ST793 Midterm 1

ID: ____________________ **Name:** ____________________

Instructions:

- Read each question carefully.
- Show all work to receive full credit.
- Make all work legible.

Honor Pledge:

I certify that I have not received or given unauthorized aid in taking this exam.

Signed: ____________________

1. *(30 points)* Denote by T survival time, and consider two independent samples of survival times: $T_{11}, \ldots, T_{1n_1} \overset{IID}{\sim} Exponential(\lambda_1)$ and $T_{21}, \ldots, T_{2n_2} \overset{IID}{\sim} Exponential(\lambda_2)$, where $T \sim Exponential(\lambda)$ denotes the exponential distribution with rate λ which has distribution function equal to $P(T \leq t) = 1 - \exp(-\lambda t)$. Suppose that we do not observe T_{ij} but instead observe (δ_{ij}, U_{ij}) for $i = 1, 2$ and $j = 1, \ldots, n_i$ where

$$U_{ij} = \min(T_{ij}, C_i)$$
$$\delta_{ij} = 1(T_{ij} \leq C_i)$$

and C_i is the fixed "potential censoring time" for items in the ith group, in the sense that we have a censored observation at time u if $C_i = u$ and $T_{ij} > u$, where $i = 1, 2$. Assume that the pairs (δ_{ij}, U_{ij}) are independent and assume non-informative censoring; that is C_i and T_{ij} are independent.

   (a) *(5 points)* Construct the likelihood function corresponding to a generic sample.

   (b) *(5 points)* Construct the score vector.

(c) *(5 points)* Find the MLE for λ_1 and λ_2.

(d) *(5 points)* Calculate the total Fisher Information matrix.

(e) *(5 points)* Derive the Wald test statistic (T_W) for testing the null hypothesis $H_0: \lambda_1 = \lambda_2$ versus $H_a: \lambda_1 \neq \lambda_2$.

(f) *(5 points)* Specify the asymptotic null distribution of T_W.

2. *(30 points)* Let $Y_{11}, \ldots, Y_{1n_1} \overset{IID}{\sim} Bernoulli(p_1)$ and $Y_{21}, \ldots, Y_{2n_2} \overset{IID}{\sim} Bernoulli(p_2)$ two independent samples, where $p_1, p_2 \in (0, 1)$ are unknown parameters.

   (a) *(5 points)* Construct the log-likelihood function corresponding to a generic sample.

   (b) *(5 points)* Construct the score vector.

(c) *(7 points)* Calculate the total Fisher information matrix.

(d) *(7 points)* Derive the score test statistic, (T_S) for testing the null hypothesis $H_0: \ p_1 = p_2 = 1/2$ versus the alternative $H_a: \ p_1 \neq 1/2$ or $p_2 \neq 1/2$.

(e) *(6 points)* Specify the asymptotic null distribution of T_S.

3. *(25 points)* Let $Y_1, \ldots, Y_n \overset{IID}{\sim}$ *Exponential* distribution with location ϕ and scale λ, which has density function given by

$$f(y;\ \phi,\ \lambda) = \lambda^{-1} \exp\{-(y - \phi)/\lambda\} \qquad y \geq \phi$$

where $\phi \geq 0$ and $\lambda > 0$.

(a) *(10 points)* Find the MLE for ϕ, $\widehat{\phi}_{MLE}$, and the MLE for λ, $\widehat{\lambda}_{MLE}$.

(b) *(8 points)* Give the asymptotic distributions of the $\widehat{\phi}_{MLE}$ and $\widehat{\lambda}_{MLE}$ respectively. Make sure you state all the results used in your derivations.

(c) *(7 points)* Give an appropriate asymptotic mean $\boldsymbol{\mu}_n$ and asymptotic variance Σ_n such that we can write that $(\widehat{\phi}_{MLE},\ \widehat{\lambda}_{MLE})^T \sim AN_2(\boldsymbol{\mu}_n, \Sigma_n)$.

4. *(5 points)* Let $Y_1, \ldots, Y_n$ be a sequence of independent normally distributed variables such that $Y_1, \ldots, Y_\phi$ have mean μ and variance σ_0^2, while $Y_{\phi+1}, \ldots, Y_n$ have mean $\mu+\beta$ and variance 1. Here $\mu,\ \beta,\ \phi$ are unknown parameters with $\mu \in \mathbb{R}$, $\beta \neq 0$, and $\phi \in \{1, \ldots, n\}$. Hence there is a change point ϕ such that the mean of all Y_j with $j > \phi$ is "increased" by β if $\phi < n$. Consider a test of hypothesis that there is no change $H_0 :\ \phi = n$ versus the alternative that there is a change, $H_a :\ \phi < n$.
There are several types of non-regularity exhibited here. Name **two reasons** for which the standard asymptotic results we studied may not apply here.

5. (a) *(5 points)* Suppose that $\boldsymbol{X}_n \rightarrow_d \boldsymbol{X}$ and $\boldsymbol{Y}_n \rightarrow_d \boldsymbol{X}$. The notation $\rightarrow_d$ denotes *convergence in distribution* and $\boldsymbol{X}$ is a non-trivial distribution. Does it mean that $\boldsymbol{X}_n - \boldsymbol{Y}_n \rightarrow_d \mathbf{0}$?
Circle Yes / No. Explain your choice.

(b) *(5 points)* Let $Y_1, \ldots, Y_n \overset{IID}{\sim}$ from some distribution $F(\cdot, \boldsymbol{\theta})$ that satisfies the regularity assumptions. Let $S(\boldsymbol{\theta}, \boldsymbol{Y})$ be the score vector, where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$. The score vector is asymptotically multivariate normal. Specify the asymptotic mean and the asymptotic variance of $S(\boldsymbol{\theta}, \boldsymbol{Y})$.