STATISTICS 630 - Solution to Final Exam

July 26, 2013

1. A manufacturer of booklets packages them in boxes of 100. Suppose that the weight X (in ounces) of a single booklet has a distribution with moment generating function

$$M_X(s) = \exp\left(s + \frac{s^2}{800}\right).$$

(a) Use the moment generating function to show that

$$E(X) = 1 \quad \text{and} \quad \mathrm{Var}(X) = \frac{1}{400}.$$

$$E(X) = \frac{dM_X(s)}{ds} = (1 + \frac{2s}{800}) \exp\left(s + \frac{s^2}{800}\right)\Bigg|_{s=0} = 1.$$

$$\begin{aligned} E(X^2) &= \frac{d^2M_X(s)}{ds^2} = \frac{d}{ds}(1 + \frac{2s}{800}) \exp\left(s + \frac{s^2}{800}\right)\Bigg|_{s=0} = 1 \\ &= (1 + \frac{2s}{800})^2 \exp\left(s + \frac{s^2}{800}\right) + \frac{2}{800} \exp\left(s + \frac{s^2}{800}\right)\Bigg|_{s=0} = 1 + \frac{1}{400} \end{aligned}$$

Thus, $\mathrm{Var}(X) = 1 + \frac{1}{400} - 1 = \frac{1}{400}$.

(b) Suppose that $X_1, \ldots, X_{100}$ are independent random variables representing the weights of 100 books with the above mean and variance. Obtain an expression for the approximate probability that the total weight of the 100 booklets exceeds 100.5 ounces.

$$P\left[\sum_{i=1}^{100} X_i > 100.5\right] = P\left[\frac{\sum_{i=1}^{100} X_i - (100)(1)}{\sqrt{100/400}} - \frac{100.5 - (100)(1)}{\sqrt{100/400}}\right] = P[Z > 1] = 1 - \Phi(1).$$

The use of the normal distribution is justified since each X_i has a normal distribution according to the mgf and the sum of independent normal rvs has a normal distribution. You could also use the CLT to justify the above as an approximate probability statement.

2. Let $X_1, \ldots, X_n$ be a random sample from the Poisson(θ) distribution for $\theta > 0$ with probability mass function

$$f(x|\theta) = \begin{cases} \frac{e^{-\theta}\theta^x}{x!}, & x = 0, 1, 2, ..., \\ 0 & \text{otherwise.} \end{cases}$$

Suppose that θ has the prior density

$$\pi(\theta) = \begin{cases} 4\theta^2 e^{-2\theta}, & \theta > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Obtain the posterior distribution of θ given $X_1 = x_1, \ldots, X_n = x_n$. Then obtain the mean and variance of the posterior distribution.

$$\pi(\theta|x) \propto L(\theta|x_1, ..., x_n) \times \pi(\theta) \propto \theta^2 e^{-2\theta} e^{-n\theta} \theta^{\sum_{i=1}^n x_i} = \theta^{2+\sum_{i=1}^n x_i} e^{-(n+2)\theta}$$

We recognize this as the kernel of the gamma($3+\sum_{i=1}^n x_i, n+2$) distribution. Thus, the posterior distribution of θ given $X_1 = x_1, \ldots, X_n = x_n$ is the gamma($3+\sum_{i=1}^n x_i, n+2$) distribution. The posterior mean and variance are

$$E[\theta|x_1, ..., x_n] = \frac{3 + \sum_{i=1}^n x_i}{n+2} \quad \text{and} \quad \text{Var}[\theta|x_1, ..., x_n] = \frac{3 + \sum_{i=1}^n x_i}{(n+2)^2}$$

3. Let $Z_1, \ldots, Z_7$ be independent standard normal random variables. Identify the distribution completely of each of the following random variables. Be sure to explain your reasoning.

   (a) $U = 5Z_1 - 6Z_2 + 3$.
   $E(U) = 3$ and $\text{Var}(U) = 5^2 + (-6)^2 = 61$. Thus, $U \sim N(3, 61)$ since the sum of independent normal rvs has a normal distribution.

   (b) $V = Z_1^2 + Z_2^2 + Z_4^2 + Z_7^2 \sim \chi^2(4)$ by definition of the chi-squared distribution.

   (c) $Y = Z_7/\sqrt{[Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2 + Z_5^2 + Z_6^2]/6}$.
   Since $T = Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2 + Z_5^2 + Z_6^2 \sim \chi^2(6)$, we can write $Y = Z_7/\sqrt{T/6} \sim t(6)$ by definition of the t distribution.

   (d) $W = (Z_1^2 + Z_2^2 + Z_3^2)/(Z_4^2 + Z_5^2 + Z_6^2)$.
   We write
   $$W = \frac{(Z_1^2 + Z_2^2 + Z_3^2)/3}{(Z_4^2 + Z_5^2 + Z_6^2)/3} \sim F(3, 3),$$
   by definition of the F distribution.

4. Suppose that the number of students (Y) enrolling in STAT 630 in a semester is a negative binomial random variable with parameters $r = 20$ and $\theta = 1/4$. Each student that enrolls passes STAT 630 with probability $\pi = 0.8$. Assume that the students' performances are independent and hence, assume that conditional on $Y = y$, the number of students X passing STAT 630 in a semester has a binomial(y, π) distribution. Find the unconditional mean and variance of X, the number of students passing STAT 630 in a semester.

From the properties of the negative binomial distribution,

$$E(Y) = \frac{20(1-1/4)}{1/4} = 60 \quad \text{and}\mathrm{Var}(Y) = \frac{20(1-1/4}{(1/4)^2} = 240.$$

Thus,

$$E[X] = E[E(X|Y)] = E[Y0.8] = 0.8E[Y] = 48$$

and

$$\begin{aligned}\mathrm{Var}(X) &= E[\mathrm{Var}(X|Y)] + \mathrm{Var}[E(X|Y)] = E[Y(0.8)(1-0.8)] + \mathrm{Var}[0.8Y] \\ &= 0.16(60) + (0.8)^2(240) = 163.2.\end{aligned}$$

5. Suppose that $X_1, \ldots, X_n$ are a random sample from a distribution with probability mass function

$$f(x|\theta) = \begin{cases} \frac{e^{-\theta^2}\theta^{2x}}{x!}, & x = 0, 1, 2, \ldots, \ 0 < \theta < \infty \\ 0 & \text{otherwise,} \end{cases}$$

and mean $E(X_i) = \theta^2$. In Test 2, you found that the maximum likelihood estimator is $\hat{\theta} = \sqrt{\sum_{i=1}^n X_i/n}$.

(a) Obtain Fisher's information for θ.

$$\frac{\partial \log(f_\theta(x))}{\partial\theta} = \frac{\partial}{\partial\theta}\left[-\theta^2 + 2x\log(\theta) - \log(x!)\right] = -2\theta + \frac{2x}{\theta}.$$

$$\frac{\partial^2 \log(f_\theta(x))}{\partial\theta^2} = \frac{\partial}{\partial\theta}\left[-2\theta + \frac{2x}{\theta}\right] = -2 - \frac{2x}{\theta^2}.$$

$$I(\theta) = E\left[-\frac{\partial^2 \log(f_\theta(x))}{\partial\theta^2}\right] = 2 + \frac{2E(X)}{\theta^2} = 2 + \frac{2\theta^2}{\theta^2} = 4.$$

(b) Use Fisher's information to construct an approximate level γ confidence interval for θ based on the maximum likelihood estimator.

$$\hat{\theta} \pm Z_{(1+\gamma)/2}\sqrt{\frac{1}{nI(\hat{\theta})}} = \hat{\theta} \pm Z_{(1+\gamma)/2}\sqrt{\frac{1}{n4}}.$$

.

6. Suppose that you have a single observation V from the beta(θ,1) distribution with probability density function

$$f(v|\theta) = \begin{cases} \theta v^{\theta-1}, & 0 < v < 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) You are interested in testing $H_0 : \theta = 1$ versus $H_a : \theta = 2$. Suppose that you reject H_0 whenever $V > 0.9$. Determine the size (or level) of this test. Then determine the power of this test.

The level of the test is

$$\alpha = P[V > 0.9 \text{ when } \theta = 1] = \int_{0.9}^{1} 1 dv = v\Big|_{0.9}^{1} = 0.1.$$

The power of the test is

$$\beta = P[V > 0.9 \text{ when } \theta = 2] = \int_{0.9}^{1} 2v dv = v^2\Big|_{0.9}^{1} = 0.19.$$

(b) Is the test in part (a) most powerful of its size for testing $H_0 : \theta = 1$ versus $H_a : \theta = 2$? Is this test uniformly most powerful for testing $H_0 : \theta = 1$ versus $H_a : \theta > 1$? Be sure to justify your answer.

The likelihood ratio is

$$LR = \frac{f_2(v)}{f_1(v)} = \frac{2v}{1} = 2v.$$

The MP test rejects when $LR > c$. We see that the test in part (a) has this form. By the NP Lemma it is MP of its size.

For any alternative $\theta > 1$, the likelihood ratio is

$$LR = \frac{f_\theta(v)}{f_1(v)} = \frac{\theta v^{\theta-1}}{1} = \theta v^{\theta-1}.$$

Thus, $LR > c$ is equivalent to $\theta v^{\theta-1} > c$. We see that this is equivalent to rejecting when $v > k = (c/\theta)^{1/(\theta-1)}$. Thus, the form of the LR test is the same for any $\theta > 1$ and the test in part (a) is UMP for testing $H_0 : \theta = 1$ versus $H_a : \theta > 1$.