Math 281AB Qualifying Exam

Professor Jelena Bradic

August 29, 2023

INSTRUCTIONS (read carefully): The total points on this test exceed the maximum points attainable; this is to provide you flexibility in choosing the problems that best align with your strengths and preparation.

Total points possible: 80

Maximum score: 60 points

For those aiming for an **MS** level pass, your target is to accumulate **up to 40 points**. You may choose to address Problems 1 and 2 exclusively, or delve into a portion of Problem 3 to reach this benchmark. If you're setting your sights on a **PhD-** level distinction, you should aim for a score **above 40 points and up to 50 points**. However, if your aspirations are set on achieving a **PhD** level pass or higher, you are expected to gather **up to 60 points**, and this necessitates the attempt of Problem 3, which carries 40 points in itself. While it's possible to achieve the maximum score of 60 points without completing all parts of Problem 3, a thorough and complete solution of both Problems 1 and 2 is then imperative.

NOTE: Ensure that your computation is thoroughly detailed. Do not use examples or exercises from the book as statements. While you can use Theorems, Propositions, and Lemmas as statements, it is imperative to clearly define all relevant terms state the name, number and a book of the result you are using. Online resources should not be referenced or copied. Be meticulous in documenting every step of your work. Ensure your handwriting is clear and legible; illegible work will result in a deduction of points. Kindly submit your work by emailing it to **jbradic@ucsd.edu**.

Problem 1 [20 pt]

Suppose $\boldsymbol{Z}_i = (X_i, Y_i)^\top \in \mathbb{R}^2$, $i = 1, 2, \dots$ is a sequence of an i.i.d. random vectors with a finite mean vector $\boldsymbol{\mu}$ (coordinate-wise) and a variance-covariance matrix $\boldsymbol{\Sigma}$ that has bounded away from zero and infinity eigenvalues. Suppose we are interested in the parameter

$$\omega = \log\Big(\frac{\mu_1}{\mu_2}\Big).$$

Consider an estimator

$$Z_n = \log\Big(\frac{\bar{X}_n}{\bar{Y}_n}\Big), \qquad \bar{X}_n = n^{-1}\sum_{i=1}^n X_i, \bar{Y}_n = n^{-1}\sum_{i=1}^n Y_i.$$

Show that $Z_n \to \omega$ in probability and derive (prove it) the asymptotic distribution of $\sqrt{n}(Z_n - \omega)$ as $n \to \infty$.

Problem 2 [20 pt]

Suppose that Θ follows a log-normal distribution with known hyperparameters $\mu_0 \in \mathbb{R}$ and $\sigma_0^2 > 0$. Suppose that given $\Theta = \theta$, $(X_1, X_2, \cdots, X_n)$ is an i.i.d. sample from Uniform distribution on an interval $(0, \theta)$.

(a) What is the posterior distribution of $\log(\Theta)$?

(b) Let δ_τ be the Bayes estimator of θ under the loss

$$L(\theta, d) = 0, \text{ if } \frac{1}{\tau} \le \frac{\theta}{d} \le \tau$$

and loss is equal to 1 otherwise. Here, $\tau > 1$ is fixed. Find a closed form expression for the limit of δ_τ as $\tau \to 1$.

Note: Part (b) concerns Bayes estimators of θ, not of $\log(\theta)$, but part (a) is still relevant.

Problem 3 [40 pt]

Consider a longitudinal study where subjects are measured at multiple time points. Let Y_{it} be the binary response of subject i (where $i = 1, ..., n$) at time t (where $t = 1, ..., T$).

Assume the logistic regression model for the probability $P(Y_{it} = 1)$:

$$\log\left(\frac{P(Y_{it} = 1)}{1 - P(Y_{it} = 1)}\right) = \mathbf{x}_{it}^T\boldsymbol{\beta}$$

where $\mathbf{x}_{it}$ is a p-dimensional covariate vector for subject i at time t, and $\boldsymbol{\beta}$ is a p-dimensional vector of coefficients. Given the repeated measures nature of the data, it is common to

assume that the observations within a subject might be correlated. Let's denote $\mathbf{Y}_i$ as the response vector of subject i, and let $\mathbf{V}_i$ be the variance-covariance matrix of $\mathbf{Y}_i$. Under the compound symmetry assumption:
1. The variances of all measurements for a given subject are equal and denoted by σ^2.
2. The covariances between any two different measurements for the same subject are equal and denoted by $\sigma^2\rho$, where ρ is the correlation coefficient.

The variance-covariance matrix $\mathbf{V}_i$ for subject i would then look like:

$$\mathbf{V}_i = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{bmatrix}$$

Here, σ^2 is the variance of each measurement, and $\sigma^2\rho$ is the covariance between any two different measurements within the same subject. Assume that $\rho \neq 1$ and $\rho \neq \frac{-1}{T-1}$.

(a) Derive the likelihood function for the longitudinal logistic regression model.

(b) Obtain the Maximum Likelihood Estimator (MLE) for $\boldsymbol{\beta}$ by maximizing the aforementioned likelihood function.

(c) Prove the asymptotic normality of the MLE, $\hat{\boldsymbol{\beta}}$, as $n \to \infty$. Specify the asymptotic mean and variance.

(d). Discuss how the presence of intra-subject correlation (the correlation of measurements within the same subject across different time points) might affect your derivations and conclusions. Would the MLE still be consistent and asymptotically unbiased?