# WPI Mathematical Sciences Ph.D. General Comprehensive Exam
## MA 541 Probability and Mathematical Statistics II
## May, 2023

Note: Please show a clear logic of each solution. If you cannot solve a problem perfectly, still show your idea on solving the problem.

1. (20 points) For $i = 1, \ldots, n$,

$$X_i \mid Z_i = 0 \overset{ind}{\sim} \text{Normal}(\mu_0, \sigma^2), \quad X_i \mid Z_i = 1 \overset{ind}{\sim} \text{Normal}(\mu_1, \sigma^2),$$

and

$$Z_i \overset{ind}{\sim} \text{Bernoulli}(p),$$

where $p, \mu_0, \mu_1, \sigma^2$ are all unknown. [Note that while $X_1, \ldots, X_n$ are observed, $Z_1, \ldots, Z_n$ may not be observed.]

(a) Write down the likelihood function in two different forms.

(b) Suppose $Z_1, \ldots, Z_n$ are observed, find the maximum likelihood estimators (MLE) of $p, \mu_0, \mu_1, \sigma^2$.

(c) Suppose $Z_1, \ldots, Z_n$ are unobserved, explain how to find MLEs of $p, \mu_0, \mu_1, \sigma^2$. What problem can arise?

(d) Suppose $\mu_0 < \mu_1$, what adjustments can you make in (c)?

2. (20 points) In genetics, the non-mutation rate $p$ (i.e., the probability that no mutations will occur) within a DNA segment is an essential measure for determining the importance of that DNA segment. To estimate this rate, we randomly sample $n$ independent DNA sequences and study the same segment in each. There are different ways to collect data and make estimations. In this case, we consider two methods:

- Estimation 1: Use the sample proportion of non-mutations among the $n$ segments: $\hat{p}$.
- Estimation 2: Count the number of mutations in each segment: $X_1, X_2, \ldots, X_n$. Assuming that these can be modeled as Poisson random variables with mean $\lambda$, we can apply the maximum likelihood estimate (MLE).

(a) For each estimation procedure, please define the estimator and find its asymptotic distribution.

(b) Obtain the asymptotic relative efficiency (ARE) of these two estimators. For this purpose, you can assume the number of mutations can be modeled by a Poisson distribution.

(c) Discuss the advantages and disadvantages of these estimators.

3. (20 points) Let $X_1, \cdots, X_n$ be i.i.d. Uniform$(\theta, \theta + 1)$ random variables where $\theta$ is real.

(a) Find a minimal sufficient statistic for $\theta$.

(b) Show whether the minimal sufficient statistic is complete or not.

4. (20 points) Let $x_1, \ldots, x_n$ be a set of known constants, and let $Y_1, \ldots, Y_n$ be independent with $Y_i \sim \mathcal{N}(\theta x_i, 1)$.

   (a) Find $\hat{\theta}$, the maximum liklihood estimator for $\theta$.

   (b) What is the distribution of $\hat{\theta}$?

   (c) Show that $\sqrt{\sum_i x_i^2}(\hat{\theta} - \theta)$ is a pivot and use it to find a $1 - \alpha$ confidence interval for $\theta$.

   Your answers for the first part may include functions $\phi(\cdot)$ and $\Phi(\cdot)$, the standard normal pdf and cdf, respectively.

5. (20 points) Suppose that $X_1, X_2, \ldots, X_n$ are iid with common density

$$f(x|\theta) = \theta^{-1} e^{x/\theta}, \quad x \geq 0,$$

   where $\theta$ is a positive parameter.

   (a) Find a minimal sufficient statistic for $\theta$.

   (b) Show that $X_{(1)} = min(X_1, X_2, \ldots, X_n)$ is not a sufficient statistic for $\theta$.

   (c) Show that $\hat{\theta} = nX_{(1)}$ is an unbiased estimator of $\theta$.

6. (20 points) Suppose $X_1, X_2, X_3, \cdots, X_n$ be independent and identically distributed random variables from a distribution with.

$$f(x) = \frac{x^2 \exp\left(\frac{-x^2}{2\sigma^2}\right)}{\sigma^3\sqrt{2}\,\Gamma(3/2)} \quad \sigma > 0, \quad x \geq 0.$$

   (a) What is the UMP (uniformly most powerful) level $\alpha$ test for

$$H_0 \;:\; \sigma = 1 \quad \text{vs} \quad H_1 \;:\; \sigma = 2.$$

   (b) If possible find the UMP (uniformly most powerful) level $\alpha$ test for

$$H_0 \;:\; \sigma = 1 \quad \text{vs} \quad H_1 \;:\; \sigma > 1.$$