STAT 583: Advanced Theory of Statistical Inference **Homework 1**

Lecture 1: Homework

Lecturer: Fang Han *Due Date: Apr. 23*

- Please send the electronic copies of your HW to Yandi Shen's email box (*ydshen@uw.edu*).
- All assignments are due at 11:59PM on the due date.
- Each student will be allotted FOUR free days which can be used throughout the semester to turn in homework assignments late without penalty. For instance, you might choose to turn in HW1 two days late, HW2 one day late, and HW3 two days late. Once your free days are used up, late homeworks will be penalized 20% per day (For instance, a homework turned in two days late will receive only 60% credit.). Exceptions to these rules will of course be made for serious illness or other emergency circumstances; in these cases, please contact me as soon as you are aware of the problem.
- Latex-ed solutions are preferred, though not required.
- You should attempt to solve all homework problems on your own before joining a group to solve them together.
- Be sure to show your work and justify your answers in a mathematically rigorous fashion.
- It is a very good idea to start early on the problem sets, at least to read them over so that you can be thinking about them in the background.
- The lecture notes might have been updated so download the most recent version.

1.1 Probability inequalities

Problem 1 (Hoeffding's Lemma and Inequality)

(1) Prove, if $\mathbb{E}X = 0$ and $\mathbb{P}(X \in [a,b]) = 1$, then

$$\mathbb{E}\exp(tX) \le \exp(t^2(b-a)^2/8).$$

(2) Use the above inequality to further show that, if $X_1, \ldots, X_n$ are mean-zero independent random variables with $\mathbb{P}(X_i \in [a_i, b_i]) = 1$, then

$$\mathbb{E}\exp\Big(t\sum_{i=1}^n X_i\Big) \le \exp\Big(t^2\sum_{i=1}^n (b_i - a_i)^2/8\Big).$$

(3) Use the above inequality to derive the following "Hoeffding's inequality": under the same setting as in the Case (2), we have, for any $t > 0$,

$$P(|\overline{X}_n| > t) \le 2\exp\Big(-\frac{2n^2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\Big),$$

where $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ represents the sample mean.

Problem 2 (Exercise 3.1.10, Page 135, GN2015)

Let $\xi_i, i \in \mathbb{N}$ be independent centred random variables such that $\mathbb{E}\exp(\lambda\xi_i) < \infty$, for all $0 < \lambda < \lambda_0$ for some $\lambda_0 < \infty$. Set $S_k = \sum_{i=1}^k \xi_i, k \in \mathbb{N}$. Show that, for $0 < \lambda < \lambda_0$, the sequence $\{(\exp(\lambda S_k), \mathcal{S}_k)\}$, where $\mathcal{S}_k = \sigma(\xi_i : i \leq k)$, is a positive submartingale, and apply Doob's martingale inequality to obtain

$$P\Big(\max_{k \leq n} S_k > t\Big) \leq \mathbb{E}\exp(\lambda S_n)/\exp(\lambda t), \quad \text{for } t > 0.$$

Use this to derive a Hoeffding's inequality for $\max_{k \leq n} S_k$.

Problem 3 (Exercise 3.1.9, Page 135, GN2015)

Let $\xi_i, i \leq n$ be nonnegative independent random variables in L^p space, namely, $\mathbb{E}\xi_i^p < \infty$, for some $p > 0$. Prove the following three assertions.

a. For all $\delta > 0$,

$$\mathbb{E}\max_{1 \leq i \leq n} \xi_i^p \leq \delta^p + p\int_\delta^\infty t^{p-1}\sum_{i=1}^n P(\xi > t)dt.$$

b. Use that $1 - x \leq \exp(-x)$ and $1 - \exp(-x) \geq x/(1+x)$ to show that

$$P(\max_i \xi_i > t) \geq \frac{\sum_i P(\xi_i > t)}{1 + \sum_i P(\xi_i > t)}.$$

c. Suppose that $\delta_0 = \inf\{t : \sum_{i=1}^n P(\xi_i > t) \leq \lambda\}$ for some $\lambda > 0$. Use the preceding inequality and the monotonicity of the function $x/(1+x)$ to deduce

$$P(\max_i \xi_i > t) \geq \frac{\sum P(\xi > t)}{1+\lambda} \quad \text{for } t \geq \delta_0,$$

and

$$P(\max_i \xi_i > t) \geq \frac{\lambda}{1+\lambda} \quad \text{otherwise.}$$

Conclude

$$\mathbb{E}\max_{1 \leq i \leq n} \xi^p \geq \frac{\lambda}{1+\lambda}\delta_0^p + \frac{p}{1+\lambda}\int_{\delta_0}^\infty t^{p-1}\sum_{i=1}^n P(\xi_i > t)dt.$$

1.2 Weak convergence

Problem 4 (Delta method)

a. State and prove the Slutsky's Theorem;

b. State and prove the Delta's method (the multivariate version);

c. Show that the Delta's method implies the Slutsky's Theorem.

Problem 5 (Lemma 2.2, V2000)

Prove the Portmanteau Lemma (Lemma 2.2 in V2000), namely, for random vectors X_n and X, the following statements are equivalent:

a. $P(X_n \leq x) \to P(X \leq x)$ for all continuity points of $x \to P(X \leq x)$;

b. $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for all bounded continuous functions f;

c. $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for all bounded Lipschitz functions f;

d. $\liminf \mathbb{E}f(X_n) \geq \mathbb{E}f(X)$ for all nonnegative continuous functions f;

e. $\liminf P(X_n \in G) \geq P(X \in G)$ for every open set G;

f. $\limsup P(X_n \in F) \leq P(X \in F)$ for every closed set F;

g. $P(X_n \in B) \to P(X \in B)$ for all Borel sets B with $P(X \in \delta B) = 0$, where δB is the boundary of B.

1.3 Application of the Donsker

Problem 6

Derive the influence function of sample quantiles, and give a second proof of the ASN of sample quantiles. (You are not required to verify the Hadamard differentiability of $\theta(Q) = Q^{-1}$).

Problem 7

Consider i.i.d. observations $X_1, \ldots, X_n$ of $X \in \mathbb{R}$. Outline the conditions such that $M_n = n^{-1/2} \sum_{i=1}^n |X_i - \overline{X}_n|^p$ (for some $p \in (0, \infty)$) is asymptotically normal under these conditions. Derive the explicit limiting distribution of M_n under the given conditions.

Problem 8

Consider i.i.d. observations $X_1, \ldots, X_n$ of $X \in \mathbb{R}$. Outline the conditions such that $M_n = n^{-1/2} \sum_{i=1}^n X_i \mathbb{1}(X_i > \overline{X}_n)$ (where $\overline{X}_n := n^{-1} \sum_{i=1}^n X_i$) is asymptotically normal under these conditions. Derive the explicit limiting distribution of M_n under the given conditions.

Problem 9 (simple linear regression, fixed design)

Assume the standard linear regression model $Y_i = X_i^T \beta_0 + \epsilon_i$ for $i \in [n]$, where $\{X_i \in \mathbb{R}^p, i \in [n]\}$ are deterministic (fixed design) and $\{\epsilon_i \in \mathbb{R}, i \in [n]\}$ are i.i.d. copies of a mean-zero random variable ϵ. Consider the simple linear regression estimator of $\beta_0 \in \mathbb{R}^p$:

$$\widehat{\beta} := \operatorname*{argmin}_{\beta \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n (Y_i - X_i^T \beta)^2.$$

a. Show, under which conditions for $\{X_i, i \in [n]\}$ and ϵ, we have $\|\widehat{\beta} - \beta_0\|_2 = o_P(1)$?

b. Show, under which conditions for $\{X_i, i \in [n]\}$ and ϵ, we have $\|\widehat{\beta} - \beta_0\|_2 = O_P(1/\sqrt{n})$?

c. Show, under which conditions for $\{X_i, i \in [n]\}$ and ϵ, we have $\sqrt{n}(\widehat{\beta} - \beta_0)$ is asymptotically normal?

Problem 10 (simple linear regression, fixed design)

Repeat the arguments in Problem 9, (b) and (c). But this time, instead of using the closed form of $\widehat{\beta}$, verify the conditions in Theorems 14 and 15.

Problem 11 (simple linear regression, random design)

Assume the standard linear regression model $Y_i = X_i^T\beta_0 + \epsilon_i$ for $i \in [n]$, where $\{X_i \in \mathbb{R}^p, i \in [n]\}$ are i.i.d. copies of a random vector X and $\{\epsilon_i \in \mathbb{R}, i \in [n]\}$ are i.i.d. copies of a mean-zero random variable ϵ. Consider the simple linear regression estimator of $\beta_0 \in \mathbb{R}^p$:

$$\widehat{\beta} := \operatorname*{argmin}_{\beta \in \mathbb{R}^p} \frac{1}{n}\sum_{i=1}^n (Y_i - X_i^T\beta)^2.$$

a. Show, under which conditions for X and ϵ, we have $\|\widehat{\beta} - \beta_0\|_2 = o_P(1)$?

b. Show, under which conditions for X and ϵ, we have $\|\widehat{\beta} - \beta_0\|_2 = O_P(1/\sqrt{n})$?

c. Show, under which conditions for X and ϵ, we have $\sqrt{n}(\widehat{\beta} - \beta_0)$ is asymptotically normal?

Problem 12 (simple linear regression, random design)

Repeat the arguments in Problem 11, (b) and (c). But this time, instead of using the closed form of $\widehat{\beta}$, verify the conditions in Theorems 14 and 15.

Problem 13 (consistency of Z-estimators)

Let $\Psi_n : \mathbb{R}^p \to \mathbb{R}^q$ be a random vector-valued function, let Ψ be a fixed vector-valued function of $\theta \in \Theta \subset \mathbb{R}^p$, and let $\|\cdot\|$ be a norm in $\mathbb{R}^q$ that is not necessarily the Euclidean norm. Assume, for every $\epsilon > 0$,

$$\sup_{\theta \in \Theta} \|\Psi_n(\theta) - \Psi(\theta)\| = o_P(1) \quad \text{and} \quad \inf_{\theta : d(\theta,\theta_0) \geq \epsilon} \|\Psi(\theta)\| > 0 = \|\Psi(\theta_0)\|.$$

Prove that any sequence of estimators $\widehat{\theta}_n$ such that $\Psi_n(\widehat{\theta}_n) = o_P(1)$ satisfies $\|\widehat{\theta}_n - \theta_0\| = o_P(1)$.

Problem 14 (one-step estimator)

Consider the Z-estimation theory described in Problem 13. Assume the following Frechet differentiability condition: For any constant $M > 0$, we assume

$$\sup_{\sqrt{n}\|\theta - \theta_0\| < M} \|\sqrt{n}(\Psi_n(\theta) - \Psi_n(\theta_0)) - \dot{\Psi}_0\sqrt{n}(\theta - \theta_0)\| = o_P(1),$$

where $\dot{\Psi}_0$ is a deterministic nonsingular matrix. Given a preliminary root-n estimator $\widetilde{\theta}_n$ of θ_0 (namely, $\sqrt{n}\|\widetilde{\theta}_n - \theta_0\| = O_P(1)$), the one-step estimator $\widehat{\theta}_n$ is defined as

$$\widehat{\theta}_n = \widetilde{\theta}_n - \dot{\Psi}_{n,0}^{-1}\Psi_n(\widetilde{\theta}_n),$$

where we assume $\dot{\Psi}_{n,0} \xrightarrow{P} \dot{\Psi}_0$. Prove, provided $\sqrt{n}\Psi_n(\theta_0) \Rightarrow Z$ for some random variable Z,

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = -\dot{\Psi}_0^{-1}\sqrt{n}\Psi_n(\theta_0) + o_P(1).$$

In other words, $\widehat{\theta}_n$ is asymptotically equivalent to the solution of $\Psi_n(\theta) = 0$.

Problem 15

Use the one-step idea described in Problem 14 to construct an asymptotically efficient estimator of the median θ_0 for the Cauchy distribution of density:

$$p_{\theta_0}(x) = \frac{1}{\pi} \cdot \frac{1}{1 + (x - \theta_0)^2}.$$

You do not need to verify the Frechet differentiability condition. (Hint: a natural choice of $\widetilde{\theta}$ is the sample median, and a difficult to solve Z-estimator is the MLE.)