Feynman Liang
STAT 210B
Lecture #2
1/19/2017

(a) Room change: Evans 330

(b) HW1 distributed today

1 Review

$$U = \frac{1}{\binom{n}{r}} \sum_\beta h(X_{\beta_1}, \dots, X_{\beta_r}) \tag{1.1}$$

$$\operatorname{Var} U = \binom{n}{r}^{-2} \sum_{c=1}^{r} \binom{n}{r} \binom{r}{c} \binom{n-r}{r-c} \zeta_c \tag{1.2}$$

U is asymptotically Gaussian, as we will show using Hajek projections: project sequence $\{T_n\}$ onto sequence $\{S_n\}$

2 Hajek Projection of U-statistics

Aside: Hilbert spaces of random variables gives notions of orthogonality. TODO: Fig 1.1 again

$T_n = T_n - S_n + S_n$, so if we can show $T_n - S_n \to 0$ then by Slutsky's theorem $T_n \to S_n$.

Theorem 2.1.

$$\frac{\operatorname{Var} T_n}{\operatorname{Var} \hat{S}_n} \to 1 \implies \underbrace{\frac{T_n - \mathbb{E} T_n}{\sigma(T_n)} - \frac{\hat{S}_n - \mathbb{E}\hat{S}_n}{\sigma(\hat{S}_n)}}_{(*)} \xrightarrow{p} 0 \tag{2.1}$$

Proof.

$$\operatorname{Var}(*) = 2 - 2\frac{Cov(T_n, \hat{S}_n)}{\sigma(T_n)\sigma(\hat{S}_n)} \tag{2.2}$$

$$T_n - \hat{S}_n \perp \hat{S}_n \implies \mathbb{E} T_n \hat{S}_n = \mathbb{E} \hat{S}_n^2 \tag{2.3}$$

$$\implies Cov(T_n, \hat{S}_n) = \mathbb{E}\hat{S}_n^2 - (\mathbb{E}\hat{S}_n)^2 = \operatorname{Var} \hat{S}_n \tag{2.4}$$

Hence

$$\operatorname{Var}(*) = 2 - 2\frac{Cov(T_n, \hat{S}_n)}{\sigma(T_n)\sigma(\hat{S}_n)} = 2 - 2\frac{\operatorname{Var} \hat{S}_n}{\sigma T_n \sigma \hat{S}_n} \tag{2.5}$$

Taking $n \to \infty$

$$\mathrm{Var}(*) = 2 - 2\frac{\mathrm{Var}\,\hat{S}_n}{\sigma T_n \sigma \hat{S}_n} \to 2 - 2 \cdot 1 = 0 \tag{2.6}$$

Since $\mathrm{Var}(*) \to 0$ and $\mathbb{E}(*) \to 0$, we have $(*) \xrightarrow{qm} 0$ which implies $(*) \xrightarrow{\mathrm{p}} 0$. $\square$

Still need to choose what element in Hilbert space to project onto. Consider

$$\mathcal{S} = \left\{ \sum_{i=1}^{n} g_i(X_i) \right\} \tag{2.7}$$

Let $T_n = U$ be the U-statistic on n samples.

Theorem 2.2. *The projection of T onto $\mathcal{S}$ is $\hat{S} = \sum_{i=1}^{n} \mathbb{E}[T \mid X_i] - (n-1)\mathbb{E}T$.*

Proof. Independence and tower property $\implies$ $\mathbb{E}[\mathbb{E}[T \mid X_i] \mid X_j] = \mathbb{E}[\mathbb{E}[T \mid X_i]] = \mathbb{E}T$ for $i \neq j$.

$\mathbb{E}[\hat{S} \mid X_i] = (n-1)\mathbb{E}T + \mathbb{E}[T \mid X_j] - (n-1)\mathbb{E}T = \mathbb{E}[T \mid X_j]$.

$\mathbb{E}[T - \hat{S}]g_j(X_j) = \mathbb{E}[\mathbb{E}[T - \hat{S}]g_j(X_j) \mid X_j] = \mathbb{E}[\mathbb{E}[(T - \hat{S} \mid X_j]g_j(X_j)] = 0$ $\square$

Back to U-statistics; project $U - \underbrace{\theta}_{=\mathbb{E}U}$ onto $\{\sum_i g_i(X_i)\}$ to get

$$\hat{U} = \sum_{i=1}^{n} \mathbb{E}[U - \theta \mid X_i] \tag{2.8}$$

Consider one term:

$$\mathbb{E}[h(X_{\beta_1}, \ldots, X_{\beta_r}) - \theta \mid X_i = x] = \begin{cases} h_1(x), & \text{if } i \in \beta \\ 0, & \text{otherwise} \end{cases} \tag{2.9}$$

The $i \in \beta$ case occurs $\binom{n-1}{r-1}$ ways. $\frac{\binom{n-1}{r-1}}{\binom{n}{r}} = \frac{r}{n}$. So

$$\hat{U} = \frac{r}{n} \sum_{i=1}^{n} h_i(X_i) \tag{2.10}$$

$$\mathrm{Var}\,\hat{U} = \frac{r^2}{n^2} n \mathbb{E} h_1^2(X_1) \tag{2.11}$$

$$= \frac{r^2}{n} \int \mathbb{E}[h(x, X_2, \ldots, X_r) - \theta]\mathbb{E}[h(x, X_2', \ldots, X_r') - \theta]P(dx) \tag{2.12}$$

$$= \frac{r^2}{n}\zeta_1 \tag{2.13}$$

Hence we have

$$\frac{\mathrm{Var}\,U}{\mathrm{Var}\,\hat{U}} = \frac{\frac{r^2}{n}\zeta_1 + O(\frac{1}{n^2})\zeta_2 + O(\frac{1}{n^r})\zeta_r}{\frac{r^2}{n}\zeta_1} = \frac{\frac{r!^2}{1!(r-1)!^2}\frac{1}{n}\zeta_1 + \cdots}{\frac{r^2}{n}\zeta_1} \tag{2.14}$$

$$\xrightarrow{n\to\infty} 1 \tag{2.15}$$

Applying CLT to $\hat{U}$ and ref thm 2.1 yields a CLT for U.

3 Concentration inequalities

Markov's inequality: $X \geq 0$. $P(X \geq t) \leq \frac{\mathbb{E}X}{t}$. TODO: Fig 2.1

Chebyshev's inequality: Assuming $\operatorname{Var} X$ exists, $P(|X-\mu| \geq t) \leq \frac{\operatorname{Var} X}{t^2}$.

Higher moment versions exist: $P(|X-\mu| \geq t) \leq \frac{\mathbb{E}|X-\mu|^k}{t^k}$

Chernoff: Let $\phi(\lambda) = \mathbb{E}e^{\lambda(X-\mu)}$ exists $|\lambda| < b$, $b > 0$. Applying Markov

$$P(X-\mu \geq t) = P(e^{\lambda(X-\mu)} \geq e^{\lambda t}) \leq \frac{\mathbb{E}e^{\lambda(X-\mu)}}{e^{\lambda t}} \tag{3.1}$$

Since this holds for any λ, we have (after taking logs)

$$\log P(X-\mu \geq t) \leq \inf_{\lambda \in [a,b)} \left(\log \mathbb{E}e^{\lambda(X-\mu)} - \lambda t \right) \tag{3.2}$$

Moment generating function (mgf) of a Gaussian

$$\mathbb{E}e^{\lambda X} = e^{\mu\lambda + \frac{\sigma^2\lambda^2}{2}} \tag{3.3}$$

$$\inf_{\lambda \geq 0} (*) = \inf_{\lambda \geq 0} \left(\frac{\lambda^2\sigma^2}{2} - \lambda t \right) = -\frac{t^2}{2\sigma^2} \tag{3.4}$$

So applying Chernoff bound, we get an exponential tail bound

$$P(X \geq \mu + t) \leq e^{-\frac{t^2}{2\sigma^2}} \tag{3.5}$$

Definition 3.1. A random variable X is *sub-Gaussian* if $\exists \sigma > 0$ such that

$$\mathbb{E}e^{\lambda(X-\mu)} \leq e^{\lambda^2\sigma^2/2} \tag{3.6}$$