Final

SDS384

Spring 2019

This exam has five short and three long questions. You will have to answer **four short questions**, **two long questions**. The assigned points are noted next to each question; the total number of points is 50. You have 180 minutes to answer the questions.

Please answer all problems in the space provided on the exam. Use extra pages if needed. Of course, please put your name on extra pages.

Read each question carefully, **show your work** and **clearly present your answers**. Note, the exam is printed two-sided - please don't forget the problems on the even pages!

Good Luck!

Name: _____

UTeid: _____

1 Short questions (28 points)

Please answer any four of the short questions.

1. (7 pts) Consider a sequence of iid random variables $X_1, \ldots, X_n$ such that $X_i \sim Beta(\theta, 1)$, where $\theta > 0$. Let $\bar{X}_n$ denote the sample mean. The method of moments estimator of θ is $\hat{\theta}_n = \bar{X}_n/(1-\bar{X}_n)$. Derive the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$. *You can use the fact that a* $Beta(\beta, 1)$ *random variable has mean* $\beta/(1+\beta)$ *and variance* $\frac{\beta}{(\beta+1)^2(\beta+2)}$. Recall that the expectation of a $beta(\beta, 1)$ random variable is $\theta/(1+\theta)$. So $\bar{X}_n \xrightarrow{P} \theta/(1+\theta)$ and variance $\sigma^2 = \frac{\theta}{(\theta+1)^2(\theta+2)}$. Now

$$\begin{aligned}\sqrt{n}(\hat{\theta}_n - \theta) &= \sqrt{n}\left(\frac{\bar{X}_n}{1-\bar{X}_n} - \theta\right) \\ &= \sqrt{n}(1+\theta)\frac{\bar{X} - \frac{\theta}{1+\theta}}{1-\bar{X}}\end{aligned}$$

Using CLT we have $\sqrt{n}(\bar{X} - \frac{\theta}{1+\theta}) \xrightarrow{d} N(0, \sigma^2)$. $1-\bar{X} \xrightarrow{P} 1/(1+\theta)$. So $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \theta(\theta+1)^2/(\theta+2))$.

2. (7 pts) What is the VC dimension of the union of k intervals on the real line? Show your calculations. The answer is 2k. It is not hard to see that any 2k distinct points on the real line can be shattered using k intervals: it suffices to shatter each of the k pairs of consecutive points with an interval. Assume now that $2k+1$ distinct points $x_1 < \cdots < x_{2k+1}$ are given. For any $i \in [1, 2k+1]$, label x_i with $(-1)^i + 1$, that is alternatively label points with 1 or -1. This leads to $k+1$ points labeled positively and requires $2k+1$ intervals to shatter the set since no interval can contain two consecutive points. Thus, no set of $2k+1$ points can be shattered by k intervals and the VC dimension of the union of k intervals is $2k$.

3. (7 pts) Let $X_1, X_2, \ldots, X_n$ be i.i.d. samples of random variable with density f on the real line. A standard estimate of f is the kernel density estimate

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

where $K : \Re \to [0, \infty)$ is a kernel function satisfying $\int_{-\infty}^{\infty} K(t)dt = 1$, and h is a bandwidth parameter. We will measure the quality of $\hat{f}$ using $\|\hat{f} - f\|_1 := \int_{-\infty}^{\infty} |\hat{f}(t) - f(t)|dt$. Prove that:

$$P(\|\hat{f} - f\|_1 \geq E\|\hat{f} - f\|_1 + \delta) \leq e^{-cn\delta^2},$$

where c is some constant. This seems like something suited for McDiarmid or the bounded differences inequality. So we will first calculate how big the differences are.

$$\begin{aligned}
&|\|\hat{f}_{X_1,\ldots,X_n}(t) - f(t)\|_1 - \|\hat{f}_{X_1',\ldots,X_n}(t) - f(t)\|_1| \\
&\leq \int |\hat{f}_{X_1,\ldots,X_n}(t) - \hat{f}_{X_1',\ldots,X_n}(t)|dt \\
&\leq 1/nh \int \left| K\left(\frac{x - X_i}{h}\right) - K\left(\frac{x - X_i'}{h}\right) \right| dt \leq 2/n
\end{aligned}$$

So, now we can use McDiarmid's inequality to get the above result.

4. (7 pts) Consider n i.i.d random variables $X_1, \ldots, X_n$ with mean μ. We are interested in estimating $E(X_1 - \mu)^3$. Construct a U statistic for estimating this. Explicitly write down the kernel.

   Let $g(x_1, x_2, x_3) = (x_1 - x_2)^2(x_1 - x_3)$. We have: Assume WLOG $\mu = 0$. $E[g(x_1, x_2, x_3)] = E(x_1^3 - 2x_1^2x_2 - 2x_1x_2x_3 + x_2^2(x_1 - x_3)) = E[x_1^3]$.

   But since this is not symmetric we can symmetrize it as:

   $h(x_1, x_2, x_3) = \frac{\sum_{ijk \in \Pi_3} g(x_i, x_j, x_k)}{6}$, where Π_3 is all permutations of $\{1, 2, 3\}$.

5. (7 pts) Consider $X_1, \ldots, X_n$, n independent Uniform($[a, b]$) random variables. Obtain an upper bound on $E[\max_i X_i]$ in terms of a, b and n.

If anyone gives the formula for subgaussian RVS and have $E[\max_i X_i] \leq (a + b)/2 + \sqrt{2 \log n}(b - a)^2/4$, give them 3 points, since this will be far larger than b as n grows.

$P(\max_i X_i < t) = \left(\frac{t-a}{b-a}\right)^n$. So

$$\begin{aligned} E[\max_i (X_i - a)] &= \int_a^b \left(1 - \left(\frac{t-a}{b-a}\right)^n\right) dt \\ &= (b-a) \int_0^1 (1 - x^n)\, dx \\ &= (b-a)\left(1 - \frac{1}{n+1}\right) \\ &= (b-a)\frac{n}{n+1} \\ E[\max_i X_i] &= a + (b-a)\frac{n}{n+1} = b - \frac{b-a}{n+1} \end{aligned}$$

2 Long questions (22 points)

Please answer any two of the long questions.

1. (11 pts) Let $X_1, \ldots, X_n$ be independent random variables with $X_n \sim N(0, \sigma_n^2)$, where $\sigma_k^2 = 2^{-(k-1)}$.

   (a) (5 pts) Does the Lindeberg condition hold? Give a proof. $B_n^2 = 2(1-2^{-n}) \in [1,2]$, and so the LC boils down to seeing if the following goes to zero.

$$\begin{aligned}\frac{.5}{1-2^{-n}} \sum_i E[X_i^2 1(|X_i| \geq \epsilon B_n)] &\geq 1/2 \sum_i E[X_i^2 1(|X_i| \geq \epsilon)] \\ &\geq 1/2 E[X_1^2 1(|X_1| \geq \epsilon)] \\ &\overset{\epsilon=1}{\geq} 1/s2E[X_1^2 1(|X_1| \geq 1)] > 0\end{aligned}$$

   So it does not hold.

(b) (4 pts) Does $\sum_i X_i$ converge to a normal distribution? Give a proof of your answer. If yes, obtain the parameters of the limiting normal distribution. Of course, the sum of n independent normals (S_n) is normally distributed with mean zero and variance $B_n^2 = 2(1 - 2^{-n})$. So we have $S_n/B_n \xrightarrow{d} N(0,1)$.

$$\begin{aligned}
\frac{S_n}{B_n} - \frac{S_n}{\sqrt{2}} &= -S_n \frac{1}{\sqrt{2}}(1 - (1 - 2^{-n})^{-1/2}) \\
\left|\frac{S_n}{B_n} - \frac{S_n}{\sqrt{2}}\right| &= \left|\frac{S_n}{B_n}\right| \sqrt{(1 - 2^{-n})}(1 - (1 - 2^{-n})^{-1/2}) \\
&= |Z| o(1) = o_P(1)
\end{aligned}$$

So $\frac{S_n}{B_n} - S_n/\sqrt{2} \xrightarrow{P} 0$ and so $S_n \xrightarrow{d} N(0,2)$

(c) (2 pts) Write in a few sentences if this contradicts the Lindeberg-Feller theorem. It does not. Since $\max_j \sigma_{nj}^2/B_n^2 \to 1/2$, the Lindeberg condition is not necessary. This is why the LF theorem is not contradicted.

2. (11 pts) Let $X_1, \ldots, X_n$ be i.i.d random variables with mean μ and variance σ^2. We are interested in estimating the variance of the quantity $U = \frac{\sum_{i<j} X_i X_j}{\binom{n}{2}}$.

   (a) (6pts) Use the Efron-Stein inequality to obtain an upper bound on the variance of U. Let U_i' be U obtained from a sample $X_1^{i-1}, X_i', X_{i+1}^n$. Note that $U - U_i' = \frac{2\sum_{j\neq i}(X_i - X_i')X_j}{n(n-1)}$

$$\begin{aligned}
\text{var}(U) &\leq 1/2 \sum_j E(U - U_j')^2 = \frac{4}{2n(n-1)^2} E[(\sum_{j\neq i}(X_i - X_i')X_j)^2] \\
&= \frac{2}{n(n-1)^2} E[(X_i - X_i')^2 (\sum_{j\neq i} X_j)^2] \\
&= \frac{2}{n(n-1)^2} (2\sigma^2) E[(\sum_{j\neq i} X_j)^2] \\
&= \frac{4(1+o(1))}{n^3} \sigma^2 E[(n-1)(\sigma^2+\mu^2) + n(n-1)\mu^2] \\
&= \frac{4(1+o(1))}{n} \sigma^2 \mu^2
\end{aligned}$$

The step is true if σ, μ are fixed constants.

(b) (4pts) What is the asymptotic variance of U? From your class notes on U statistics, the variance is $4\sigma^2\mu^2/n$.

(c) (1pt) Is the upper bound obtained using the Efron Stein inequality tight? Explain in a few sentences.

It is. Since it matches the asymptotic variance as $n \to 0$.

3. (11 pts) Let $X_i \in \mathbb{R}^p, i = 1 \dots n$ be i.i.d random variables such that $X_i \sim N(0, I_{p\times p})$ where $I_{p\times p}$ is the $p \times p$ identity matrix. Define the function class $\mathcal{F} = \{f : \mathbb{R}^p \to \mathbb{R} | f(x_1, \dots, x_p) = \beta^T x; \|\beta\|_1 \leq R\}$, where $\beta^T x = \sum_{i=1}^p \beta_i x_i$. We will do a direct proof of $\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_i f(X_i) - E[f(X_1)]\right| \xrightarrow{P} 0$.

   (a) (4pts) Show that

$$\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_i f(X_i) - E[f(X_1)]\right| \leq \frac{R}{n}\|\sum_i X_i\|_\infty.$$

   Hint: you can use the fact that $\sup_{\|u\|_1\leq 1} |u^T v| = \|v\|_\infty$, *where* $u, v \in \mathbb{R}^p$.

$$\begin{aligned}\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_i f(X_i) - E[f(X_1)]\right| &= \sup_{\|\beta\|_1\leq R} \left|\frac{1}{n}\sum_i \beta^T X_i\right| \\ &= \frac{R}{n}\sup_{\|u\|_1\leq 1} \left|\frac{1}{n}\sum_i u^T X_i\right| \\ &= \frac{R}{n}\|\sum_i X_i\|_\infty\end{aligned}$$

(b) (2pts) Show that

$$\frac{R}{n}\|\sum_i X_i\|_\infty = \frac{R}{\sqrt{n}} \max_{1\le j\le p} Z_j,$$

where Z_j's are i.i.d standard normal random variables. Jack, there is a typo here it should be $|Z_j|$. So I would suggest a bit more lenient grading. If they prove Z_j thats fine.

Note that $\sum_i X_i/\sqrt{n} \sim N(0, I_{p\times p})$. This is easy to see because $\sum_i X_i(a)/\sqrt{n} \sim N(0,1)$ and since all X_i are generated from a gaussian with identity as covariance matrix, the coordinates of the sum will also be uncorrelated. Therefore,

$$\frac{R}{n}\|\sum_i X_i\|_\infty = \frac{R}{\sqrt{n}} \max_{1\le j\le p} |Z_j|,$$

(c) (5pts) Now show that, as long as $R\sqrt{\log p/n} \to 0$,

$$\sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_i f(X_i) - E[f(X_1)] \right| \xrightarrow{P} 0.$$

$$\begin{aligned} P(\max_j |Z_j| \ge \epsilon\sqrt{n}/R) &\le pP(|Z_1| \ge \epsilon\sqrt{n}/R) \\ &\le 2p\exp(-c\epsilon^2 n/R^2) \end{aligned}$$

So as long as $\log p \ll c\epsilon^2 n/R^2$, we have the failure probability going to zero. This is satisfied when,

$$R\sqrt{\log p/n} \to 0$$

If they show this without the absolute value, they should get full score.