STATISTICS 630 - Solutions to Test II **November 6, 2013**

1. Suppose that the random variables (X, Y) have joint probability density function (pdf)

$$f(x,y) = \begin{cases} 15x^2y, & 0 < x < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

You may use without proof the results from the first exam that the marginal pdfs of X and Y are

$$f_X(x) = \begin{cases} \frac{15}{2}x^2(1-x^2), & 0 < x < 1, \\ 0, & \text{otherwise,} \end{cases} \quad \text{and} \quad f_Y(y) = \begin{cases} 5y^4, & 0 < y < 1, \\ 0, & \text{otherwise,} \end{cases}$$

and also that the conditional pdf of X given $Y = y$ for $0 < y < 1$ is given by

$$f_{X|Y}(x|y) = \frac{3x^2}{y^3},\ 0 < x < y.$$

(a) Obtain $E(X)$.

$$E(X) = \int_0^1 x\frac{15}{2}x^2(1-x^2)dx = \frac{15}{2}\left(\frac{x^4}{4} - \frac{x^6}{6}\right)\Big|_0^1 = \frac{5}{8}$$

(b) Compute $E[X|Y = y]$. Then compute $E[E[X|Y]]$ to verify directly that $E(X) = E[E(X|Y)]$.

$$E[X|Y = y] = \int_0^y x\frac{3x^2}{y^3}dx = \frac{3x^4}{4y^3}\Big|_0^y = \frac{3y}{4}.$$

Then

$$E[E[X|Y]] = \int_0^1 \frac{3y}{4}5y^4dy = \frac{15y^6}{24}\Big|_0^1 = \frac{5}{8}.$$

2. Suppose that $X_1, \ldots, X_n$ is a random sample from a distribution with probability density function

$$f_X(x) = \begin{cases} 2\theta x e^{-\theta x^2}, & 0 < x < \infty, \\ 0, & \text{otherwise,} \end{cases}$$

where $\theta > 0$.

(a) Find the maximum likelihood estimator of θ.

The log-likelihood is

$$\ell(\theta) = \log\left(\Pi_{i=1}^n 2\theta x_i e^{-\theta x_i^2}\right) = n\log(2) + n\log(\theta) + \sum_{i=1}^n \log(x_i) - \theta\sum_{i=1}^n x_i^2.$$

The likelihood equation is

$$\frac{\partial \ell(\theta)}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^{n} x_i^2 = 0.$$

This has solution $\hat{\theta} = n/\sum_{i=1}^{n} x_i^2$. Since $\frac{\partial^2 \ell(\theta)}{\partial \theta^2} = -n/\theta^2 < 0$ at $\theta = \hat{\theta}$, we have found a maximum. Thus, the maximum likelihood estimator is

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} X_i^2}.$$

(b) Find the method of moments estimator based on the mean of the distribution,

$$E(X) = \sqrt{\pi}/(2\sqrt{\theta}).$$

First we solve the above expression for θ:

$$\sqrt{\theta} = \sqrt{\pi}/(2E(X)) \quad \text{or} \quad \theta = \pi/(2E(X))^2.$$

We replace $E(X)$ by $\bar{X}_n$ to get the method of moments estimator,

$$\hat{\theta} = \frac{\pi}{4\bar{X}_n^2}.$$

3. The statistical software package SAS is used in an applied statistics course where many of the students are not experienced programmers. Suppose that the number of errors X made by a randomly chosen student for a given SAS program has the following probability mass function:

x	0	1	2	3
$p_X(x)$	0.4	0.3	0.2	0.1

(a) Verify that $E(X) = 1$ and Var(X) = 1 (be sure to show your calculations).

$$E(X) = (0)(0.4) + (1)(0.3) + (2)(0.2) + (3)(0.1) = 1$$
$$E(X^2) = (0^2)(0.4) + (1^2)(0.3) + (2^2)(0.2) + (3^2)(0.1) = 2$$
$$\mathrm{Var(X)} = \mathrm{E(X^2)} - (\mathrm{E(X)})^2 = 2 - 1 = 1.$$

(b) Suppose that a class of 64 students is taking this course. Suppose that the numbers of errors committed by the students in the class on the given SAS program can be considered to be a random sample from the above distribution. Obtain an expression for the approximate probability that the total number of errors committed by the 64 students in the class exceeds 60.

$$P[\sum_{i=1}^{n} X_i > 60] = P\left[\frac{\sum_{i=1}^{n} X_i - (64)1}{\sqrt{64 \cdot 1}} > \frac{60 - (64)1}{\sqrt{64 \cdot 1}}\right] \approx P[Z > -0.5] = 1 - \Phi(-0.5).$$

The approximation holds due to the Central Limit Theorem.

4. Suppose that $X_1 \sim N(2, 2^2)$, $X_2 \sim N(0, 3^2)$, and $X_3 \sim N(-1, 1^2)$ are independent random variables.

   (a) Let $U = X_1 - 2X_2 + 5X_3 + 7$. Find the distribution of U.

   $E(U) = E(X_1) - 2E(X_2) + 5E(X_3) + 7 = 2 - (2)(0) + (5)(-1) + 7 = 4$ and $\mathrm{Var(U)} = \mathrm{Var(X_1)} + (-2)^2\mathrm{Var(X_2)} + 5^2\mathrm{Var(X_3)} = 4 + (4)(9) + (25)(1) = 65$. Since U is a sum of independent normal rvs, $U \sim N(4, 65)$.

   (b) Find values of $C_1,\ C_2,\ C_3,\ C_4,\ C_5$, and C_6, (where $C_2 \neq 0$ and $C_4 \neq 0$) so that

   $$\frac{X_1 + C_1}{\sqrt{C_2(X_2 + C_3)^2 + C_4(X_3 + C_5)^2}} \sim t(C_6).$$

   First, we standardize $X_1,\ X_2$, and X_3:

   $$\frac{X_1 - 2}{2},\ \frac{X_2}{3},\ \text{and}\ \frac{X_3 + 1}{1}.$$

   Thus,

   $$\frac{\frac{X_1-2}{2}}{\sqrt{\left(\left(\frac{X_2}{3}\right)^2 + \left(\frac{X_3+1}{1}\right)^2\right)/2}} = \frac{X_1 - 2}{\sqrt{\frac{2}{9}(X_2 + 0)^2 + 2(X_3 + 1)^2}} \sim t(2),$$

   and $C_1 = -2, /C_2 = 2/9,\ C_3 = 0,\ C_4 = 2,\ C_5 = 1$, and $C_6 = 2$.

5. Let T_1 and T_2 be unbiased estimators of μ. Consider the estimator $T = \frac{1}{2}T_1 + \frac{1}{2}T_2$ of μ. Suppose that $\mathrm{Var(T_1)} = 2$ and $\mathrm{Var(T_2)} = 1$.

   (a) Show that T is an unbiased estimator of μ.

   $$E(T) = \frac{1}{2}E(T_1) + \frac{1}{2}E(T_2) = \frac{1}{2}\mu + \frac{1}{2}\mu = \mu.$$

   (b) Obtain mean squared error of T under each of the following scenarios. Use this information to determine which scenario would lead to the best estimation for μ. Explain why.

   i. $\mathrm{Cov(T_1, T_2)} = 0$
   ii. $\mathrm{Cov(T_1, T_2)} = 1$
   iii. $\mathrm{Cov(T_1, T_2)} = -1$

   Since T is unbiased,

   $$MSE(T) = \mathrm{Var(T)} = \frac{1}{4}\mathrm{Var(T_1)} + \frac{1}{4}\mathrm{Var(T_2)} + \frac{1}{2}\mathrm{Cov(T_1, T_2)} = \frac{1}{2} + \frac{1}{4} + \frac{1}{2}\mathrm{Cov(T_1, T_2)}.$$

   Substitute in $\mathrm{Cov(T_1, T_2)}$ for each scenario:

   i. $MSE(T) = 3/4$
   ii. $MSE(T) = 3/4 + 1/2 = 5/4$
   iii. $MSE(T) = 3/4 - 1/2 = 1/4$

   Thus, the third scenario results in the smallest MSE and best estimation of μ.