2014 USC Marshall Statistics PhD Screening Exam
Part I: in class (9AM-1:30PM, June 20, 2016)

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.
- 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. Let $\{V_n\}$ be i.i.d. non-negative random variables. Fixing $r > 0$ and $q \in (0, 1]$, consider the sequence $W_0 = 1$ and $W_n = (qr + (1 - q)V_n)W_{n-1}$, $n \geq 1$.

   A motivating example of W_n is recording the relative growth of a portfolio where a constant fraction q of one's wealth is re-invested each year in a risk-less asset that grows by r per year, with the remainder re-invested in a risky asset whose annual growth factors are random V_n.

   (a) Show that $n^{-1} \log W_n \overset{a.s.}{\to} w(q)$, where $w(q) = \mathbb{E} \log(qr + (1 - q)V_1)$.

(b) Show that $q \mapsto w(q)$ is concave on $(0, 1]$.

(c) Using Jensen's inequality show that $w(q) \le w(1)$ in case $\mathbb{E}V_1 \le r$. Further, show that if $\mathbb{E}V_1^{-1} \le r^{-1}$, then the almost sure convergence applies also for $q = 0$ and $w(q) \le w(0)$.

2. Suppose $X \sim N(\mu, 1)$ [the normal distribution with mean μ and variance 1], and μ is drawn from a mixture prior $(1-\omega)\,\delta_0(d\mu) \;+\; \omega\,\gamma(\mu)\,d\mu$, where δ_0 is point mass at 0 and $\gamma(\mu)$ is a density which is unimodal and symmetric around 0. Show that the posterior median $\hat{\mu}(x)$ is a shrinkage rule, i.e, it satisfies

$$0 \leq \hat{\mu}(x) \leq x \qquad \text{for all } x \geq 0 \ .$$

Hint: Show first that if $x > 0$, $v > 0$ and $x \neq v$, then for the posterior distribution we have:

$$\pi(\mu = x - v | X = x) \geq \pi(\mu = x + v | X = x) \ .$$

Continue your answers for Question 2 if needed.

3. Define an estimator $\widehat{\theta}_n$ of location as the center of an interval of length 2 that contains the largest possible fraction of the observations. Assuming consistency of $\widehat{\theta}_n$, derive its rate of convergence. Feel free to assume reasonable regularity conditions. A perfectly rigorous proof is not absolutely necessary: just make a convincing partially rigorous argument for a particular rate of convergence.

Continue your answers for Question 3 if needed.

4. This problem contains four subquestions that are not related to each other.

   (a) Minimize $f_0(x) = x_1^2 + x_2^2$ subject to $h_1(x) = x_1/(x_1^2 + x_2^2) \leq 0$ and $h_2(x) = (x_1 + x_2)^2 = 0$.

(b) Suppose f and g are convex functions, and c_1 and c_2 are positive constants. Show by the definition of convex functions that $c_1 f + c_2 g$ is convex. (Note: please show complete work; do not just quote some intermediate results).

(c) Describe the following algorithms (use mathematical notations when necessary):

i. Newton's method.

ii. Coordinate decent.

iii. Gradient decent.

5. We draw a sample of n independent and identically distributed observations $\mathbf{x}_1, \cdots, \mathbf{x}_n$ from a p-variate Gaussian distribution $N(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is a $p \times p$ nonsingular covariance matrix. Let us consider the problem of estimating the covariance matrix $\mathbf{\Sigma}$ based on the given sample.

   (a) What estimator is a natural unbiased estimator of $\mathbf{\Sigma}$? Write down this estimator by assuming that the mean of the p-dimensional random vector is known to be $\mathbf{0}$ (so there is no need to estimate its mean vector here). We call this estimator $\widehat{\mathbf{\Sigma}}$.

(b) What are the potential issues of the estimator $\widehat{\mathbf{\Sigma}}$ constructed in part a) when the dimension p is large compared to the sample size n? You can consider two cases: 1) $p \leq n$ and p is roughly of the same order as n; 2) $p > n$ or even $p \gg n$ (that is, p is much larger than n, meaning $p/n \to \infty$ as $n \to \infty$). These issues represent the main challenges of large covariance matrix estimation.

(c) Let us consider the special case of $\boldsymbol{\Sigma} = I_p$ and assume that $p/n \to \gamma \in (0, \infty)$ as $n \to \infty$. What can we say about the strong convergence (that is, almost sure convergence) of the largest eigenvalue of $\widehat{\boldsymbol{\Sigma}}$? What can we say about the limiting distribution of the largest eigenvalue of $\widehat{\boldsymbol{\Sigma}}$? Please provide some details when needed.

(d) We now consider the general case where $\mathbf{\Sigma}$ may not be the identity matrix but has all eigenvalues sandwiched between c and c^{-1} for some positive constant $c \in (0, 1]$. What can we say about the limiting behavior of the largest eigenvalue of $\widehat{\mathbf{\Sigma}}$ when $p/n \to \gamma \in (0, \infty)$ as $n \to \infty$? Please provide some details of your reasoning.