Department of Statistics
University of Wisconsin, Madison
PhD Qualifying Exam Option A
August 30, 2016
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 X and Y be independent random variables, $X \sim N(0,1)$, $Y \sim N(0,1)$, and

$$Z = \begin{cases} |Y| & X \geq 0 \\ -|Y| & X < 0 \end{cases}$$

(a) Show that $Z \sim N(0,1)$.

(b) Show that

$$P(\max(X,Z) \leq t) = \begin{cases} [2\Phi(t)-1]^2/2 + 1/2 & t \geq 0 \\ 2[\Phi(t)]^2 & t < 0 \end{cases}$$

where Φ is the cumulative distribution function of $N(0,1)$.

(c) Show by the definition of the conditional expectation that

$$E(Z|X) = \begin{cases} \sqrt{2/\pi} & X \geq 0 \\ -\sqrt{2/\pi} & X < 0 \end{cases} \quad \text{a.s.}$$

(d) Show that X and Z are not jointly normal.

(e) Let $(X_1, Y_1, Z_1), ..., (X_n, Y_n, Z_n)$ be independent and identically distributed as (X, Y, Z).

i. Show that, as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} X_i Z_i \to a \text{ a.s.}$$

and identify the value of a.

ii. Show that, as $n \to \infty$,

$$\left(\frac{1}{n}\sum_{i=1}^{n} X_i Z_i - a_n\right) \Big/ b_n \to N(0,1) \text{ in distribution}$$

and identify the values of a_n and b_n.

2. Consider the family of univariate natural parameter exponential families, with probability density function of the form $f_\theta(x) = h(x)\exp(\theta x - A(\theta))$. Let $\mu(\theta) = \mathbb{E}_\theta[X]$, where X has a natural parameter exponential family distribution. Let $X_1, X_2, ..., X_n$ be independent draws from $f_\theta(x)$.

   (a) Let X_1 and X_2 be two independent random variables with pdf f_θ. Show directly (without using any theorem for exponential families) that the probability density function (pdf) of $Y = X_1 + X_2$ is of the form

   $$f_\theta^{(2)}(y) = h_2(y)\exp(\theta y - 2A(\theta)), \quad -\infty < y < \infty$$

   where $h_2(y)$ is a function of y.

   (b) Extend the previous result to a general integer $n \geq 3$, i.e., if $X_1, ..., X_n$ are independent random variables with pdf f_θ, show directly (without using any theorem for exponential families) that the pdf of $Y = X_1 + X_2 + \cdots + X_n$ is of the form

   $$f_\theta^{(n)}(y) = h_n(y)\exp\{\theta y - nA(\theta)\}, \quad -\infty < y < \infty$$

   where $h_n(y)$ is a function of y.

   (c) Find both the maximum likelihood and method of moments estimators for $\mu(\theta)$ and θ given the data $\{X_i\}_{i=1}^n$ for general A.

   (d) Let $\widehat{\mu}^{ML}$ denote the maximum likelihood estimator for μ. Compute the bias, variance, and mean-squared error for $\widehat{\mu}^{ML}$ for general A. Find a sharp lower bound on the variance of any unbiased estimator of μ .

   (e) Consider the *shrinkage estimator* $\widehat{\mu}^{Shr,\alpha} = (1-\alpha)\widehat{\mu}^{ML}$ for some $0 \leq \alpha \leq 1$. Find the bias, variance, and mean-squared error for $\widehat{\mu}^{Shr,\alpha}$ in terms of μ and α.

   (f) For the Poisson(λ) distribution with probability mass function $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$, find $A(\theta)$. Show that $\widehat{\mu}^{Shr,\alpha}$ is a maximum a posteriori (MAP) estimator. Find and name its corresponding prior distribution.

   Hint: For a prior distribution $q(\mu)$, the MAP estimator is:

   $$\widehat{\mu}^{MAP} := \arg\max_{\mu}\{q(\mu) f_{\mu(\theta)}(x)\}.$$

3. Let (X_i, Y_i), $i = 1, 2, \ldots, n$, be independent and identically distributed bivariate normal random vectors with

$$EX_i = EY_i = 0, \quad EX_i^2 = EY_i^2 = 1.$$

Consider the estimation of $\rho = EX_iY_i$.

(a) Find the minimal sufficient statistics, and show that they are minimal sufficient.

(b) Show that the minimal sufficient statistics are not complete.

(c) Let R_n denote the maximum likelihood estimator (MLE) of ρ. Show that $n^{1/2}(R_n - \rho)$ converges in distribution to a normal distribution with mean 0 and variance τ^2 as $n \to \infty$ and give the value of τ^2 as a function of ρ.

(d) Find a variance stabilizing transformation $g(R_n)$ such that the asymptotic variance of $n^{1/2}(g(R_n) - g(\rho))$ is independent of ρ.

Hint: You may use the result:

$$\begin{aligned}\int \frac{\sqrt{1+x^2}}{1-x^2}\,dx &= \frac{1}{\sqrt{2}}\log\left\{\left(\frac{1+x}{1-x}\right)\left(\frac{\sqrt{2}\sqrt{1+x^2}+1+x}{\sqrt{2}\sqrt{1+x^2}+1-x}\right)\right\} \\ &\quad - \log(\sqrt{1+x^2}+x).\end{aligned}$$

4. Suppose a population consists of N distinct values $a_1, \ldots, a_N$. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with distribution

$$P(X_i = a_j) = \frac{1}{N}, \quad \forall i = 1, \ldots, n; \quad j = 1, \ldots, N.$$

Here, $N, a_1, \ldots, a_N$ are unknown parameters and $n \geq 2$. We wish to estimate the population mean $\mu = EX_j$ by minimizing the squared error loss.

(a) Let the random variable n^* be the number of distinct values which are observed (i.e., the number of different X_i's). Also, let $X_1^*, \ldots, X_{n^*}^*$ be these different values of the observations. For any $a_{j_1}, \ldots, a_{j_n} \in \{a_1, \ldots, a_N\}$, derive the conditional probability

$$P\left(X_1 = a_{j_1}, \ldots, X_n = a_{j_n} \middle| n^*, X_1^*, \ldots, X_{n^*}^*\right).$$

(b) Suppose $\bar{X}^* = \sum_{i=1}^{n^*} X_i^*/n^*$ and $\bar{X} = \sum_{i=1}^{n} X_i/n$. Show that $\bar{X}^*$ is a **strictly** better estimator of μ than $\bar{X}$ with respect to the squared error loss.

(c) If there are same values among $a_1, \ldots, a_N$, does your conclusion for Part (b) change? Prove or give a counter-example for your claim.