MATHEMATICAL STATISTICS
QUALIFYING EXAM
Summer 2014

NAME:____________________________                                    ID:__________________________________

Instructions: There are 4 problems – you are required to solve them all. Please show detailed work for full credits. This is a close book exam. Please do NOT use calculator or cell phone during the exam.

1. (a) For any random variable X and Y, prove that the correlation between them \( \rho_{XY} \) is between 0 and 1, i.e.
\[
0 \leq \rho_{XY} \leq 1
\]
(b) Suppose that \((X,Y)\) is distributed as a bivariate normal with the following
\[
\left( \begin{array}{c} X \\ Y \end{array} \right) \sim \mathcal{N} \left( \left( \begin{array}{c} \mu_X \\ \mu_Y \end{array} \right), \left( \begin{array}{cc} \sigma_X^2 & \rho \sigma_X \sigma_Y \\ \rho \sigma_X \sigma_Y & \sigma_Y^2 \end{array} \right) \right)
\]
Show that the correlation between X and Y is \( \rho \).
(c) Let \((X_1, \ldots, X_n)\) have a multinomial distribution with \(m\) trials and cell probabilities \(p_1, \ldots, p_n\). Show that for every \(i\) and \(j\),
\[
\text{Cov}(X_i, X_j) = -mp_i p_j.
\]

2. Let \(X_1, \ldots, X_n\) be iid random variables such that its moment generating function \(M_{X_i}(t) < \infty\) for \(|t| < h\), for some \(h > 0\). Let \(EX_i = \mu\) and \(Var(X_i) = \sigma^2 > 0\), and define \(\bar{X_n} = \frac{\sum_{i=1}^{n} X_i}{n}\). Show that:
\[
\frac{\sqrt{n}(\bar{X_n} - \mu)}{\sigma} \overset{d}{\rightarrow} \mathcal{N}(0,1)
\]

3. Let \(X_1, \ldots, X_n\) be iid with common pdf
\[
f(x|\theta) = \begin{cases} \theta x^{\theta-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}
\]
where \(\theta > 0\).
(a) Show that the ML estimator of \(\frac{1}{\theta}\) is \(\delta(X_1, \ldots, X_n) = -\frac{1}{n} \sum_{i=1}^{n} \log X_i\).
(b) Calculate the expected value and the variance of \(\delta(X_1, \ldots, X_n)\).
(c) Find the Cramer-Rao lower bound for the variance of an unbiased estimator of \(\frac{1}{\theta}\). What can you conclude?
(d) Find a pivotal quantity that depends on the data only through \(\delta\).
(e) Suppose that \(n = 1\) and that \(\alpha \in (0, 1)\). Use the result in (d) to construct an interval estimator for \(\frac{1}{\theta}\) with confidence coefficient \(1 - \alpha\). Give explicit formulas for the endpoints that involve only \(X_1\) and \(\alpha\).

4. Let \(\{\tilde{y}_i\}_{i=1}^{\mathcal{N}}\) be a random sample and \(\tilde{y}_i = (y_{i1}, y_{i2})\) denote the pair of observations for subject \(i\). Suppose that \(y_{i1}\) and \(y_{i2}\) are independent of each other, and \(y_{it} \in \{0, 1\}\), for \(i = 1, \ldots, \mathcal{N}\) and \(t = 1,2\). Further suppose that \(Y_{it}\) satisfy the following model:
\[
\text{logit}[P(Y_{it} = 1)] = \alpha_i + \beta * I(t = 2),
\]
where \(I(t = 2)\) is an indicator function such that \(I(t = 2) = \begin{cases} 1 & \text{if } t = 2 \\ 0 & \text{otherwise} \end{cases}\).
(a) Find the sufficient statistics \(\tilde{\alpha}_i\) for \(\alpha_i\).
(b) Construct the conditional likelihood function (CLL), i.e., \(\prod_{i=1}^{\mathcal{N}} P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2} | \tilde{\alpha}_i)\).
(c) Find the conditional maximum likelihood estimation of \(\beta\) that maximizes the CLL in (b).