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 7 pages.

<table>
<thead>
<tr>
<th>Question</th>
<th>Points Available</th>
<th>Points Earned</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>70</strong></td>
<td></td>
</tr>
</tbody>
</table>
1. [10 points] Let $X_n$ be the maximum of a random sample $Y_1, \ldots, Y_n$ from the density $p(x) = 2(1-x)I(0 \leq x \leq 1)$. Find constants $a_n$ and $b_n$ such that $b_n(X_n - a_n)$ converges in distribution to a non-degenerate limit.

**Solution:** $a_n = 1$ and $b_n = \sqrt{n}$. For any $x \leq 0$, we have

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

**Solution:** Denote $Z = (Z_1, \ldots, Z_n)^\top$, so $U = Z / \|Z\|_2$. Given that $Z \sim \mathcal{N}(0, I_n)$, it’s easy to see that for any orthogonal transformation $O$ of $\mathbb{R}^n$, $OZ \sim \mathcal{N}(0, I_n)$ and $\|OZ\|_2 = \|Z\|_2$. So

$$OU = \frac{OZ}{\|Z\|_2} = \frac{OZ}{\|OZ\|_2} \sim \frac{Z}{\|Z\|_2} = U.$$
3. [10 points] Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed with mean $\mu$ and finite variance $\sigma^2$. Find the asymptotic distribution of $\bar{X}_n^2$ (after it is properly normalized), where $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$.

**Solution:** By central limit theorem, we have

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

To get the asymptotic distribution of $\bar{X}_n^2$, we discuss two cases.

**Case 1:** $\mu \neq 0$. Applying delta method gives

$$\sqrt{n}(\bar{X}_n^2 - \mu^2) \xrightarrow{d} \mathcal{N}(0, 4\mu^2 \sigma^2).$$

**Case 2:** $\mu = 0$. Then

$$\frac{\sqrt{n}}{\sigma} \bar{X}_n \xrightarrow{d} \mathcal{N}(0,1).$$

Taking square on both sides with continuous mapping gives

$$\frac{n \bar{X}_n^2}{\sigma^2} \xrightarrow{d} \chi_1^2.$$
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:** $\frac{d}{du} \sin^{-1}(u) = 1/\sqrt{1-u^2}$.

**Solution:**

(a). $g(x)$ is not differentiable at $1/2$, so we need to discuss two cases.

**Case 1:** $p \neq 1/2$. Central limit theorem gives us

$$\sqrt{n}(X_n/n - p) \xrightarrow{d} \mathcal{N}(0, p(1-p)).$$

Since $g(x)$ is differentiable at $p$, and $g'(p) = \pm 1$. Applying delta method yields

$$\sqrt{n}(g(X_n/n) - g(p)) \xrightarrow{d} \mathcal{N}(0, p(1-p)).$$

**Case 2:** $p = 1/2$. Denote

$$Y_n = \frac{X_n}{n} - \frac{1}{2},$$

then by central limit theorem, we have $\sqrt{n}Y_n \to \mathcal{N}(0,1/4)$. Notice that

$$g(X_n/n) - g(1/2) = g(Y_n + 1/2) - 1/2 = \min\{1/2 + Y_n, 1/2 - Y_n\} - 1/2 = 1/2 - |Y_n| - 1/2 = -|Y_n|.$$

Since the absolute value function is continuous, by continuous mapping, we have

$$\sqrt{n}(g(X_n/n) - g(1/2)) = -\sqrt{n}|Y_n| \xrightarrow{d} -|Z|,$$

where $Z \sim \mathcal{N}(0,1/4)$.

The absolute value of normal distribution is called folded normal distribution.

(b). It can be found that

$$h'(x) = \frac{1}{2\sqrt{x(1-x)}}.$$ 

So $h'(p)^2p(1-p) = 1/4$ is a constant, and $h(x)$ is a variance-stabilizing transformation for $X_n/n$. 

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 $p_{Y|X}(y|x) = 1/(1 + e^{-y^t x_0})$. Define the log-loss function $\ell_\theta(y|x) = \log(1 + e^{-y^t x_\theta})$. Let $\hat{\theta}_n$ minimize the empirical logistic loss $L_n(\theta) = (1/n) \sum_{i=1}^n \ell_{\theta}(Y_i|X_i) = (1/n) \sum_{i=1}^n \log(1 + e^{-Y_i^t X_i^t \theta})$ from pairs $(X_i, Y_i)$ drawn from the logistic model with parameter $\theta_0$. Assume that the covaraites $X_i \in \mathbb{R}^d$ are i.i.d. and satisfy $E(X_i X_i^t) = \Sigma > 0$ and $E\|X_i\|^4_2 < \infty$.

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

**Solution:** We can interchange expectation and derivative since $\ell_\theta(\cdot)$ is smooth enough.

\[
\nabla L(\theta) = E[(-1/(1 + e^{X^t \theta}))Y X],
\]
\[
\nabla^2 L(\theta) = E[(e^{X^t \theta} / (1 + e^{X^t \theta})^2)XX^t].
\]

For any $u \in \mathbb{R}^d$, we have
\[
u^t \nabla^2 L(\theta) u = E[(e^{X^t \theta} / (1 + e^{X^t \theta})^2)(u^t X)^2] \geq 0.
\]

To further show that $\nabla^2 L(\theta)$ is positive definite, if there is a $u \in \mathbb{R}^d$ such that $u^t \nabla^2 L(\theta) u = 0$, then $u^t E(X X^t) u = 0$. This is a contradiction with $E(X_i X_i^t) = \Sigma$ being positive definite.

(b) Under these assumptions show that $\hat{\theta}_n$ is consistent estimator of $\theta_0$ as $n \to \infty$. Provide details of your work.

**Solution:** It can be checked that $\nabla L(\theta_0) = 0$ and since $\nabla^2 L(\theta)$ is positive definite, $\theta_0$ is the unique root of $\nabla L(\theta) = 0$. Now $\hat{\theta}_n$ is the MLE of $\theta_0$, $\theta$ is identifiable and $L_n(\theta) \xrightarrow{p} L(\theta)$ for every $\theta$ by weak law of large numbers. These conditions guarantee that $\hat{\theta}_n$ is consistent of $\theta_0$.

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

**Solution:** When $d = 1$, the Fisher information is
\[
I_{\theta_0} = E[X^2 e^{X \theta_0} / (1 + e^{X \theta_0})^2],
\]
and under regularity conditions, we have
\[
\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, 1/I_{\theta_0}).
\]
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 \( \hat{F}_h(x) = \int_{-\infty}^{x} \hat{f}_h(t) dt \). Equivalently, we have
\[
\hat{F}_h(x) = \frac{1}{n} \sum_{i=1}^{n} H\left(\frac{x - X_i}{h}\right), \quad \text{where} \quad H(x) = \int_{-\infty}^{x} K(t) dt.
\]
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,
\[
\text{MISE}(\hat{F}_h) = \int_{-\infty}^{\infty} E_f \{(\hat{F}_h(x) - F(x))^2\} dx.
\]
What is the order of the optimal bandwidth?

**Solution:** With integration by parts, change of variable and Taylor expansion, the bias is
\[
\mathbb{E}_f \hat{F}_h(x) - F(x) = \int H\left(\frac{x - t}{h}\right) dF(t) - F(x) = \frac{1}{h} \int F(t) K\left(\frac{x - t}{h}\right) dt - F(x)
\]
\[
\sim \frac{h^2}{2} f'(x) \int y^2 K(y) dy := h^2 B_f(x).
\]
The variance is
\[
\text{Var}_f \hat{F}_h(x) = \frac{1}{n} \int H^2\left(\frac{x - t}{h}\right) dF(t)
\]
\[
\sim \frac{1}{n} F(x)(1 - F(x)) - \frac{2h}{n} f(x) \int yH(y) K(y) dy
\]
\[
:= \frac{1}{n} F(x)(1 - F(x)) - \frac{h}{n} V_f(x).
\]
Combining the above two display gives
\[
\text{MISE}(\hat{F}_h) \sim h^4 \int B_f^2(x) dx + \frac{1}{n} \int F(x)(1 - F(x)) dx - \frac{h}{n} \int V_f(x) dx,
\]
and the optimal choice of \( h \) is
\[
h = \left( \frac{\int V_f(x) dx}{4n \int B_f^2(x) dx} \right)^{1/3} \sim n^{-1/3}.
\]