UC Berkeley
Department of Statistics

STAT 210A: Introduction to Mathematical Statistics

Problem Set

Issued: Nov. 18, Wednesday
Due: Dec. 10, Thursday (at beginning of class)

Policy

This is a special homework. There are 5 problems as usual but each of them has 10 subproblems. **Please select 30 subproblems out of 50.** For example, you can work on three entire problems or 6 subproblems from each. More generally you can allocate them in any pattern you want and only the total number of subproblems matters.

Do not worry if you get stuck in some problems. You can use the result from the parts that you do not solve. For example, you can skip Problem 9.1 (a)-(d) and go directly to (e) assuming the previous parts have been solved. If you find any typo or mistake, feel free to post it on piazza and we will correct them as soon as possible.

Preview

Hopefully, you have already seen the elegance of theoretical statistics in this course, which offers you a whirlwind tour in this fantastic world. These problems are designed for you to understand several most astonishing parts of statistics, which has been covered to some degree in the second half semester. Here is a short description of the problems.

Problem 1 Most popular estimators in statistics can be obtained by minimizing a convex function. For example, mean is the minimizer of L2 loss, median is the minimizer of L1 loss and quantiles are minimizers of a group of piecewise linear functions. It turns out that there exists a generic way to analyze them and establish the consistency under weak regularity conditions. Problem 9.1 decompose the fantastic idea, initialized by Pollard in 1990s, into small pieces so that each piece is doable. You can learn how to apply this technique to prove the consistency of mean, median, quantiles and even M estimator, LASSO estimator and ridge estimator. In this problem, we only deal with location estimation problem but it is easy to extend these ideas to corresponding regression estimators.

Problem 2 This is a continuation of Problem 9.1. After refining the above result a little bit, the asymptotic distribution of various estimators can be calculated without paying much more effect. You will see an alternative proof , originated by Pollard and Knight in 1990s and developed by Koenker in 2000s, of the asymptotic normality of median and this technique can be easily extended to establish the asymptotic normality for LAD or quantile regression.

Problem 3 Bootstrap is a generic methodology to approximate the finite sample behavior of estimators and is able to deal with complicated statistics whose limiting distribution is hard to be obtained analytically. The far-reaching paper by Efron (1979) reveals the intrinsic relationship between Bootstrap and Jacknife, which is a popular way to estimate the bias and variance of common estimators before. In this problem, you will compare these two techniques case by case. Usually the answers can not be obtained analytically, however, these problems are carefully designed and take advantage of some elegant properties of some distributions so that the final result can be written explicitly.

Problem 4 One of the most important technical tool in modern statistics is concentration inequalities and the union bound. you may feel overwhelmed at the first time you learn sub-Gaussianity, sub-exponentiality and various inequalites. But after enough exercises you will see how natural and, most importantly, how qualitative they are even though they look like very mathematical. This problem aims to establish the bounds for the maximum of random variables. It involves various techinques in the lectures. After suffering for a while these techniques will become your kindly friends and accompany you in the rest of your research life.

Problem 5 Random matrix theory, although used to be a tough topic, become increasingly more important in modern statistics. By using concentration inequalities and underlying geometry, you can figure out the finite sample behaviors of the random matrix accurately. In this problem, you will learn the relationship between sub-Gaussian random variables and sub-exponential random variables by using the techinques developed by Vershynin and then apply these results to obtain the bound for largest eigenvalues of sample covariance.

ENJOY THE HOMEWORK AND
HAPPY THANKSGIVNG!

Problem 9.1

Let $\lambda_n(\theta)$ be a random convex function of $\theta \in \Theta$ where Θ is an open subset of $\mathbb{R}^p$. Suppose $\lambda(\theta)$ is a deterministic convex function such that for any given θ,

$$\lambda_n(\theta) \xrightarrow{d} \lambda(\theta).$$

Let K be any compact subset K of Θ.

(a) Show that for any $\varepsilon > 0$,

$$P\left(\sup_{\theta \in K} \lambda_n(\theta) - \lambda(\theta) > \varepsilon\right) \to 0.$$

Hint: you may use the following facts: (i) a compact set can be covered by finite number of p-dimensional cubes; (ii) a convex function is always continuous and hence uniformly continuous in any compact set.

(b) Show that for any $\varepsilon > 0$,

$$P\left(\inf_{\theta \in K} \lambda_n(\theta) - \lambda(\theta) < -\varepsilon\right) \to 0.$$

(c) Conclude from (a) and (b) that

$$\sup_{\theta \in K} |\lambda_n(\theta) - \lambda(\theta)| \xrightarrow{p} 0.$$

(d) Assume that $\lambda(\theta)$ has a unique minimizer, denoted by θ_0 and let $\hat{\theta}_n$ be a minimizer of $\lambda_n(\theta)$. Show that

$$\hat{\theta}_n \xrightarrow{p} \theta_0.$$

Hint: you may use the fact that for any convex function $f(x)$ with unique minimizer x^*,

$$h(\delta) \triangleq \inf_{|x-x^*|>\delta} (f(x) - f(x^*)) > 0$$

for any $\delta > 0$. (This is not hard to prove, but you can use this directly.)

(Cont.) Let $X_1, \ldots, X_n \in \mathbb{R}^p$ be i.i.d. random variables, and ρ be a convex function. Consider the following location estimator

$$\hat{\theta}_n = \arg\min_\theta \frac{1}{n} \sum_{i=1}^n \rho(X_i - \theta).$$

(e) Use the previous result to show that

$$\hat{\theta}_n \xrightarrow{p} \arg\min_\theta \mathbb{E}\rho(X_1 - \theta)$$

provided the right-handed side exists and is finite.

(f) Let $\rho(x) = ||x||_2^2 = \sum_{i=1}^p x_i^2$, and assume $\mathbb{E}||X_1||_2^2 < \infty$. Find the limit of $\hat{\theta}_n$;

(g) Assume that $p = 1$, X_1 has continuous distribution with $\mathbb{E}|X_1| < \infty$ and let $\rho(x) = |x|$. Find the limit of $\hat{\theta}_n$;

(h) Assume that $p = 1$, X_1 has continuous distribution with $\mathbb{E}|X_1| < \infty$. Let $\tau \in (0, 1)$ and define $\rho(x)$ as the following piecewise linear function

$$\rho(x) = \tau x I(x \geq 0) + (\tau - 1) x I(x < 0).$$

Find the limit of $\hat{\theta}_n$;

(Cont.) Now consider a penalized least square estimator

$$\hat{\theta}_n^\lambda = \arg\min_\theta \frac{1}{n} \sum_{i=1}^n ||X_i - \theta||_2^2 + \lambda G(\theta)$$

where G is a convex function on $\mathbb{R}^p$.

(i) Let $G(\theta) = ||\theta||_2^2$, what is the limit of $\hat{\theta}_n^\lambda$?

(j) Let $G(\theta) = ||\theta||_1 \triangleq \sum_{i=1}^p |\theta_i|$, what is the limit of $\hat{\theta}_n^\lambda$?

Problem 9.2

In this problem, we will explore the limit distribution of the minimizer of random convex functions.

(a) (A finer version of problem 9.1(d)) Let $\lambda_n(\theta)$ and $\tilde{\lambda}_n(\theta)$ be two sequences of random convex functions. Assume that $\tilde{\lambda}_n(\theta)$ has a unique minimizer $\tilde{\theta}_n$ for each n (note that $\tilde{\theta}_n$ is a random variable) and let θ_n be a minimizer of $\lambda_n(\theta)$. Let

$$\Delta_n(\delta) = \max_{||\theta - \tilde{\theta}_n| \leq \delta} |\lambda_n(\theta) - \tilde{\lambda}_n(\theta)|$$

and

$$h_n(\delta) = \inf_{||\theta - \tilde{\theta_n}|| > \delta} \tilde{\lambda}_n(\theta) - \tilde{\lambda}_n(\tilde{\theta}_n).$$

Show that

$$P(||\theta_n - \tilde{\theta}_n|| > \delta) \leq P\left(2\Delta_n(\delta) \geq h_n(\delta)\right).$$

Hint: Modify the proof of Problem 9.1(d).

(b) Assume that $\lambda_n(\theta)$ is convex and can be written as

$$\lambda_n(\theta) = \frac{1}{2}\theta^T V \theta - \theta^T U_n + C_n + r_n(\theta)$$

where $V \in \mathbb{R}^p$ is a positive-definite matrix, $U_n = O_p(1)$, C_n be some random variables only depending on n and $r_n(\theta)$ is a random function such that for any given θ, $r_n(\theta) = o_p(1)$. Show that

$$|\theta_n - V^{-1}U_n| \xrightarrow{p} 0,$$

where $\theta_n = \arg\min_\theta \lambda_n(\theta)$.

Hint: Let $\tilde{\lambda}_n(\theta) = \frac{1}{2}\theta^T V \theta - \theta^T U_n + C_n$ *and use Problem 9.1(c) and Problem 9.2(a).*

(c) Under the settings of (b), assume further that $U_n \xrightarrow{d} U$ for some random variable U. Show that

$$\theta_n \xrightarrow{d} V^{-1}U.$$

(Cont.) Let $X_1, \ldots, X_n \in \mathbb{R}^p$ be i.i.d. random variables, and ρ be a convex function. Consider the following location estimator

$$\hat{\theta}_n = \arg\min_\theta \frac{1}{n}\sum_{i=1}^n \rho(X_i - \theta).$$

Assume that $\theta_0 = \arg\min_\theta \mathbb{E}\rho(X_1 - \theta)$ exists and is finite, then from Problem 9.1(e) we know that $\hat{\theta}_n \xrightarrow{p} \theta_0$.

(d) Suppose $p = 1$ and ρ has a bounded third derivative, prove that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\left(0, \frac{\mathbb{E}\rho'(X_1 - \theta_0)^2}{\left(\mathbb{E}\rho''(X_1 - \theta_0)\right)^2}\right).$$

(e) Let $\rho(x) = ||x||_2^2$, identify the limiting distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$.

(Cont.) Let $p = 1$, and $\rho(x) = |x|$. Then it is easy to show that $\hat{\theta}_n$ is the sample median. Now we establish the asymptotic normality for $\hat{\theta}_n$ using the previous results.

(f) (Knight's entity) Show that for any u and t,

$$|u - t| - |u| = -t(I(u > 0) - I(u \leq 0)) + 2\int_0^t (I(u \leq s) - I(u \leq 0))ds.$$

(g) Let θ_0 be the median of X_1, $Y_i = X_i - \theta_0$ and

$$\lambda_n(\eta) = \sum_{i=1}^n \left(\left|Y_i - \frac{\eta}{\sqrt{n}}\right| - \left|Y_i\right| \right).$$

Show that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \arg\min_\eta \lambda_n(\eta).$$

(h) Using Knight's entity,

$$\lambda_n(\eta) = -\frac{\eta}{\sqrt{n}} \sum_{i=1}^n (I(Y_i > 0) - I(Y_i \leq 0)) + 2\sum_{i=1}^n \int_0^{\frac{\eta}{\sqrt{n}}} \left(I\left(Y_i \leq s\right) - I\left(Y_i \leq 0\right)\right) ds \triangleq \Lambda_{1n} + \Lambda_{2n}.$$

Assume that X_1 has a density at θ_0, denoted by $f(\theta_0)$. Show that $\Lambda_{1n} = \eta U_n$, where

$$U_n \xrightarrow{d} N(0, 1).$$

(i) Show that

$$\Lambda_{2n} = f(\theta_0)\eta^2 + r_n(\eta)$$

where $r_n(\eta) = o_p(1)$ for any given η.

Hint: Calculate $\mathbb{E}\Lambda_{2n}$ *and* $Var(\Lambda_{2n})$. *You can assume the interchangeability of integral and expectation.*

(j) Identify the limiting distribution of $\hat{\theta}_n$.

Problem 9.3
Jacknife and Bootstrap are two popular methods to estimate the bias and the variance of a statistic. In this problem we will compare them in different situations.

Let $\hat{\theta} = \hat{\theta}(F_n)$ be an estimator of $\theta = \theta(F)$ where F_n is the empirical distribution of i.i.d. samples $X_1, \ldots, X_n \sim F$. The bias of $\hat{\theta}$ is defined as $\mathbb{E}\hat{\theta} - \theta$. Let $\hat{\theta}_{(i)} = \hat{\theta}(F_n^{(i)})$ where $F_n^{(i)}$ is the empirical distribution of $X_j, j \neq i$. For example, if $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n X_i$, then $\hat{\theta}_{(i)} = \frac{1}{n-1}\sum_{j\neq i} X_j$. The Jacknife estimate of bias is defined as

$$\widehat{bias}_{jack} = (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta})$$

where $\hat{\theta}_{(\cdot)} = \sum_{i=1}^n \hat{\theta}_{(i)}/n$. The Jacknife estimate of variance is defined as

$$\widehat{var}_{jack} = \frac{n-1}{n}\sum_{i=1}^n (\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2.$$

Another definition is based on pseudo observations $\tilde{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{(i)}$. The estimate of bias is defined by

$$\widehat{bias}'_{jack} = \hat{\theta} - \tilde{\theta}_{(\cdot)}$$

where $\tilde{\theta}_{(\cdot)} = \frac{1}{n}\sum_{i=1}^n \tilde{\theta}_i$ and the estimate of variance is defined by

$$\widehat{var}'_{jack} = \frac{1}{n(n-1)}\sum_{i=1}^n (\tilde{\theta}_i - \tilde{\theta}_{(\cdot)})^2.$$

In the following context, we use $bias_F(\theta)$ to represent the bias of $\hat{\theta}$, i.e. $\mathbb{E}_F(\hat{\theta}) - \theta$ and $var_F(\theta)$ to represents the variance of $\hat{\theta}$.

(a) Show that two definitions agree with each other, i.e.

$$\widehat{bias}_{jack} = \widehat{bias}'_{jack}, \quad \widehat{var}_{jack} = \widehat{var}'_{jack}.$$

(b) Let $\theta = \mathbb{E}X_1$ and $\hat{\theta} = \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and assume $\sigma^2 = Var(X_1) < \infty$. Calculate $bias_F(\hat{\theta})$, $var_F(\hat{\theta})$, $\widehat{bias}_{jack}(\hat{\theta})$ and $\widehat{var}_{jack}(\hat{\theta})$. Then compute $\mathbb{E}_F\widehat{bias}_{jack}(\hat{\theta}) - bias_F(\hat{\theta})$ and $\mathbb{E}_F\widehat{var}_{jack}(\hat{\theta}) - var_F(\hat{\theta})$.

(c) Let $\theta = Var(X_1) < \infty$ and $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$. Calculate $bias_F(\hat{\theta})$ and $\widehat{bias}_{jack}(\hat{\theta})$. What is the difference between $bias_F(\hat{\theta})$ and $\mathbb{E}_F\widehat{bias}_{jack}(\hat{\theta})$?

(d) Let $X_1 \sim Unif(0,1)$, $\theta = 0$ and $\hat{\theta} = X_{(1)}$. Calculate $bias_F(\hat{\theta})$ and $\widehat{bias}_{jack}(\hat{\theta})$. Does the ratio of $\mathbb{E}_F\widehat{bias}_{jack}(\hat{\theta})$ and $bias_F(\hat{\theta})$ converge to 1?

Hint: you may use Homework 7.5 (a).

(e) Under the settings of (d), calculate $var_F(\hat{\theta})$ and $\widehat{var}_{jack}(\hat{\theta})$. Does the ratio of $\mathbb{E}_F\widehat{var}_{jack}(\hat{\theta})$ and $var_F(\hat{\theta})$ converge to 1?

(f) Let $X_1 \sim exp(1)$ and $n = 2m$ be an even integer. Denote the median of X_1 by θ and then the sample median $\hat{\theta} = (X_{(m)} + X_{(m+1)})/2$ where $X_{(m)}$ and $X_{(m+1)}$ are the m-th largest and $(m+1)$-th largest order statistics. Calculate $var_F(\hat{\theta})$ and $\widehat{var}_{jack}(\hat{\theta})$. Does the ratio of $\mathbb{E}_F\widehat{var}_{jack}(\hat{\theta})$ and $var_F(\hat{\theta})$ converge to 1?

Hint: you may use the hint of Homework 3.3 (d).

(Cont.) Similar to Jacknife, we can define bootstrap bias and bootstrap variance. Let $\hat{\theta}^*$ be the statistic evaluated on a bootstrap sample $x_1^*, \ldots, x_{(n)}^*$. Then the bootstrap bias can be defined as

$$\widehat{bias}_{boot}(\hat{\theta}) = \mathbb{E}(\hat{\theta}^* | X_1, \ldots, X_n) - \hat{\theta},$$

and the bootstrap variance is defined as

$$\widehat{var}_{boot}(\hat{\theta}) = Var(\hat{\theta}^* | X_1, \ldots, X_n)$$

(g) Let $\hat{\theta} = \bar{X}$. Show that $\hat{\theta}^*$ can be written as a weighted sum with random weight

$$\hat{\theta}^* = \frac{1}{n}\sum_{i=1}^{n} W_i X_i,$$

where $(W_1, \ldots, W_n) \sim Multi\left(n; \frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$.

(h) Under the settings of (b), calculate $bias_F(\hat{\theta})$, $var_F(\hat{\theta})$, $\widehat{bias}_{boot}(\hat{\theta})$ and $\widehat{var}_{boot}(\hat{\theta})$. Then compute $\mathbb{E}_F\widehat{bias}_{boot}(\hat{\theta}) - bias_F(\hat{\theta})$ and $\mathbb{E}_F\widehat{var}_{boot}(\hat{\theta}) - var_F(\hat{\theta})$. Compare these results to part (b).

(i) Under the settings of (c), calculate $bias_F(\hat{\theta})$ and $\widehat{bias}_{boot}(\hat{\theta})$. What is the difference between $bias_F(\hat{\theta})$ and $\mathbb{E}_F\widehat{bias}_{boot}(\hat{\theta})$? Compare it to part (c).

(j) Under the settings of (d), calculate $bias_F(\hat{\theta})$ and $\widehat{bias}_{boot}(\hat{\theta})$. Does the ratio of $\mathbb{E}_F\widehat{bias}_{boot}(\hat{\theta})$ and $bias_F(\hat{\theta})$ converge to 1? Compare it to part (d).

Problem 9.4

One of the most important task in modern statistics is to bound the tail probability or expectation of the maximum. In this problem we will explore the properties of maximum in different situations.

(a) Let X be a non-negative random variable such that

$$P(X \geq t) \leq c_1 e^{-c_2 t^\alpha}$$

for some $c_1 > 1, c_2 > 0, \alpha \geq 1$ and any $t > 0$. Show that

$$\mathbb{E}X = \int_0^\infty P(X \geq t)dt \leq \left(\frac{\log c_1}{c_2}\right)^{\frac{1}{\alpha}} \left(1 + \frac{1}{\alpha \log c_1}\right).$$

Hint: you may separate the integral region into two parts $\int_0^{t_*}$ *and* $\int_{t_*}^\infty$ *where* $c_1 e^{-c_2 t_*^\alpha} = 1$.

(b) Let $X_1, \ldots, X_n$ be n mean-zero σ-sub-Gaussian random variables, i.e.

$$\mathbb{E}e^{\lambda X_i} \leq e^{\frac{\lambda^2 \sigma^2}{2}}, \quad \forall \lambda \in \mathbb{R}.$$

Use part (a) to give a bound for $\mathbb{E} \max_i |X_i|$.

(c) Let $\psi : \mathbb{R} \to \mathbb{R}^+$ be a convex functionand strictly increasing on $\mathbb{R}^+$. Let $X_1, \ldots, X_n$ are n random variables. Show that

$$\mathbb{E} \max_i X_i \leq \inf_{\lambda > 0} \frac{1}{\lambda} \psi^{-1} \left(\sum_{i=1}^n \mathbb{E}\psi(\lambda X_i) \right).$$

Hint: use Jensen's inequality.

(d) Let $X_1, \ldots, X_n$ be n mean-zero σ-sub-Gaussian random variables. Show that

$$\mathbb{E} \max_i X_i \leq \sqrt{2\sigma^2 \log n}, \quad \mathbb{E} \max_i |X_i| \leq \sqrt{2\sigma^2 \log 2n}.$$

Compare it to the result in part (b).

Hint: use part (c) and let $\psi(x) = e^x$.

(e) Let $X_1, \ldots, X_n$ be n mean-zero (σ, b)-sub-exponential random variables, i.e.

$$\mathbb{E}e^{\lambda X_i} \leq e^{\frac{\lambda^2 \sigma^2}{2}}, \quad \forall |\lambda| < \frac{1}{b}$$

Show that

$$\mathbb{E} \max_i X_i \leq b \log n + \frac{\sigma^2}{2b}, \quad \mathbb{E} \max_i |X_i| \leq b \log 2n + \frac{\sigma^2}{2b}$$

(f) Let $X_1, \ldots, X_n$ be n random variables such that $(\mathbb{E}|X_i|^q)^{\frac{1}{q}} \leq M$ for some $q \geq 1$. Show that

$$\mathbb{E} \max_i |X_i| \leq M n^{\frac{1}{q}}.$$

Hint: use part (c) and let $\psi(x) = x^q$.

(g) Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(0,1)$. From part (d), we already know that

$$\mathbb{E} \max_i |X_i| \leq \sqrt{2\log(2n)}.$$

Show that

$$\lim_{n\to\infty} \frac{\mathbb{E}\max_i |X_i|}{\sqrt{2\log 2n}} = 1.$$

Hint: you may use the following inequality

$$\left(\frac{1}{z} - \frac{1}{z^3}\right) e^{-\frac{z^2}{2}} \leq \int_z^\infty e^{-\frac{x^2}{2}}\, dx \leq \frac{1}{z} e^{-\frac{z^2}{2}}.$$

This is not hard to prove but you can use it directly without proof.

(h) Let $X_1, \ldots, X_n$ be independent σ-sub-Gaussian random variables. Denote $X = (X_1, \ldots, X_n)$, then

$$||X||_2 \triangleq \sqrt{X_1^2 + \ldots + X_n^2} = \sup_{\alpha \in \mathbb{S}^{n-1}} \alpha^T X$$

where $\mathbb{S}^{n-1}$ is the n-dimensional unit sphere, i.e.

$$\mathbb{S}^{n-1} = \{\alpha \in \mathbb{R}^n : ||\alpha||_2 = 1\}.$$

Let $\mathcal{T}$ be a finite subset of $\mathbb{S}^{n-1}$ such that for any $\alpha \in \mathbb{S}^{n-1}$, there exists $\tilde{\alpha} \in \mathcal{T}$ such that $||\alpha - \tilde{\alpha}||_2 \leq \frac{1}{2}$. Show that

$$||X||_2 \leq 2 \max_{\alpha \in \mathcal{T}} \alpha^T X.$$

(i) Suppose we know that there exists a subset $\mathcal{T}$ satisfying the above condition and $|\mathcal{T}| \leq 5^n$, show that

$$\mathbb{E}||X||_2 \leq 4\sigma\sqrt{n}.$$

Hint: use part (d).

(j) Can you show that there exists a subset $\mathcal{T}(\varepsilon) \subset \mathbb{S}^{n-1}$ such that for any $\alpha \in \mathbb{S}^{n-1}$ there exists $\tilde{\alpha} \in \mathcal{T}(\varepsilon)$ such that $||\alpha - \tilde{\alpha}|| < \varepsilon$ and

$$|\mathcal{T}(\varepsilon)| \leq c(\varepsilon)^n$$

for some function $c(\varepsilon)$? (e.g. $c = 1 + 2/\varepsilon$.)

Problem 9.5
For given random variable X, the sub-Gaussian norm $||X||_{\psi_2}$ is defined as

$$||X||_{\psi_2} = \sup_p p^{-\frac{1}{2}} \left(\mathbb{E}|X|^p\right)^{\frac{1}{p}},$$

and the sub-exponential norm $||X||_{\psi_1}$ is defined as

$$||X||_{\psi_1} = \sup_p p^{-1} \left(\mathbb{E}|X|^p\right)^{\frac{1}{p}},$$

where the supremum is taken over all positive integers p.

(a) Show that if X is mean zero and σ-sub-Gaussian, then $||X||_{\psi_2} \le c_1\sigma$ for some universal constant c_1.

(b) Show that if $||X||_{\psi_2} < \infty$, then X is $c_2||X||_{\psi_2}$-sub-Gaussian.

Hint: use Taylor expansion.

(c) Show that if X is mean zero and (σ, b)-sub-exponential, then $||X||_{\psi_1} \le c_3 \max\{\sigma, b\}$ for some universal constant c_3.

(d) Show that if $||X||_{\psi_1} < \infty$, then X is $(c_4||X||_{\psi_1}, c_4||X||_{\psi_1})$-sub-exponential for some universal constant c_4.

Hint: use Taylor expansion.

(e) Show that

$$||X||_{\psi_2}^2 \le ||X^2||_{\psi_1} \le 2||X||_{\psi_2}^2.$$

Conclude that if X is σ-sub-Gaussian, then X^2 is $(c_5\sigma^2, c_5\sigma^2)$-sub-exponential for some universal constant c_5.

(f) Let $X_1, \ldots, X_n$ be independent sub-exponential random variables with parameters (σ_i, b_i). Show that $\sum_{i=1}^n X_i$ is sub-exponential with parameters

$$(\sigma^*, b^*) = \left(\sqrt{\sum_{i=1}^n \sigma_i^2}, \max_i b_i\right).$$

(Cont.) A random vector $Z \in \mathbb{R}^p$ is called σ-sub-Gaussian if for any $\alpha \in \mathbb{S}^{p-1}$, $\alpha^T Z$ is σ-sub-Gaussian. Let $x_1, \ldots, x_n \in \mathbb{R}^p$ be i.i.d. zero-mean σ-sub-Gaussian random vectors with covariance matrix $\mathbb{E}x_1x_1^T = \Sigma$. The sample covariance matrix is defined as

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n x_i x_i^T.$$

(g) For given $\alpha \in \mathbb{S}^{p-1}$, show that $\mathbb{E}\alpha^T\hat{\Sigma}\alpha = \alpha^T\Sigma\alpha$ and $\alpha^T\hat{\Sigma}\alpha$ is $(c_6\sigma^2/\sqrt{n}, c_6\sigma^2/n)$-sub-exponential for some universal constant c_6, i.e.

$$\mathbb{E}e^{\lambda(\alpha^T\hat{\Sigma}\alpha - \alpha^T\Sigma\alpha)} \leq e^{\frac{\lambda^2(c_6\sigma^2)^2}{2n}}, \quad \forall|\lambda| < \frac{n}{c_6\sigma^2}.$$

Hint: you may use part (e), part (f) as well as the properties of sum of independent sub-exponential random variables in the textbook

(h) Recall that the maximum eigenvalue of a symmetric matrix S (or the operator norm of S) can be expressed as

$$||S||_{op} = \lambda_{max}(S) = \sup_{\alpha\in\mathbb{R}^{p-1}} |\alpha^T S\alpha|.$$

Let $\varepsilon < \frac{1}{3}$ and $\mathcal{T}(\varepsilon)$ be the set defined in Problem 9.4 (j) (with dimension n replaced by p), show that

$$||\hat{\Sigma} - \Sigma||_{op} \leq \frac{1}{1-3\varepsilon} \max_{\alpha\in\mathcal{T}(\varepsilon)} \left|\alpha^T\hat{\Sigma}\alpha - \alpha^T\Sigma\alpha\right|.$$

(i) Show that

$$\mathbb{E}||\hat{\Sigma} - \Sigma||_{op} \leq \frac{c_7 p\sigma^2}{n} + c_8\sigma^2,$$

for some universal constants c_7 and c_8.

Hint: you may use problem 9.4 (e), problem 9.4 (j) and problem 9.5 (g).

(j) Show that for some universal constant c_9,

$$\mathbb{E}e^{\lambda||\hat{\Sigma}-\Sigma||_{op}} \leq e^{\frac{\lambda^2(c_6\sigma^2)^2}{n} + c_9 p}, \quad \forall|\lambda| < \frac{n}{c_6\sigma^2}.$$