Math 281A **2019 Fall Final** **Dec 12, 2019**

Name: ______________________ **PID:** ____________________

Do not turn the page until told to do so.

1. No calculators, tablets, phones, or other electronic devices are allowed during this exam.
2. Read each question carefully and answer each question completely.
3. Show all of your work. No credit will be given for unsupported answers, even if correct.
4. **Please answer questions within the spaces provided.** If you do need some more room, use the back side of the same piece of paper and clearly label the question.
5. If you are unsure of what a question is asking for, **do not hesitate to ask an instructor or course assistant for clarification**.
6. This exam has 6 pages.

Question	*Points Available*	*Points Earned*
1	10	
2	10	
3	10	
4	10	
5	15	
6	15	
TOTAL	70	

1. [10 points] Let X_n be the maximum of a random sample $Y_1, \ldots, Y_n$ from the density $p(x) = 2(1-x)I(0 \le x \le 1)$. Find constants a_n and b_n such that $b_n(X_n - a_n)$ converges in distribution to a non-degenerate limit.

2. [10 points] Let $Z_1, \ldots, Z_n$ be independent standard normal variables. Show that the vector $U = (Z_1, \ldots, Z_n)^\intercal / N$, where $N^2 = \sum_{i=1}^n Z_i^2$, is uniformly distributed over the unit sphere $\mathbb{S}^{n-1}$ in $\mathbb{R}^n$ in the sense that U and OU are identically distributed for every orthogonal transformation O of $\mathbb{R}^n$.

3. [10 points] Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed with mean μ and finite variance σ^2. Find the asymptotic distribution of $\bar{X}_n^2$ (after it is properly normalized), where $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$.

4. [10 points] Suppose $X_n \sim \text{binomial}(n, p)$, where $0 < p < 1$. (a). Find the asymptotic distribution of $g(X_n/n) - g(p)$, where $g(x) = \min\{x, 1-x\}$. (b) Show that $h(x) = \sin^{-1}(\sqrt{x})$ is a variance-stabilizing transformation for X_n/n. This is called the *arcsine transformation* of a sample proportion. **Hint**: $\frac{d}{du}\sin^{-1}(u) = 1/\sqrt{1-u^2}$.

5. [15 points] Suppose that we observe data in pairs $(X, Y) \in \mathbb{R}^d \times \{\pm 1\}$, where the data come from a logistic model with $X \sim P_0$ and $p_{Y|X}(y|x) = 1/(1 + e^{-y \cdot x^\intercal \theta_0})$. Define the log-loss function $\ell_\theta(y|x) = \log(1 + e^{-y \cdot x^\intercal \theta})$. Let $\hat{\theta}_n$ minimize the empirical logistic loss $L_n(\theta) = (1/n) \sum_{i=1}^n \ell_\theta(Y_i|X_i) = (1/n) \sum_{i=1}^n \log(1 + e^{-Y_i X_i^\intercal \theta})$ from pairs (X_i, Y_i) drawn from the logistic model with parameter θ_0. Assume that the covaraites $X_i \in \mathbb{R}^d$ are i.i.d. and satisfy $\mathbb{E}(X_i X_i^\intercal) = \Sigma > 0$ and $\mathbb{E}\|X_i\|_2^4 < \infty$.

   (a) Let $L(\theta) = \mathbb{E}_{\theta_0}\{\ell_\theta(Y|X)\}$ be the population logistic loss. Show that the second order derivative evaluated at θ_0 is positive definite.

   (b) Under these assumptions show that $\hat{\theta}_n$ is consistent estimator of θ_0 as $n \to \infty$. Provide details of your work.

   (c) Find the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$, provided that it is consistent. You may assume $d = 1$.

6. [15 points] Let $X_1, \ldots, X_n$ be a data sample of a continuous random variable X with distribution function F and density f. From the kernel density estimator $\hat{f}_h(x) = (1/n)\sum_{i=1}^n K_h(x - X_i)$, where $K_h(u) = K(u/h)/h$, one can construct a kernel estimator for the distribution function as $\hat{F}_h(x) = \int_{-\infty}^x \hat{f}_h(t)\mathrm{d}t$. Equivalently, we have

$$\hat{F}_h(x) = \frac{1}{n}\sum_{i=1}^n H\left(\frac{x - X_i}{h}\right), \quad \text{where} \quad H(x) = \int_{-\infty}^x K(t)\mathrm{d}t.$$

Assume K is non-negative, symmetric around 0 and integrates to 1. Under smoothness conditions, find the leading term of the mean integrated square error (MISE) of $\hat{F}_h$, that is, $\mathrm{MISE}(\hat{F}_h) = \int_{-\infty}^\infty \{\hat{F}_h(x) - F(x)\}^2 \mathrm{d}x$. What is the order of the optimal bandwidth?