_n := \frac{\sum_{i<j}^{n} (X_i - X_j)^2/2}{n(n-1)/2} = s_n^2$$

$\square$

Sample variance as U-statistic

- Is its expectation the variance?
- $\frac{1}{2}E[(X_1 - X_2)^2] = \frac{1}{2}E(X_1 - \mu - (X_2 - \mu))^2 = \sigma^2$

U-statistics examples: Wilcoxon one sample rank statistic

Example

$U_n = \sum_i R_i 1(X_i > 0)$, where R_i is the rank of X_i in the sorted order $|X_1| \leq |X_2| \ldots$.

- This is used to check if the distribution of X_i is symmetric around zero.
- Assume X_i to be distinct.
- $R_i = \sum_{j=1}^{n} 1(|X_j| \leq |X_i|)$

U-statistics examples: Wilcoxon one sample rank statistic

Example

$T_n = \sum_i R_i 1(X_i > 0)$, where R_i is the rank of X_i in the sorted order $|X_1| \le |X_2| \dots$.

$$
\begin{aligned}
T_n = \sum_i R_i 1(X_i > 0) &= \sum_{i=1}^n \sum_{j=1}^n 1(|X_j| \le |X_i|) 1(X_i > 0) \\
&= \sum_{i=1}^n \sum_{j=1}^n 1(|X_j| \le X_i) 1(X_i \neq 0) = \sum_{i \neq j}^n 1(|X_j| \le X_i) + \sum_{i=1}^n 1(X_i > 0) \\
&= \sum_{i<j} 1(|X_j| < X_i) + \sum_{i<j} 1(|X_i| < X_j) + \sum_{i=1}^n 1(X_i > 0) \\
&= \sum_{i<j} 1(X_i + X_j > 0) + \sum_{i=1}^n 1(X_i > 0) = \binom{n}{2} U_2 + n U_1
\end{aligned}
$$

- Asymptotically dominated by the first term, which is an U statistic.

Properties of the U-statistic

- The U is for unbiased.
- Note that $E[U] = Eh(X_1, \ldots, X_r)$
- $\text{var}(U(X_1, \ldots, X_r)) \leq \text{var}(h(X_1, \ldots, X_r))$ (Rao Blackwell theorem)
    - Just $h(X_1, \ldots, X_r)$ is an unbiased estimator of $\theta(F)$.
    - But averaging over many subsets reduces variance.

Properties of U-statistics

- Let $X_{(1)} \dots, X_{(n)}$ denote the order statistics of the data.
- The empirical distribution puts $1/n$ mass on each data point.
- So we can think about the U statistic as

$$U_n = E[h(X_1, \dots, X_r)|X_{(1)}, \dots, X_{(n)}]$$

- We also have:

$$\begin{aligned} E[(U-\theta)^2] &= E\left[\left(E[h(X_1, \dots, X_r) - \theta|X_{(1)}, \dots, X_{(n)}]\right)^2\right] \\ &\leq E[E[(h(X, \dots, X_r) - \theta)^2|X_{(1)}, \dots, X_{(n)}]] \\ &= \mathrm{var}(h(X_1, \dots, X_r)) \end{aligned}$$

- Rao-Blackwell theorem says that the conditional expectation of any estimator given the sufficient statistic has smaller variance than the estimator itself.
- For $X_1, ;X_n \overset{iid}{\sim} P$, the order statistics are sufficient. (why?)

More novel examples

Example (Gini's mean difference/ mean absolute deviation)

Let $\theta(F) := E[|X_1 - X_2|]$; the corresponding U statistic is
$U_n = \frac{\sum_{i<j} |x_i - x_j|}{\binom{n}{2}}$.

Example (Quantile Statistic)

Let $\theta(F) := P(X_1 \leq t) = E[1(X_1 \leq t)]$; the corresponding U statistic is
$U_n = \frac{\sum_i 1(X_i \leq t)}{n}$.