STATISTICS 630 - Solution to Final Exam **July 24, 2012**

1. Let the independent random variables X and Y both be unbiased measurements of a quantity μ; that is, $E(X) = E(Y) = \mu$. Suppose we combine the two measurements using the weighted average

$$T_\alpha = \alpha X + (1-\alpha)Y,$$

where $0 \le \alpha \le 1$. Suppose that the variances of X and Y are $\sigma_X^2 = 2$ and $\sigma_Y^2 = 1$, respectively. First show that T_α is unbiased estimator of μ. Then find the mean squared error of T_α as an estimator of μ and the value of α that minimizes $MSE(T_\alpha)$.

Since $E(T_\alpha) = E(\alpha X + (1-\alpha)Y) = \alpha E(X) + (1-\alpha)E(Y) = \mu$, T_α is unbiased for μ. Next

$$MSE(T_\alpha) = \mathrm{Var}(Z) = \mathrm{Var}(\alpha X + (1-\alpha)Y) = \alpha^2\mathrm{Var}(X) + (1-\alpha)^2\mathrm{Var}(Y) = 2\alpha^2 + (1-\alpha)^2$$

Call this function of α, $g(\alpha)$. Then $g'(\alpha) = 4\alpha - 2(1-\alpha) = 6\alpha - 2$. Then $g'(\alpha) = 0$ implies that $\alpha = 1/3$. Since $g''(\alpha) = 6 > 0$, this solution provides a minimum.

2. Let $X_1, \ldots, X_n$ be a random sample from the Weibull distribution with density

$$f(x|\theta) = 3\theta x^2 e^{-\theta x^3}, \qquad x > 0, \quad 0 < \theta < \infty.$$

Obtain the maximum likelihood estimator of θ and Fisher's information for θ. Use these to construct an approximate level γ confidence interval for θ.

The joint log likelihood is

$$\log[f_\theta(x))] = n\log(3) + n\log(\theta) + \sum_{i=1}^n \log(x_i) - \theta\sum_{i=1}^n x_i^3.$$

Take the derivative, set equal to zero and solve for θ:

$$\frac{\partial \log[f_\theta(x))]}{\partial\theta} = \frac{n}{\theta} - \sum_{i=1}^n x_i^3 = 0 \Longrightarrow \hat\theta = \frac{n}{\sum_{i=1}^n x_i^3}.$$

The second derivative is $\frac{\partial^2 \log(f_\theta(x)))}{\partial\theta^2} = -\frac{n}{\theta^2} < 0$ when $\theta > 0$. Thus, $\hat\theta$ maximizes the likelihood and is the m.l.e.

To find Fisher's information, we take $nI(\theta) = I_n(\theta) = -E\left[\frac{\partial^2\log(f_\theta(X))}{\partial\theta^2}\right] = \frac{n}{\theta^2}$. The approximate $\gamma-$confidence interval for θ is given by

$$\hat\theta \pm Z_{(1+\gamma)/2}\frac{1}{\sqrt{nI(\hat\theta)}} = \hat\theta \pm Z_{(1+\gamma)/2}\frac{\hat\theta}{\sqrt{n}}$$

3. Let $X_1, \ldots, X_{16}$ be a random sample from a normal distribution with unknown mean μ and known variance $\sigma^2 = 4$. It is of interest to test the hypotheses

$$H_0 : \mu = 10 \quad \text{vs.} \quad H_a : \mu > 10$$

at level of significance α. Define $\bar{X} = \sum_{i=1}^{n} X_i/n$. Find the critical value c_α for a level α test of the form:

"Reject H_0 if $\bar{X} \geq c_\alpha$."

Then obtain an expression in terms of Φ (the standard normal cdf) for the power curve associated with your test. (You will get full credit for a correct expression in terms of Φ and $Z_{1-\alpha}$.)

To find the critical value, set

$$\alpha = P[\bar{X} \geq c_\alpha \text{ when } \mu = 10] = P\left[\frac{\bar{X} - 10}{2/\sqrt{16}} \geq \frac{c_\alpha - 10}{2/\sqrt{16}}\right] = P\left[Z \geq \frac{c_\alpha - 10}{2/\sqrt{16}}\right].$$

Thus, $Z_{1-\alpha} = \frac{c_\alpha - 10}{2/\sqrt{16}}$ which implies that $c_\alpha = 10 + Z_{1-\alpha}\frac{2}{4}$.

The power of the test at $\mu = \mu'$ is given by

$$\begin{aligned} \beta(\mu') &= P[\bar{X} \geq c_\alpha \text{ when } \mu = \mu'] = P\left[\frac{\bar{X} - \mu'}{2/\sqrt{16}} \geq \frac{10 + Z_{1-\alpha}\frac{2}{4} - \mu'}{2/\sqrt{16}}\right] \\ &= P\left[Z \geq Z_{1-\alpha} + \frac{10 - \mu'}{2/4}\right] = 1 - \Phi\left(Z_{1-\alpha} + 20 - 2\mu'\right) \end{aligned}$$

4. Suppose that (X, Y) have the joint probability density function

$$f_{X,Y}(x, y) = \begin{cases} x + y & 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{otherwise,} \end{cases}$$

and marginal probability density functions

$$f_X(x) = \begin{cases} \frac{1}{2} + x & 0 \leq x \leq 1, \\ 0 & \text{otherwise,} \end{cases} \qquad f_Y(y) = \begin{cases} \frac{1}{2} + y & 0 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find $E[X|Y = y]$ and $\text{Var}[X|Y = y]$.

The conditional pdf of X given $Y = y$ is

$$f_{Y|X}(y|x) = \begin{cases} \frac{x+y}{y+1/2} & 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{otherwise,} \end{cases}$$

Then for $0 \leq y \leq 1$,

$$E[X|Y = y] = \int_0^1 x\frac{x + y}{y + 1/2}dx = \frac{1}{y + 1/2}\left(\frac{x^3}{3} + \frac{x^2 y}{2}\right)\Bigg|_{x=0}^{x=1} = \frac{y/2 + 1/3}{y + 1/2} = \frac{3y + 2}{6y + 3}.$$

Similarly,

$$E[X^2|Y=y] = \int_0^1 x^2 \frac{x+y}{y+1/2} dx = \frac{1}{y+1/2}\left(\frac{x^4}{4}+\frac{x^3 y}{3}\right)\Bigg|_{x=0}^{x=1} = \frac{y/3+1/4}{y+1/2}.$$

Thus,

$$\mathrm{Var}[X|Y=y] = \frac{y/3+1/4}{y+1/2} - \left(\frac{y/2+1/3}{y+1/2}\right)^2 = \frac{1+6y+6y^2}{18(1+2y)^2}.$$

5. Let $X_1, \ldots, X_n$ be a random sample from the geometric distribution with probability mass function

$$p_\theta(x) = \theta(1-\theta)^x, x = 0, 1, 2, \ldots, \quad 0 < \theta < 1.$$

Suppose that θ has the prior density

$$\pi(\theta) = 6\theta(1-\theta), \quad 0 < \theta < 1.$$

Obtain the posterior distribution of θ given $X = x$. Obtain the mean of the posterior distribution and compare this to the mean of the prior distribution.

The joint pmf of $X_1, \ldots, X_n$ is $L(\theta) = \theta^n(1-\theta)^{\sum x_i}$.
The posterior pdf is proportional to

$$L(\theta) \times \pi(\theta) = c\theta^n(1-\theta)^{\sum x_i} \times \theta(1-\theta) = c\theta^{n+1}(1-\theta)^{\sum x_i + 1}$$

This is the kernel of the beta$(n+2, \sum_{i=1}^n x_i + 2)$ distribution. Thus, the posterior distribution is the beta$(n+2, \sum_{i=1}^n x_i + 2)$ distribution. The posterior mean equals

$$\frac{n+2}{n+2+\sum_{i=1}^n x_i + 2} = \frac{n+2}{n+\sum_{i=1}^n x_i + 4}.$$

We can rewrite the posterior mean as

$$\left(\frac{\sum_{i=1}^n x_i}{n+\sum_{i=1}^n x_i + 4}\right)\left(\frac{n}{n+\sum_{i=1}^n x_i}\right) + \left(\frac{4}{n+\sum_{i=1}^n x_i + 4}\right)\left(\frac{2}{4}\right).$$

Thus, the posterior mean is a weighted average of the prior mean, 1/2, and the sample proportion of successes, $n/(n+\sum_{i=1}^n x_i)$.

6. Let $X \sim N(2, 4)$ and $Y \sim N(-3, 5)$ be independent normal random variables. (Note: The notation $N(a, b)$ indicates a normal distribution with mean a and variance b.)

   (a) Let $U = 2X + 3Y - 1$ and $V = X - CY$ where C is a constant. Identify the distributions of U and V.

   $E(U) = 2(2) + 3(-3) - 1 = -6$, $\text{Var}(U) = 4(4) + 9(5) = 61$, $E(V) = 2 + 3C$, $\text{Var}(V) = 4 + 5C^2$. Thus, $U \sim N(-6, 61)$ and $V \sim N(2 + 3C, 4 + 5C^2)$.

   (b) For U and V defined in part (a), what is the value of C that makes U and V independent?

   $\text{Cov}(U, V) = \text{Cov}(2X+3Y-1, X-CY) = 2\text{Var}(X)+3\text{Cov}(X, Y)-2C\text{Cov}(X, Y)-3C\text{Var}(Y) = 2(4) + 0 - 0 - 3C(5) = 8 - 15C$. U and V are independent iff $\text{Cov}(X, Y) = 0$, so $8 - 15C = 0$ implies that $C = 8/15$.

   (c) Let $W = C_1(X + C_2)^2 + C_3(Y + C_4)^2$. Find values of C_1, C_2, C_3, C_4, and C_5 (with $C_1 \neq 0$ and $C_3 \neq 0$) so that W has a chi-squared distribution with C_5 degrees of freedom.

   We need to express W as a sum of squared standard normal rvs. Thus, $C_2 = -2$ and $C_4 = 3$ to center the variables at zero. Next we need $\text{Var}[\sqrt{C_1}X] = \text{Var}[\sqrt{C_3}Y] = 1$. Thus, $C_1 = 1/\text{Var}(X) = 1/4$ and $C_3 = 1/\text{Var}(Y) = 1/5$. $C_5 = 2$ since we are adding two squared independent standard normal rvs.