UNIVERSITY OF MASSACHUSETTS
Department of Mathematics and Statistics
Basic Exam - Statistics
Monday, August 31, 2015

Work all problems. 60 points are needed to pass at the Masters Level and 75 to pass at the Ph.D. level.

1. Let $X_1, \ldots, X_n$ be a random sample from a Bernoulli distribution with parameter θ.

   (a) Find the maximum likelihood estimator and the moment estimator of θ.

   (b) Compare the mean squared errors (MSEs) of the above two estimators.

   (c) Find the Fisher information for θ and the Cramer-Rao lower bound for estimating θ.

   (d) Describe how to construct the $100(1-\alpha)\%$ likelihood-based interval for θ.

   For the following parts (e), (f), (g), consider the flat prior distribution for θ, $f(\theta) = 1$.

   (e) Find the posterior distribution of θ. Note that a random variable Y with a Beta distribution with parameters α and β, denoted by $Y \sim Beta(\alpha, \beta)$, has the mean $\frac{\alpha}{\alpha+\beta}$ and the mode $\frac{\alpha-1}{\alpha+\beta-2}$, and the pdf as follows:

   $$f(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}(1-y)^{\beta-1}y^{\alpha-1} \qquad y \in [0,1].$$

   (f) Find the posterior mode of θ.

   (g) Describe how to construct the $100(1-\alpha)\%$ Highest Posterior Density (HPD) interval for θ.

2. Assume that the n pairs of observations, $\left[\binom{X_1}{Y_1}, \binom{X_2}{Y_2}, \cdots, \binom{X_n}{Y_n}\right]$, are independently sampled from a bivariate normal distribution with mean vector $(\mu_X, \mu_Y)'$ and covariance matrix $\begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$ where $-\infty < \mu_X = E(X_i), \mu_Y = E(Y_i) < \infty$, $0 < \sigma_X^2 = Var(X_i), \sigma_Y^2 = Var(Y_i) < \infty$ and $-1 < \rho = Corr(X_i, Y_i) < 1$ is the correlation between X_i and Y_i. The maximum likelihood estimator (MLE) of ρ is the sample coefficient of correlation,

   $$\hat{\rho} = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \Big/ \sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2},$$

   where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, and $\hat{\rho}$ is a consistent estimator of ρ.

(a) The asymptotic distribution of $\hat{\rho}$ is $\sqrt{n}(\hat{\rho}-\rho) \xrightarrow{D} N(0,(1-\rho^2)^2)$. Derive the approximate $100(1-\alpha)\%$ confidence interval of ρ.

Consider the following transformation $g(\rho) = \frac{1}{2}\log\frac{1+\rho}{1-\rho}$ for the next three questions, (b), (c) and (d).

(b) What is the MLE of $g(\rho)$?

(c) Derive the asymptotic distribution of the MLE of $g(\rho)$.

(d) Construct the approximate $100(1-\alpha)\%$ confidence interval of ρ using the transformation $g(\rho)$ and its asymptotic distribution derived in (c).

(e) The bivariate normal pdf is given by

$$f(x,y|\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2,\rho)$$
$$= \{2\pi\sigma_X\sigma_Y\sqrt{(1-\rho^2)}\}^{-1}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}-\frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right\}$$

Write out the joint likelihood $L(\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2,\rho)$ of the observations $\left[\binom{X_1}{Y_1},\cdots,\binom{X_n}{Y_n}\right]$.

(f) It is known that the maximum likelihood estimators (MLEs) of the parameters $\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2$, and ρ are respectively $\hat{\mu}_X=\bar{X},\hat{\mu}_Y=\bar{Y},\hat{\sigma}_X^2=\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^2,\hat{\sigma}_Y^2=\frac{1}{n}\sum_{i=1}^n(Y_i-\bar{Y})^2$, and $\hat{\rho}$. Plug in these MLEs to the joint likelihood and show that the maximized likelihood is

$$\begin{aligned}\max_{\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2,\rho} L(\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2,\rho) &= L(\hat{\mu}_X,\hat{\mu}_Y,\hat{\sigma}_X^2,\hat{\sigma}_Y^2,\hat{\rho}) \\ &= \{2\pi\hat{\sigma}_X\hat{\sigma}_Y\sqrt{(1-\hat{\rho}^2)}\}^{-n}\exp\{-n\}.\end{aligned}$$

(g) For the hypothesis

$$H_0:\rho=0 \text{ vs } H_1:\rho\neq 0,$$

Show that the likelihood ratio test is equivalent to

$$\text{Reject } H_0 \text{ if and only if } (1-\hat{\rho}^2)^{n/2} < k$$

for some constant k.

(h) Let $T=\frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}}$. It is known that the above likelihood ratio test is equivalent to

$$\text{Reject } H_0 \text{ if and only if } |T| > c$$

for some constant c. It is also known that T has the Students's t distribution with $n-2$ degrees of freedom when $\rho=0$. Find the value c to make the likelihood ratio test a level α test.

3. Suppose we conduct an experiment by giving rats one of five possible doses of a drug, denoted by $Y_1 < Y_2 < \cdots < Y_5$. For each dose level Y_i, n rats are used and the number of dead rats, denoted by X_i, is observed. Here we have five independent binomial random variables, that is X_i has a binomial distribution with the probability of death, p_i ($X_i \sim Binom(n,p_i)$) where $i=1,\cdots,5$ and $p_1\leq p_2\leq\cdots\leq p_5$. The main interest in this experiment is to estimate the dose at which the rats have a 50 percent chance of dying: $\theta=\min\{i: p_i\geq 0.5\}$, which is a function of $p_1,\cdots,p_5$. We decide to take a Bayesian approach by using a flat prior distribution truncated over the parameter space, $H=\{(p_1,\ldots,p_5): p_1\leq p_2\leq\cdots\leq p_5\}$.

(a) Write the joint likelihood $L(p_1, \ldots, p_5)$.

(b) Describe how to estimate the posterior mean of θ,

$$E(\theta \mid X_1, \ldots, X_5) = \int \cdots \int_H \theta f(p_1, \ldots, p_5 \mid X_1, \ldots, X_5) dp_1 \cdots dp_5.$$