Final Exam IEDA 5270

Name ______________________ Student ID ______________________

Question 1 (20 points)
Let $(X_1, \ldots, X_n)$ be a random sample from the Poisson distribution truncated at 0, i.e.,

$$P_\theta(X = x) = (e^\theta - 1)^{-1}\theta^x/x!, \quad x = 1, 2, \ldots, \quad \text{and} \quad \theta > 0.$$

(a) (10 points) Find the UMVUE of θ when $n = 1$.

(b) (10 points) Find the UMVUE of θ when $n = 2$. [Hint: First, show that $T = X_1 + X_2$ is complete and sufficient for θ. Then, find a $h(T)$ which is unbiased for θ. You may want to use the Taylor expansion of $e^\theta - 1$.]

Question 2 (15 points)
Suppose X has an exponential distribution with rate λ, i.e.,

$$p_\lambda(x) = \lambda e^{-\lambda x}, \forall x \geq 0.$$

(a) (5 points) Determine the most powerful test of "$H_0 : \lambda = 1$" versus "$H_1 : \lambda = 2$" with level $\alpha = 0.05$.

(b) (10 points) Determine the most powerful test of "$H_0 : \lambda = 1$ or 4" versus "$H_1 : \lambda = 2$" with level $\alpha = 0.05$.

Question 3 (20 points)
Let $(X_1, \ldots, X_n)$ be a random sample from $N(\mu, \sigma^2)$.

(a) (10 points) Suppose that $\sigma^2 = \gamma\mu^2$ with unknown $\gamma > 0$ and unknown $\mu \in \mathbb{R}$. Obtain a confidence set for γ with confidence coefficient $1 - \alpha$ by inverting the acceptance regions of likelihood ratio tests for $H_0 : \gamma = \gamma_0$ versus $H_1 : \gamma \neq \gamma_0$.

(b) (10 points) Repeat part (a) when $\sigma^2 = \gamma\mu$ with unknown $\gamma > 0$ and unknown $\mu > 0$.

Question 4 (20 points)
Zero-Inflated Poisson (ZIP) regression is a model for count data with excess zeros. In a ZIP, the response variable Y follows

$$Y_i \sim \begin{cases} 0 & \text{if } Z_i = 1 \\ \text{Poisson}(\lambda_i) & \text{if } Z_i = 0 \end{cases}$$

where

$$Z_i \sim \text{Bernoulli}(p_i).$$

Recall that Poisson pmf is given by $P_\lambda(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$. The regression model is specified by parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ such that

$$\log(\lambda_i) = \boldsymbol{b}_i^\top \boldsymbol{\beta}, \quad \text{logit}(p_i) = \log(p_i/(1 - p_i)) = \boldsymbol{g}_i^\top \boldsymbol{\gamma}.$$

(a) (10 points) Suppose we can observe the variable Z_i's, write down the log-likelihood with the complete data $(\boldsymbol{Y}, \boldsymbol{Z}; \boldsymbol{B}, \boldsymbol{G})$ and find the MLE. [Here, $\boldsymbol{B}$ is the design matrix consists of $\boldsymbol{b}_i$ and $\boldsymbol{G}$ is the design matrix consists of $\boldsymbol{g}_i$.]

(b) (10 points) Suppose Z_i is not observed, write down the incomplete log-likelihood with the incomplete data $(\boldsymbol{Y}; \boldsymbol{B}, \boldsymbol{G})$ and describe an Expectation-Maximization algorithm to solve an approximate MLE.

Question 5 (25 points)

James-Stein shrinkage estimator. Let $\boldsymbol{X} = (X_1, \ldots, X_d) \sim N(\boldsymbol{\theta}, I_d)$ be a multivariate (with $d \geq 3$) normal r.v. A natural estimator for $\boldsymbol{\theta}$ is just $\boldsymbol{X}$. This is the MLE and an unbiased estimator with minimum variance. James and Stein show the following shocking result: Consider the estimator

$$\boldsymbol{\delta}^{\mathrm{JS}}(\boldsymbol{X}) = \left(1 - \frac{d-2}{\|\boldsymbol{X}\|^2}\right)\boldsymbol{X},$$

and let

$$\mathrm{MSE}(\boldsymbol{\theta}, \boldsymbol{\delta}) = \mathbb{E}_{\boldsymbol{\theta}}[\|\boldsymbol{\theta} - \boldsymbol{\delta}\|^2],$$

then

$$\mathrm{MSE}(\boldsymbol{\theta}, \boldsymbol{\delta}^{\mathrm{JS}}) \leq \mathrm{MSE}(\boldsymbol{\theta}, \boldsymbol{X}).$$

Here $\|\boldsymbol{x}\|^2 = \sum_i x_i^2$ is the squared l_2 norm. To give an example, suppose that we were estimating: the mean number of Justin Bieber records sold in the Bahamas, the mean profit Trader Joe's makes from its almond butter, and the mean number of deep learning papers appearing on arXiv each month. Why should our estimate for the number of Bieber albums sold have anything to do with how much Trader Joe's charges for its almond butter, or how many deep learning papers are written?

We are going to prove this result.

(a) (5 points) (Uni-variate Stein's lemma) Suppose $X \sim N(\theta, 1)$, and $h : \mathbb{R} \to \mathbb{R}$ is differentiable and $\mathbb{E}[|h'(X)|] < \infty$. Show that

$$\mathbb{E}[(X - \theta)h(X)] = \mathbb{E}[h'(X)].$$

(b) (5 points) (Multi-variate Stein's lemma) Suppose $\boldsymbol{X} = (X_1, \ldots, X_d) \sim N(\boldsymbol{\theta}, I_d)$ with $\boldsymbol{\theta} \in \mathbb{R}^d$, and let $h : \mathbb{R}^d \to \mathbb{R}^d$ be differentiable with $\mathbb{E}[|\frac{\partial h_i}{\partial x_j}(\boldsymbol{X})|] < \infty$. Then

$$\mathbb{E}[(\boldsymbol{X} - \boldsymbol{\theta})^\top h(\boldsymbol{X})] = \sum_{i=1}^d \mathbb{E}\left[\frac{\partial h_i}{\partial x_i}(\boldsymbol{X})\right].$$

[Hint: consider $\mathbb{E}[(X_i - \theta_i)h_i(\boldsymbol{X})] = \mathbb{E}[\mathbb{E}_{X_i}[(X_i - \theta_i)h_i(\boldsymbol{X})]]$.]

(c) (10 points) Let $h(\boldsymbol{X}) = \boldsymbol{X} - \delta(\boldsymbol{X})$, show that

$$\mathrm{MSE}(\boldsymbol{\theta}, \boldsymbol{\delta}) = d + \mathbb{E}_{\boldsymbol{\theta}}[\|h(\boldsymbol{X})\|^2] - 2\sum_{i=1}^d \mathbb{E}\left[\frac{\partial h_i}{\partial x_i}(\boldsymbol{X})\right].$$

(d) (5 points) Find the expression of $\mathrm{MSE}(\boldsymbol{\theta}, \boldsymbol{\delta}^{\mathrm{JS}})$ and show that it is smaller than $\mathrm{MSE}(\boldsymbol{\theta}, \boldsymbol{X})$.