# Homework 10 solutions: Chapter 6

November 17, 2005

**Problem 4**
**(a)**
We are given $X$ which is equal to the number of Yes answers given. So, we can think of $X$ as the sum of two variables. So, $X = Y + Z$ where $Y$ is the number of people who flipped a heads and $Z$ is the number of people who flipped tails and truthfully answered Yes (So $Z = p * tails$.) To find an unbiased estimator for $p$, it might be a good idea to first look at the expected value of $X$. So...

$$E[X] = E[Y] + E[Z] = \frac{n}{2} + \frac{n}{2} * p$$

We know that expected value is linear. Thus we can subtract constants and multiply by constants inside the expected value and then take them out. So, using this fact, we can find an unbiased estimator, $\hat{p}$, by replacing $p$ with $\hat{p}$ in the above equation and setting it equal to $X$ instead of $E[X]$. Thus...

$$\hat{p} = \frac{2X - n}{n} \tag{1}$$

And so
$E[\hat{p}] = E[\frac{2X-n}{n}] = \frac{2}{n}E[X] - \frac{n}{n} = \frac{2}{n}(\frac{n}{2} + \frac{n}{2} * p) - 1 = 1 + p - 1 = p$

**(b)**
To find the standard error of the estimator, we take the square root of its variance. So..

$$Var(\hat{p}) = Var(\frac{2X-n}{n}) = \frac{4}{n^2}Var(X)$$

Thus we need the variance of X. Since we assume that the coin flip is independent of the having the characteristic, we can treat $Y$ and $Z$ as two independent binomial variables($Y\ Bin(n, \frac{1}{2})$ and $Z\ Bin(\frac{n}{2}, p)$) and sum their variances.

$$Var(X) = Var(Y) + Var(Z) = \frac{n}{4} + \frac{n}{2}p(1-p)$$

And so,

$$Var(\hat{p}) = \frac{1 + 2p - 2p^2}{n} \tag{2}$$

**Problem 16**
Given are the following measurements of the ozone level that are assumed to have a normal distribution. $0.06, 0.07, 0.08, 0.11, 0.12, 0.14, 0.21$ To find a 95% confidence interval for $\mu$, we need both $\bar{X}$ and $s$.

$$\bar{X} = \frac{1}{7}\sum X_i \approx 0.113$$
$$s = \sqrt{\frac{1}{6}\sum X_i - \bar{X}} \approx .05$$

The confidence interval is then

$$\bar{X} \pm t\frac{s}{\sqrt{7}} \approx 0.113 \pm 0.05 \tag{3}$$

where $F_{t_6}(t) = .975$ (so $t = 2.45$)

**Problem 20**

Assuming $\sigma^2$ is known for a sample from the normal distribution, we can use a z statistic to construct a confidence interval since

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{(n)}}} \sim N(0,1)$$

Thus for a 95 % confidence interval, we first find $z$ such that $F_z(z) = .975$. So the confidence interval is constructed as

$$\bar{X} \pm 1.96 * \frac{\sigma}{\sqrt{n}} \tag{4}$$

**Problem 21**

We are given the following weights $999.4, 999.8, 1000.4, 1000.8, 1001.0$ We first assume that the measurement errors are normally distributed with known variance of 1. So, using the formula above, we find the 95% confidence interval for the true weight $\mu$.

$$\bar{X} \pm 1.96 * \frac{\sigma}{\sqrt{n}} \approx 1000.3 \pm 1.96 * \frac{1}{\sqrt{5}} \approx 1000.3 \pm 0.88 \tag{5}$$

When we assume an unknown variance, we must follow the same procedure that occurred in Problem 16 where $t = 2.78$ since the degrees of freedom is now 4. Also, we find $s \approx 0.672$. The confidence interval is then

$$\bar{X} \pm t\frac{s}{\sqrt{5}} \approx 1000.3 \pm 0.84 \tag{6}$$

The smaller interval is affected by the fact that the sample variance is much smaller than 1. If $s^2 \approx \sigma^2$, we would expect a larger confidence interval for the interval using t.

**Problem 29**

**(a)** Basically as $r \longrightarrow \infty$, the $X_1, X_2, ..., X_n$ $N(\mu, \sigma^2)$ We know, from the book, the ${\chi^2}_r$ variable divided by r goes to one, leaving $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ which is distributed as a standard normal.

**(b)** $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim Z$ where $Z$ is a standard normal. We also know that $\frac{(n-1)s^2}{\sigma^2} \sim {\chi^2}_{n-1}$. So we use these to show

$$\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} = \frac{\bar{X}-\mu}{\sqrt{\frac{s^2}{n}}} = \frac{\bar{X}-\mu}{\sqrt{\frac{s^2}{n}}} * \frac{\sigma}{\sigma} = \frac{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{s^2}{\sigma^2} * \frac{n-1}{n-1}}} = \frac{Z}{\sqrt{\frac{(n-1)s^2}{\frac{\sigma^2}{(n-1)}}}} = \frac{Z}{\sqrt{\frac{Y}{n-1}}} \sim t_{n-1} \tag{7}$$

Where $Z$ is a random variable with a standard normal distribution and $Y$ has a ${\chi^2}_{n-1}$ distribution.

**Problem 36**

**(a)** From Problem 124 in Chapter 3, we know that $Cov(X_1, X_2) = -np_1p_2$ when $X_1$ and $X_2$ are from a multinomial distribution. We want:

$$Cov(\hat{p_G}, \hat{p_G}) = E[\hat{p_G}\hat{p_G}] - E[\hat{p_G}]E[\hat{p_G}] = \frac{1}{n^2}(E[BG] - E[B]E[G]) = \frac{1}{n^2}Cov(B, G) = \frac{-p_Bp_G}{n} \tag{8}$$

**b** The approximate normal distribution is $N(\hat{p_B} - \hat{p_G}, Var(\hat{p_B} - \hat{p_G}))$ where $Var(\hat{p_B} - \hat{p_G}) = Var(\hat{p_B}) + Var(\hat{p_G}) - 2Cov(\hat{p_B}, \hat{p_G}) \approx \frac{\hat{p_B}(1-\hat{p_b})}{n} + \frac{\hat{p_B}(1-\hat{p_b})}{n} + 2\frac{\hat{p_B}\hat{p_G}}{n}$

**c** Thus, the approximate 95% confidence interval (using the values from 6.3.5) is $0.3 \pm 0.0398$

**Problem 50** $X_1, X_2, .., X_n$ are a sample with pdf $f_\theta(x) = \theta x^{\theta-1}, 0 \leq x \leq 1$. To find the MLE, we need to calculate the likelihood and take its log.

$$L(X) = \Pi\theta {x_i}^{\theta-1} = \theta^n(\Pi x_i)^{\theta-1} \tag{9}$$

$$l(\theta) = n\ln\theta + (\theta - 1)\sum \ln x_i \tag{10}$$

Then, we take the derivative with respect to $\theta$:

$$l'(\theta) = \frac{n}{\theta} + \sum \ln x_i \tag{11}$$

To find the MLE solve for $\theta$

$$0 = \frac{n}{\hat{\theta}} + \sum \ln x_i \Rightarrow -\sum \ln x_i = \frac{n}{\hat{\theta}} \Rightarrow \hat{\theta} = -\frac{n}{\sum \ln x_i} \tag{12}$$