STATISTICS Ph. D. COMPREHENSIVE EXAM

January, 2016

General Instructions: Write your Code Word on your answer sheets. Do not put your name or id number on any of your answer sheets. All solutions should be rigorously explained. Show all work. If you cannot solve a problem, at least explain what the problem is about and what approach may lead to its solution.

1. Consider estimating a parameter θ on the real line with weighted squared error loss $L(\theta, a) = w(\theta)(\theta - a)^2$ where $w(\theta) > 0$ for all θ. Given some data $X|\theta$ where θ is random, show that the optimal Bayes decision rule is $\delta(X) = \mathrm{E}[w(\theta)\theta|X]/\mathrm{E}[w(\theta)|X]$.

2. Suppose we have a random sample $(X_1, X_2, \ldots, X_n)$ with i.i.d. $X_i \sim N(\mu, \sigma^2)$. We are interested in the two-sided test of $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma \neq \sigma_0$.

   (a) Find the maximum likelihood estimators of the unknown parameters under H_1 and H_0.

   (b) Find the likelihood ratio statistic, Λ, for testing H_0 versus H_1.

   (c) Show that the size α likelihood ratio test region for testing H_0 versus H_1 is equivalent to a region defined in terms of a statistic based on $T = \sum_{i=1}^{n}(X_i - \bar{X})^2$. (You will not be able to find the α level test, but we have not asked you to.)

   (d) How would you find critical values for a size α test of H_0 versus H_1 based on T. Justify the distribution that you are using.

3. (a) Show that if $\xi \sim N(0, 1)$ then $P(|\xi| > z) \leq exp(-z^2/2)$ for any $z > 2/\sqrt{2\pi}$. (Try substitution and antidifferentiation.)

   (b) Let $\{\xi_n\} \sim N(0, 1)$ iid be independent standard normal random variables. For what values of $c > 0$ (if any) is it true that only finitely-many of the events

   $$A_n = \{|\xi_n| > c\sqrt{log\ n}\}$$

   will occur?

4. (a) Let $Y = \sigma(\rho|U| + \sqrt{1-\rho^2}V)$ where U, V are independent $N(0, 1)$ variables and $\rho \in (-1, 1)$, $\sigma \in (0, \infty)$ are constants. Demonstrate that the pdf of Y is given by

   $$f(y|\rho, \sigma) = \frac{a}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{y^2}{2\sigma^2}\right) \phi\left(\frac{\rho y}{b\sigma\sqrt{1-\rho^2}}\right), -\infty < y < \infty$$

   for some positive constant a, b. Identify the numerical values of a and b. In the above expression, $\phi(z)$ the standard normal CDF.

   (b) What is the distribution of Y when $\rho = 0$?

   (c) Consider n observations $Y_1, Y_2, \ldots, Y_n$ modeled as iid $Y_i \sim f(y_i|\rho, \sigma)$, $\rho \in (-1, 1), \sigma \in (0, \infty)$. Does a maximum likelihood estimate of $(\rho, \sigma) \in (-1, 1) \times (0, \infty)$ always exist?

(d) For the same model suppose we are interested in testing $H_0 : \rho = 0$ (against $H_1 : \rho \neq 0$). What is the value of c (exact or approximate) so that the test that rejects H_0 if

$$\frac{\sqrt{n}\bar{Y}}{S_Y} > c$$

has size 5%? Identify c as a specific quantile of a named distribution (S_Y is the sample standard deviation of $Y_1, Y_2, \ldots, Y_n$).

(e) For $n = 100$, the power of the above 5% test at $(\rho, \sigma) = (0.2, 1)$ is calculated to be 0.36. What can you say about the power of this test at $(\rho, \sigma) = (0.2, 5)$ Justify your answer.

5. Let $\{X_j\} \sim \text{Pois}(\lambda)$ are independent Poisson-distributed random variables, then the Central Limit Theorem asserts that, when properly normalized, the partial sum $S_n \equiv \sum_{j=1}^{n} X_j$ and sample average $\bar{X}_n \equiv S_n/n$ are asymptotically normal with means and variance that can be computed easily.

   (a) For any smooth (say, twice continuously differentiable) function g with non-vanishing derivative $g^{'}(x) \neq 0$, the random variables

   $$Y_n \equiv g(\bar{X}_n)$$

   are also normally distributed for large n (you don't have to prove that). Use a Taylor expansion of $g(\cdot)$ to find the mean and variance of Y_n, in terms of n, λ, g and its derivatives. Try to make your answers **correct to order** $1/n$.

   (b) Find a power $p \neq 0$ for which the function $g(x) \equiv |x|^p$ leads to random variables $Y_n \equiv g(\bar{X}_n)$ whose approximate variance does not depend on λ. Such a "variance stabilizing transformation" is sometimes used to make regression models behave better. Give the approximate mean and variance of $Y_n = |\bar{X}_n|^p$

6. Consider a linear model

$$Y = X\beta + e, \qquad \text{E}(e) = 0$$

with the (not necessarily estimable) linear constraint $\Lambda'\beta = d$. Consider two solutions to the constraint, b_1 and b_2, so that $\Lambda' b_k = d$, $k = 1, 2$. Define appropriate least squares fitted values $\hat{Y}_k$ from the model $Y = X_0\gamma + Xb_k + e$ where $X_0 = XU$ with $C(U) = C(\Lambda)^{\perp}$. Show that $\hat{Y}_1 = \hat{Y}_2$. Hint: After finding $\hat{Y}_k$, show that $(I - M_0)X(b_1 - b_2) = 0$.