Math 466/566 - Homework 3

1. We defined the chi-squared distribution as follows. Let $Z_1, Z_2, \cdots, Z_n$ be independent standard normal RV's. Let

$$X = \sum_{i=1}^{n} Z_i^2 \tag{1}$$

Then X has the chi-squared distribution with n degrees of freedom. Recall that you computed the distribution of a single Z_i^2 in the first homework. Use this to show that the distribution of one Z_i^2 (note the typo) is a gamma distribution with a suitable choice of α and β and then show that the chi-squared distribution is the gamma distribution with $\alpha = n/2$ and $\beta = 1/2$.

Solution: The previous homework showed the density of a single Z_i^2 is

$$\frac{1}{\sqrt{2\pi x}} e^{-x/2} \tag{2}$$

Comparing with the gamma distribution we see that this is the gamma distribution with $\alpha = 1/2$ and $\beta = 1/2$. Note that you do not need to check that the $\frac{1}{\sqrt{2\pi}}$ agrees with the constant in the gamma distribution, since both functions have total integral of 1.

It follows that the mgf of a single Z_i^2 is

$$\left(\frac{1/2}{1/2 - t}\right)^{1/2} \tag{3}$$

So if we add up n independent copies the mgf is

$$\left(\frac{1/2}{1/2 - t}\right)^{n/2} \tag{4}$$

This is the mgf of a gamma distribution with $\alpha = n/2$ and $\beta = 1/2$.

2. Problem 4 in chapter 3 in the book.

Solution: The mean and variance of a chi-squared distribution with $n-1$ degrees of freedom are

$$E[\chi^2_{n-1}] = n - 1, \quad Var[\chi^2_{n-1}] = 2(n-1) \tag{5}$$

Now s^2 is $\sigma^2/(n-1)$ times χ^2_{n-1}, so

$$E[s^2] = (n-1)\sigma^2/(n-1) = \sigma^2 \tag{6}$$

$$Var[\chi^2_{n-1}] = 2(n-1)\left(\frac{\sigma^2}{n-1}\right)^2 = \frac{2\sigma^4}{n-1} \tag{7}$$

3. (This problem is very close to problem 1 in chapter 3 in the book.) Data set 3.1 has a sample of 25 which is drawn from a population with unknown mean μ and variance σ^2. The population consists of stars and the measurements are indicies of brightness in a certain frequency range. Estimate μ and σ. Give a 95% confidence interval for the estimate of μ.

Solution: we find that the sample mean and variance are

$$\bar{X}_n = -52.58, \quad s^2 = 116.27 \tag{8}$$

The 95% confidence interval is

$$[\bar{X}_n - 1.96\frac{s}{\sqrt{n}}, \bar{X}_n + 1.96\frac{s}{\sqrt{n}}] = [-57.03, -48.13] \tag{9}$$

4. (This problem is very close to problem 2 in chapter 3 in the book.) The people in a population are classified into those that have knowledge of some public health issue and those that do not. p is the proportion of the population that has knowledge of the public health issue. Data set 3.2 has a sample of 900 people. A 1 indicates they have knowledge of the issue, a 0 that they do not. Estimate p with a 90% confidence interval.

Solution: We find the sample proportion is

$$f = 0.4178 \tag{10}$$

The 90% confidence interval is

$$[f - 1.645\sqrt{\frac{p(1-p)}{n}}, f + 1.645\sqrt{\frac{p(1-p)}{n}}] \tag{11}$$

In this formula we estimate p by f and so obtain a confidence interval of $[0.39076, 0.44484]$.

5. Problem 2 in chapter 4 in the book.

Solution:

$$\bar{X}_{1,n_1} = 5.726828, \quad s_1^2 = 30.11883 \tag{12}$$

$$\bar{X}_{2,n_2} = 11.04099, \quad s_2^2 = 58.30184 \tag{13}$$

The book suggests assuming the distributions are exponential. For an exponential distribution the standard deviation equals the mean. If the null hypothesis is true, the two populations have the same mean and the best way to estimate it is to pool the samples. The pooled sample mean is

$$\frac{n_1\bar{X}_{1,n_1} + n_2\bar{X}_{2,n_2}}{n_1 + n_2} = 8.97 \tag{14}$$

So we use this as the standard deviation. So we take $s = 8.97$. The statistic is then

$$t = \frac{\bar{X}_{2,n_2} - \bar{X}_{1,n_1}}{\sqrt{s^2/n_1 + s^2/n_2}} = 1.85 \tag{15}$$

With a level of 5%, the test is that we reject the null hyp if $t > 1.645$. So we reject the null hyp.

If we ignore the books assumption that the distributions are exponential and do not pool the samples to estimate a common σ we have

$$t = \frac{\bar{X}_{2,n_2} - \bar{X}_{1,n_1}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} = 2.588580 \tag{16}$$

So we again conclude we should reject the null hyp. If we do pool the samples to estimate σ, then the pooled estimate of σ is 6.89 and we get $t = 2.41$, and again conclude we should reject the null hyp.

6. Problem 3 in chapter 4 in the book.

Solution: This is a "matched pair" experiment. There is only one population (the pairs) and we are looking at the proportion of the population for which the treated twin is healthier. The null hypothesis is $p = 1/2$. If we assume that the treatment will not make one's health worse, then the alternative hypothesis is $p > 1/2$.

$$f_n = 0.51 \tag{17}$$

The statistic is

$$Z_n = \frac{f_n - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{f_n - 0.5}{\sqrt{0.5(1-0.5)/100}} = 0.2 \tag{18}$$

With a level of 5% and a one sided alternative we should reject null if $Z_n > 1.645$. So we accept the null hyp.

7. Problem 4 in chapter 4 in the book.

Solution:

$$f_1 = 0.2, \quad n_1 = 50 \tag{19}$$

$$f_2 = 0.3272727, \quad n_2 = 55 \tag{20}$$

To estimate the variance we compute the pooled estimate of p:

$$f = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2} = 0.26 \tag{21}$$

Then the statistic is

$$Z = \frac{f_1 - f_2}{\sqrt{f(1-f)/n_1 + f(1-f)/n_2}} = 1.4996 \tag{22}$$

With a level of 5% and a one-sided alternative we reject the null hyp if $Z > 1.645$. So we accept the null hyp.