2018 USC Marshall Statistics PhD Screening Exam
Part I: in class (9AM-1:30PM, June 25, 2018)

Instructions:

- This open-book exam is 270 minutes long. Write your answers on the exam sheets. Make sure to hand in all pages.
- There are five questions, each counts 20 points and is expected to be finished in 54 minutes.
- Read the questions carefully. You must show your work to get full credit. Give specific references if you are using any result from class. If you cannot finish a problem within the given exam period, explain your ideas and list the key steps of your proposed solution.

1. This problem has two separate parts (a) and (b).

   (a) For each $x > 0$, let $M(x)$ be a real-valued random variable and set $M(0) = 0$. Assume that the random function $M(x)$ is monotone non-decreasing on $[0, \infty)$. Define $T(y) \equiv \inf\{x \geq 0 : M(x) \geq y\}$. Suppose that $e^{-y}T(y)$ converges in law to an Exponential(λ) random variable when $y \to \infty$.

      (i) Find non-random $a(x)$ and $b(x) > 0$ such that $(M(x) - a(x))/b(x)$ converges in law for $x \to \infty$ to a non-degenerate random variable.

(ii) Provide the distribution function of the limit random variable. What is the name of this distribution?

(b) Let X be $\text{Bin}(n, p)$. Find $\mathbb{E}[1/(i+X)]$, for $i = 1, 2$. Hint: Recall that $\int x^a \delta x = x^{a+1}/(a+1) + C$ for $a \neq -1$.

2. Let $X_1, \cdots, X_n$, be a random sample from a Bernoulli distribution with parameter $p \in (0,1)$. In other words,

$$P(X_i = 1) = p, \text{ and } P(X_i = 0) = 1 - p\,.$$

(a) Find a sufficient statistic $T_n(X_1, \cdots, X_n)$ for p.

(b) Justify $T_n(X_1, \cdots, X_n)$ is sufficient using the definition of sufficiency.

(c) Find the maximum likelihood estimator (MLE) of p, and determine its asymptotic distribution by a direct application of the Central Limit Theorem.

3. This is a statistical inference problem. You may need to do some posterior calculation.

   Suppose that given the vector μ, the random vector X has a normal distribution in $\mathbb{R}^n$ with mean μ and identity covariance matrix. We want to make inference about $\|\mu\|^2$. Please show your work for each part of the question in details. Partial credit will be awarded.

   (a) Find an unbiased estimate of $||\mu||^2$. Call this estimator $\hat{\delta}_{\mathsf{unbiased}}$.

(b) Suppose that a Bayesian has a proper prior distribution for μ that is Gaussian with mean vector 0 and covariance kI, where k is any fixed positive real number and I is the identity matrix. He wants to minimize mean squared error (MSE). The estimator minimizing the MSE is the posterior mean of $||\mu||^2$, i.e, $\mathbb{E}(||\mu||^2|X)$. Find this estimator. Call this estimator $\hat{\delta}_{\mathsf{proper}}$.

(c) Suppose now the Bayesian use the uniform prior (which is also called "flat" or "non-informative" prior) for μ. Report $\mathbb{E}(||\mu||^2|X)$ in this case. Call it $\hat{\delta}_{\mathsf{flat}}$. Report $\hat{\delta}_{\mathsf{flat}} - \hat{\delta}_{\mathsf{unbiased}}$.

(d) Now, if the true distribution of μ is indeed Gaussian with mean vector 0 and covariance kI, then show that with respect to the unconditional (i.e., marginal) distribution of X, the Bayes estimator $\hat{\delta}_{\mathsf{proper}}$ is closer in euclidean distance to $\hat{\delta}_{\mathsf{unbiased}}$ than it is to $\hat{\delta}_{\mathsf{flat}}$ when n is large, i.e., show

$$\mathbb{E}\big(\hat{\delta}_{\mathsf{proper}} - \hat{\delta}_{\mathsf{unbiased}}\big)^2 < \mathbb{E}\big(\hat{\delta}_{\mathsf{flat}} - \hat{\delta}_{\mathsf{unbiased}}\big)^2 \text{ for large } n \ ,$$

where the expectation is over the unconditional distribution of X which is

$$\int_{\mathbb{R}^n} f(x|\mu)\pi(\mu)d\mu$$

with $f(x|\mu) = N_n(\mu, I)$ and $\pi(\mu) = N_n(0, kI)$.
Hint: Let $\hat{D} = \hat{\delta}_{\mathsf{proper}} - \hat{\delta}_{\mathsf{unbiased}}$. Compute the mean and variance of $\hat{D}$ under the unconditional distribution of X, and draw the appropriate conclusion.

You can use this space to continue your answers for Part d).

4. Assume that we have a sample of n independent and identically distributed observations $\mathbf{x}_1, \cdots, \mathbf{x}_n$ from a p-dimensional random vector $\mathbf{x} = (X_1, \cdots, X_p)^T$ with Gaussian distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = (\sigma_{jk})_{1 \leq j,k \leq p}$ is a $p \times p$ nonsingular covariance matrix.

   (a) Denote by $\boldsymbol{\Omega} = (\omega_{jk})_{1 \leq j,k \leq p} = \boldsymbol{\Sigma}^{-1}$ the precision matrix (or inverse covariance matrix). If the covariance matrix $\boldsymbol{\Sigma}$ is sparse with some entries being exactly zero (meaning that some pairs of variables X_j and X_k are uncorrelated), what can we say about the sparsity of the precision matrix $\boldsymbol{\Omega}$? Give a proof if you think $\boldsymbol{\Omega}$ is also sparse, or provide a simple example showing the opposite.

(b) Let us look at the (j,k)th entry ω_{jk} of the precision matrix $\mathbf{\Omega}$. If $\omega_{jk} = 0$ holds what is the statistical intuition of such an equation? Provide some necessary technical details. And can we conclude that the same equation $\sigma_{jk} = 0$ holds for the (j,k)th entry σ_{jk} of the covariance matrix $\mathbf{\Sigma}$?

(c) We now consider the problem of precision matrix estimation for the case of $p > n$. Assume for simplicity that the mean vector of $\mathbf{x}$ is known to be $\mathbf{0}$ to us. How would you estimate $\boldsymbol{\Omega}$? Write down the optimization problem you have in mind and provide some brief discussions on the underlying rationale that motivated your suggested algorithm.

(d) (Optional with bonus credit). Suppose we are only interested in estimating a single entry $\omega_{1,2}$ as opposed to estimating the entire precision matrix $\boldsymbol{\Omega}$ as in part c). What alternative procedure would you suggest for such a purpose (with higher accuracy)? Write down the optimization problem you have in mind and provide some brief discussions on the underlying rationale that motivated your suggested algorithm.

5. Consider the optimization problem

$$\text{minimize}_{\beta \in \mathbb{R}^p} \ \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1, \tag{1}$$

where $y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times p}$, and $\lambda > 0$. This question has four main parts. Each part should be doable even if you were not able to solve the previous part.

(a) Show that the following problem is a dual of (1):

$$\text{maximize}_{u \in \mathbb{R}^n} \ \frac{1}{2}\|y\|_2^2 - \frac{1}{2}\|y - u\|_2^2 \text{ s.t. } \|X^T u\|_\infty \leq \lambda. \tag{2}$$

Also, show that $\hat{u} = y - X\hat{\beta}$, where $\hat{\beta}$ is a solution of (1) and $\hat{u}$ is a solution of (2).

Hint: Consider rewriting (1) *as minimize*$_{\beta\in\mathbb{R}^p, z\in\mathbb{R}^n}$ $\frac{1}{2}\|y - z\|_2^2 + \lambda\|\beta\|_1$ *s.t.* $z = X\beta$.

(b) Which of the following is necessarily unique? (Carefully justify each answer.)

(i) $\hat{\beta}$

(ii) $\hat{u}$

(iii) $\|y - X\hat{\beta}\|_2^2$

(iv) $\|\hat{\beta}\|_1$

(c) Suppose $y = X\beta^* + \epsilon$, and suppose we have chosen the tuning parameter λ to satisfy $\lambda \geq \|X^T \epsilon\|_\infty$.

(i) Show that

$$\frac{1}{2}\|y - X\hat{\beta}\|_2^2 + \lambda\|\hat{\beta}\|_1 \leq \frac{1}{2}\|\epsilon\|_2^2 + \lambda\|\beta^*\|_1.$$

(ii) Use the result in part (5a) of this question to show that

$$\frac{1}{2}\|y - X\hat{\beta}\|_2^2 + \lambda\|\hat{\beta}\|_1 \geq \frac{1}{2}\|y\|_2^2 - \frac{1}{2}\|X\beta^*\|_2^2$$

(iii) [optional with bonus credit] Use the result from the previous part to show that

$$\frac{1}{2}\|y - X\hat{\beta}\|_2^2 + \lambda\|\hat{\beta}\|_1 \geq \frac{1}{2}\|\epsilon\|_2^2 - \lambda\|\beta^*\|_1$$

(iv) [optional with bonus credit] Use parts (i) and (iii) of this question to suggest an estimator for $\sigma^2 = \mathbb{E}\left[\frac{1}{n}\|\epsilon\|_2^2\right]$.

(d) Show that the function $f(\theta) = (\|\theta\|_1)^2$ is convex.