Ma 612 Mathematical Statistics
Midterm Examination

due via email or hand in to my mailbox on Friday March 30, 2012

(1) The random variables $X_1, X_2, \dots, X_n$ are i.i.d., they take values in a finite set
$$A_m = \{1, 2, \dots, m\}$$
and are uniformly distributed over A_m. The random variables $X_i's$ could be numbers that you see, e.g., the number of questions in an exam.

(a) Assume that m is not known and you want to estimate it. Write down the estimator obtained (i) by the method of moments and (ii) by maximizing the likelihood.

(b) If $\hat{m}$ is the MLE, show that $\hat{m}$ is biased and consistent for m, i.e., $E_m(\hat{m}) \neq m$, but $\hat{m} \to m$ if m is kept fixed and $n \to \infty$.

(c) Suppose m and n both tend to infinity at the same rate, i.e., $m = k \cdot n, n \to \infty$, with k a fixed constant. Show that $(\hat{m} - m)$ converges in distribution and study the limiting distribution.

(2) Let the variables $X_1, X_2, \dots, X_n$ i.i.d. from the distribution:
$$f_1(x|\theta_1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
where $\theta_1 = (\mu, \sigma^2)^T \in \mathbb{R} \times (0, \infty)$, and let $Y_1, Y_2, \dots, Y_n$ i.i.d. from the distribution:
$$f_2(y|\theta_2) = \frac{1}{2\sqrt{2\pi\sigma_1^2}} e^{-\frac{(y-\mu_1)^2}{2\sigma_1^2}} + \frac{1}{2\sqrt{2\pi\sigma_2^2}} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}},$$

where $\theta_2 = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)^T \in \mathbb{R}^2 \times (0, \infty)^2$. f_2 is called the normal mixture density.

(a) Show that both functions above are legitimate pdf's

(b) Calculate the likelihood functions in both cases $L_1(\theta_1|x_1, \ldots, x_n)$, and $L_2(\theta_2|y_1, \ldots, y_n)$

(c) Show that the first likelihood function is always bounded on $\mathbb{R} \times (0, \infty)$, but the second is always unbounded on $\mathbb{R}^2 \times (0, \infty)^2$

(d) What does part (c) imply about the MLE estimators?

(3) You want to estimate the mean score μ of a large population of financial engineering students on an aptitude test. You base your work on the following information:

(a) The population contains 80% males and 20% females.

(b) The performance of men and women on the test differs. You expect from past data that the standard deviations of test scores are

$$\sigma_M = 2 \text{ (men)}, \qquad \sigma_F = 4 \text{ (women)}.$$

(c) Test scores vary normally in each subpopulation.

Your task: choose sizes n_M and n_F for random samples of men and women that minimize the total sample size $n = n_M + n_F$, subject to the requirement that the standard deviation of your estimate of μ is 0.1.

(a) What is the MLE $\hat{\mu}$ of μ?

(b) Use the estimator $\hat{\mu}$. What are the required n_M and n_F?

(4) Let $X_1, X_2, \ldots, X_n$ be Bernoulli random variables with the following joint distribution, depending on $\theta = (\theta_1, \theta_{11}, \theta_{10})$;

$$P(X_1 = 1) = \theta_1$$

$$P(X_i = 1|X_1, \ldots, X_{i-1}) = \begin{cases} \theta_{11}, & \text{if } X_{i-1} = 1 \\ \theta_{10}, & \text{if } X_{i-1} = 0 \end{cases}$$

for $i = 1, 2, \ldots, n$. Find a four dimensional sufficient statistic.

(5) Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables uniform on $[0, \theta]$ with $\theta \geq 1$. We wish to estimate $g(\theta) = \theta^{-p}$ for $p > n$.

   (a) Show that

$$T(X_{(n)}) = \begin{cases} 1 & X_{(n)} \leq 1 \\ X_{(n)} & X_{(n)} > 1 \end{cases}$$

   is a sufficient statistic ($X_{(n)} = \max_i\{x_i\}$).

   (b) Find an unbiased estimator for $g(\theta)$. Calculate the variance of this estimator.

(6) Let $X_1, X_2, \dots, X_n$ be a random sample from the population with pdf:

$$f(x|\theta) = (\theta + 1)x^\theta \, \mathbf{1}_{[0,1]}(x)$$

where the parameter space is $\Theta = (-1, \infty)$.

   (a) Find an estimator for θ using the method of moments, call this estimator $\tilde{\theta}_n$.

   (b) Find the limiting distribution of $\sqrt{n}(\tilde{\theta}_n - \theta)$ as $n \to \infty$.

   (c) Find the maximum likelihood estimator $\hat{\theta}_n$ for θ.

   (d) Find the limiting distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ as $n \to \infty$.

   (e) Compare the variances of the limiting distributions for the two estimators. If smaller variance means a better estimator which is the better estimator?

(7) Let $X_1, X_2, \dots, X_n, \dots$ be independent random variables with distribution $X_n \sim Poisson(\lambda_n)$ where $\lambda_n = 2 + \sqrt{n-1} - \sqrt{n}$ for every n. Find the limiting distribution of $\sqrt{n}(\overline{X}_n - 2)$ where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ is the average of the first n random variables.

(8) Let $X_1, \dots, X_n$ be iid with common density $f(x) = \frac{1}{2}e^{-|x|}$, $x \in \mathbb{R}$. This density is a simple version of the Laplace density also called the double exponential. Show that

$$\lim_{n\to\infty} \sqrt{n}\frac{\sum\limits_{i=1}^n X_i}{\sum\limits_{i=1}^n X_i^2} = Z,$$

with convergence in distribution. Find the distribution of Z.