Feynman Liang
STAT 210A
Lecture #1
8/25/2016

1 Logistics

Syllabus: on Piazza.

Textbook: Wainwright: *Concentration Inequalities & High Dimensional Statistics*. At copy center on Hearst.

GSI: `hmania@berkeley.edu`

2 Introduction

A story of n and p

$$y = X\beta + \varepsilon, \qquad X \in \mathbb{R}^{n \times p} \tag{2.1}$$

Statistics try to create estimators $\hat{\beta}$, prove rates of convergence, find distributions of estimators (e.g. probability of errors, confidence interval).

Asymptotics: clasically $n \to \infty$, in this class will also consider $p \to \infty$ and the two going at different rates.

3 U-statistic

Random sample $X_1, \ldots, X_n \overset{\text{iid}}{\sim} P$. Parameter $\theta(P)$.

$$\theta = \mathbb{E} h(X_1, \ldots, X_r) \tag{3.1}$$

One possible estimator is the *empirical average* $\hat{\theta} = h(X_1, \ldots, X_r)$. But this throws away data $\{X_k : k > r\}$.

Assume, wlog (can symmetrize by add up all possible permutations to get a new function), that h is *permutation symmetric*.

Definition 3.1. A *U-statistic*

$$U = \binom{n}{r}^{-1} \sum_{\beta} h(X_{\beta_1}, \ldots, X_{\beta_r}) \tag{3.2}$$

β ranges over all unordered substs of $\{1, \ldots, n\}$ of size r.

U is unbiased: $\mathbb{E}U = \theta$.

Order statistics $\{X_{(1)}, \ldots, X_{(n)}\}$ sufficient, and since U itself is a sum over unordered subsets, it is the Rao-Blackwell estimator

$$U = \mathbb{E}(h(X_1, \ldots, X_r) \mid X_{(1)}, \ldots, X_{(n)}) \tag{3.3}$$

Theorem 3.2 (Asymptotic Normality of U statistics)**.** *If* $\mathbb{E}h^2(X_1, \ldots, X_r) < \infty$*, then*

$$\sqrt{n}(U - \theta) \xrightarrow{d} N(0, \_) \tag{3.4}$$

Example 3.3. $r = 1$. Then $U = n^{-1} \sum_{i=1}^n h(X_i)$.

Example 3.4. $r = 2$. $h(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$ (sometimes called *kernel*).

Then h is an unbiased estimator of $\theta = \mathrm{Var}X_1$ because

$$\mathbb{E}\frac{1}{2}(X_1 - X_2)^2 = \mathbb{E}\frac{1}{2}((X_1 - \mu) - (X_2 - \mu))^2 = \mathrm{Var}X_1 \tag{3.5}$$

The U-estimator

$$\begin{aligned} U &= \binom{n}{2}^{-1} \sum_{i<j} \frac{1}{2}(X_i - X_j)^2 && (3.6)\\ &= \frac{2}{n(n-1)} \sum_{i<j} \frac{1}{2}((X_i - \bar{X}) - (X_j - \bar{X}))^2 && (3.7)\\ &= TODO : algebra && (3.8)\\ &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 && (3.9) \end{aligned}$$

Example 3.5 (Signed rank statistic)**.**

$$\begin{aligned} h(x_1, x_2) &= 1_{x_1 + x_2 > 0} && (3.10)\\ \theta &= P(X_1 + X_2 > 0) && (3.11)\\ U &= \binom{n}{2}^{-1} \sum_{i<j} 1_{X_i + X_j > 0} && (3.12) \end{aligned}$$

Example 3.6 (Kendall's τ)**.** Data $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$.

$$\tau = \frac{4}{n(n-1)} \sum_{i<j} 1_{(Y_j - Y_i)(X_j - X_i)} \tag{3.13}$$

3.1 Computing the variance

$$\text{Var } U = \binom{n}{r}^{-2} \sum_{\beta} \sum_{\beta'} \text{Cov}(h(X_{\beta_1}, \ldots, X_{\beta_r}), h(X_{\beta'_1}, \ldots, X_{\beta'_r})) \tag{3.14}$$

Whenever $\beta \cap \beta' = \emptyset$ we get the same constant value, 0 if X_i independent.

By iid, if $\#(\beta \cap \beta') = 1$, it doesn't matter which k satisfies $X_{\beta_k} = X_{\beta'_k}$. This holds for any number of variables in the overlap, so define

$$\xi_c = \text{Cov when there are } c \text{ variables in common in } \beta \text{ and } \beta' \tag{3.15}$$

And we have

$$\text{Var } U = \binom{n}{r}^{-2} \sum_{c=0}^{r} \underbrace{\binom{n}{c}}_{\text{Num overlap variables}} \underbrace{\binom{r}{c}}_{\text{Which of } r \text{ variables in } \beta \text{ overlap}} \underbrace{\binom{n-r}{r-c}}_{\text{Remaining choices in } \beta'} \xi_c \tag{3.16}$$

$$= \sum_{i=1}^{r} \frac{r!^2}{c!(r-c)!^2} \frac{(n-r)\cdots(n-2r+c+1)}{n(n-1)\cdots(n-r+1)} \tag{3.17}$$

For $c = 1, r = 3$

$$\text{Var } U = \frac{(n-3)(n-4)}{n(n-1)(n-2)} \in O(n^{-1}) \tag{3.18}$$

For $c = 2, r = 3$, Var $U \in O(n^{-2})$.

Looks like a series expansion on variance, but not quite central limit theorem.

3.2 Hajek projectio

- Want the limiting distribution of random variables $\{T_n\}$
- Relate $\{T_n\}$ to sequence $\{S_n\}$ for which we know the limit

$$T_n = (T_n - S_n) + S_n \tag{3.19}$$

- Show $T_n - S_n \xrightarrow{\text{p}} 0$ and apply Slutsky's theorem to get $T_n \xrightarrow{d} S$

Aside: modes of convergence TODO: Fig 1.1

Let $\mathcal{S}$ be a linear space of random vectors with finite second moments.

Definition 3.7. $\hat{S}$ is a *projection* of T onto $\mathcal{S}$ iff $\hat{S} \in \mathcal{S}$ and $\mathbb{E}(T - \hat{S})S = 0$ for $S \in \mathcal{S}$. TODO: Fig 1.2

If $\mathcal{S}$ contains constants, then

$$\mathbb{E}T = \mathbb{E}\hat{S}, \qquad \text{Cov}(T - \hat{S}, S) = 0 \tag{3.20}$$

Theorem 3.8. *If* $\frac{Var\ T_n}{Var\ \hat{S_n}} \to 1$*, then*

$$\frac{T_n - \mathbb{E}T_n}{\sqrt{Var\ T_n}} - \frac{\hat{S}_n - \mathbb{E}\hat{S}_n}{\sqrt{Var\ \hat{S}_n}} \xrightarrow{p} 0 \tag{3.21}$$

Need to prove in probability, but TODO: Ref 1.1 means suffices to show convrgence in quadratic mean.

Proof. $\mathbb{E}$ of eq. (3.21) $= 0$.

Var(eq. (3.21)) $= 2 - 2\frac{\mathrm{Cov}(T_n, \hat{S}_n)}{\sqrt{\mathrm{Var} T_n}\sqrt{\mathrm{Var} \hat{S}_n}}$.

TODO: Finish last three lines, show going to zero □