Department of Statistics
University of Wisconsin, Madison
PhD Qualifying Exam Option A
August 29, 2017
12:30-4:30pm, Room 133 SMI

- There are a total of FOUR (4) problems in this exam. Please do all FOUR (4) problems.
- Each problem must be done in a separate exam book.
- Please turn in FOUR (4) exam books.
- Please write your code name and **NOT** your real name on each exam book.

1. Let Y be a random variable, X be a random vector, and T be a binary random variable, all defined in a common probability space. Suppose that $Y \perp T \mid X$, i.e., Y and T are independent conditional on X.

   (a) Let (Y_i, X_i, T_i), $i = 1, ..., n$, be a random sample from the distribution of (Y, X, T). Define $\mu = E(Y)$ and

   $$\widehat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \frac{T_i Y_i}{\pi(X_i)},$$

   where $\pi(X) = P(T = 1|X)$. Show that, as $n \to \infty$,

   $$\sqrt{n}(\widehat{\mu} - \mu) \to N(0, \sigma^2) \quad \text{in distribution,}$$

   where

   $$\sigma^2 = E\left\{\frac{E(Y^2 \mid X)}{\pi(X)}\right\} - \mu^2$$

   (b) Suppose that $S = h(X)$, where h is a measurable function, and $Y \perp X \mid S$, i.e., Y and X are independent conditional on S. Show that $Y \perp T \mid S$.

   (c) Assume the conditions in 1 and 2. Let $S_i = h(X_i)$ and $\rho(S) = P(T = 1|S)$. Define

   $$\widetilde{\mu} = \frac{1}{n} \sum_{i=1}^{n} \frac{T_i Y_i}{\rho(S_i)}.$$

   Show that

   $$\sqrt{n}(\widetilde{\mu} - \mu) \to N(0, \tau^2) \quad \text{in distribution}$$

   for some τ^2 with $\tau^2 < \sigma^2$ unless $S = X$.

2. (a) (Part 1) Let P be a distribution with continuous cumulative distribution function F, continuous density f, and finite second moment. Let θ and μ be the median and mean of P, respectively.

   i. If $X \sim P$, prove that

$$E(X-\mu)^2 \leq E(X-\theta)^2, \quad E|X-\mu| \geq E|X-\theta|.$$

   ii. Moreover, when $\theta \neq \mu$ and $f(\theta) > 0$, prove that the strict inequality holds, i.e.,

$$E(X-\mu)^2 < E(X-\theta)^2, \quad E|X-\mu| > E|X-\theta|.$$

(b) (Part 2) Let X_1 and X_2 be independently drawn from distribution P, where P satisfies all the conditions in Part 1. Let

$$Y = |F(X_1) - F(X_2)| \quad \text{and} \quad Z = \min\{F(X_1), F(X_2)\}.$$

   i. Prove that Y and Z have the same distribution.
   ii. If P is not necessarily a continuous distribution, what is your conclusion for Part 2(i)? Prove your result or give a counter-example.

(c) (Part 3) Suppose X is a random variable with finite second moment, but does NOT necessarily have a continuous density f. Let a be a constant such that $0 < P(X > a) < 1$, $Z = \min\{X, a\}$. Let θ_X and θ_Z be the medians of X and Z, respectively.

   i. Prove that

$$E|X-\theta_X| > E|Z-\theta_Z| \quad \text{and} \quad \mathrm{Var}(X) > \mathrm{Var}(Z). \tag{1}$$

   NOTE: Remember to show that the inequality STRICTLY holds.
   ii. When $P(X > a) = 0$ or $P(X > a) = 1$, does (1) still strictly hold? Prove your result or give a counter-example.

3. Consider a linear regression problem where $Y \sim \mathcal{N}(X\beta, I_{n\times n})$, $Y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times p}$ and $\beta \in \mathbb{R}^p$. Without loss of generality, let $\text{rank}(X) = r$ and $X = U\Sigma V^T$ be the singular value decomposition (SVD) for X, where $U \in \mathbb{R}^{n\times r}$, $\Sigma \in \mathbb{R}^{r\times r}$, and $V \in \mathbb{R}^{p\times r}$.

   (a) For an estimator $\widehat{\beta}$, define the mean-squared prediction risk as $\mathcal{R}(\beta, \widehat{\beta}) = \mathbb{E}[\|X(\widehat{\beta} - \beta)\|_2^2]$. Prove that for any estimator, $\mathcal{R}(\beta, \widehat{\beta}) = \text{Bias}^2(X\widehat{\beta}) + \text{Var}(X\widehat{\beta})$ where $\text{Bias}^2(X\widehat{\beta}) := \|\mathbb{E}[X\widehat{\beta}] - X\beta\|_2^2$ and $\text{Var}(X\widehat{\beta}) := \mathbb{E}[\|X\widehat{\beta} - \mathbb{E}[X\widehat{\beta}]\|_2^2]$.

   (b) Determine the system of linear equations that the maximum likelihood estimator $\widehat{\beta}_{MLE}$ satisfies. Under what conditions is $\widehat{\beta}_{MLE}$ unique? Determine $\text{Bias}^2(X\widehat{\beta}_{MLE})$, $\text{Var}(X\widehat{\beta}_{MLE})$, and the overall mean-squared error.

   (c) An iterative algorithm for computing the maximum likelihood estimator is *gradient descent* applied to the negative log-likelihood objective function $-\log \text{lik}(\beta)$. For an objective function $f(\beta)$, the gradient descent update with constant step-size α is:

   $$\widehat{\beta}^{t+1} = \widehat{\beta}^t - \alpha \nabla f(\widehat{\beta}^t) \quad \text{for } t = 0, 1, 2, \ldots.$$

   Write down the gradient descent update for the negative log-likelihood objective function $-\log \text{lik}(\beta)$.

   (d) Recall the SVD for X, $X = U\Sigma V^T$. Consider the gradient descent update from (c), assume $\widehat{\beta}^0 = 0$ and that the step-size α satisfies $0 < \alpha \max_j \Sigma_{jj}^2 < 1$. For all $t = 0, 1, 2, \ldots$, calculate $\text{Bias}^2(X\widehat{\beta}^t)$, $\text{Var}(X\widehat{\beta}^t)$, and the mean-squared error $\mathcal{R}(\beta, \widehat{\beta}^t)$. Your answers should be expressed in terms of α, β, t, and the diagonal elements of Σ. Looking at your expressions, what happens to the squared bias and variance as t increases?

   (e) Find $\text{Bias}^2(X\widehat{\beta}^t)$ and $\text{Var}(X\widehat{\beta}^t)$ for $t = 0$ and $t \to \infty$. Are there any values of t for which $\widehat{\beta}^t$ is dominated by the maximum likelihood estimator (that is $\mathcal{R}(\beta, \widehat{\beta}^t) \geq \mathcal{R}(\beta, \widehat{\beta}_{MLE})$ for all $\beta \in \mathbb{R}^p$ and there exists a β such that $\mathcal{R}(\beta, \widehat{\beta}^t) > \mathcal{R}(\beta, \widehat{\beta}_{MLE})$)?

4. Let $f(x, \mu, \sigma)$ denote the density function of a lognormal variable X, i.e., $\log X$ is normally distributed with mean μ and variance σ^2 and let $g(x, \tau) = \tau^{-1}\exp(-x/\tau)$, $0 < x < \infty$. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with density function $h(x)$. Consider the problem of testing $H_0 : h(x) = f(x, \mu, \sigma)$ versus $H_1 : h(x) = g(x, \tau)$.

   (a) Suppose $\sigma = \sigma_0$ is known but μ and τ are unknown.

      i. Find the log-likelihood ratio statistic S_n for testing $H_0 : h(x) = f(x, \mu, \sigma_0)$ versus $H_1 : h(x) = g(x, \tau)$. You do not need to center S_n to have mean zero nor derive its null distribution and critical value.

      ii. Prove or disprove that S_n is scale invariant.

   (b) Now suppose that μ, σ, and τ are all unknown. Let $\widehat{\mu}$, $\widehat{\sigma}$ and $\widehat{\tau}$ denote their maximum likelihood estimates (MLEs).

      i. Find the log-likelihood ratio statistic $T_n^{(1)}$ for testing H_0 versus H_1. Express it in terms of the MLEs.

      ii. What do you think is the asymptotic null distribution of $T_n^{(1)}$? Explain heuristically; no proofs are needed.

      iii. Let $\tau_0(\mu, \sigma)$ be the limiting value of $\widehat{\tau}$ as $n \to \infty$ under H_0. An alternative test statistic is $T_n^{(2)} = \log(\tau_0(\widehat{\mu}, \widehat{\sigma})/\widehat{\tau})$.

         A. Show that $ET_n^{(2)} \to 0$ as $n \to \infty$ under H_0.

         B. Under H_1, is the limiting value, as $n \to \infty$, of $ET_n^{(2)}$ positive or negative? Justify your answer.

You may use the fact that the Gumbel distribution with density $e^{-x}\exp(-e^{-x})$, $-\infty < x < \infty$, has mean $\gamma \approx 0.5772$ and variance $\pi^2/6 \approx 1.6449$. γ is known as the Euler constant.