Josh Dean
Math 728
February 14, 2013
Homework 2

Problem 4.2.1: *Let the observed value of the mean, $\overline{X}$, and the sample variance, S^2, of a random sample of size 20 from a distribution that is $N(\mu, \sigma^2)$ be 81.2 and 26.5 respectively. Find 90%, 95% and 99% confidence intervals for μ. Note how the lengths of the confidence intervals increase as the confidence increases.*

Solution: Because the true variance is unknown, the sample variance is known and the sample is relatively small, the confidence intervals are given by:

$$\left[\overline{X} - t_{\frac{\alpha}{2}}(n-1)\sqrt{\frac{S^2}{n}},\ \overline{X} + t_{\frac{\alpha}{2}}(n-1)\sqrt{\frac{S^2}{n}}\ \right]$$

Substituting in the observed values of n, $\overline{X}$, and S^2 gives:

$$\left[81.2 - t_{\frac{\alpha}{2}}(19)\sqrt{\frac{26.5}{20}},\ 81.2 + t_{\frac{\alpha}{2}}(19)\sqrt{\frac{26.5}{20}}\ \right] \tag{1}$$

Now in order to find the confidence intervals, it is only necessary to substitute in the respective values of $t_{\frac{\alpha}{2}}(19)$: the value of a t-statistic with 19 is the degrees of freedom and probability of $\frac{\alpha}{2}$ in each tail. For the desired intervals of 90%, 95% and 99% confidence, these are given respectively by:

$t_{\frac{0.1}{2}}(19) = t_{0.05}(19) = 1.729 \qquad t_{\frac{0.05}{2}}(19) = t_{0.025}(19) = 2.093 \qquad t_{\frac{0.01}{2}}(19) = t_{0.005}(19) = 2.861$

Finally, combining these values with (1) gives the desired confidence intervals for μ:

90% Confidence: $[79.2098,\ 83.1902]$

95% Confidence: $[78.7908,\ 83.6092]$

99% Confidence: $[77.9067,\ 84.4933]$

Problem 4.2.12: *Let Y be $b(300, p)$. If the observed value of Y is $y = 75$, find an approximate 90% confidence interval for p.*

Solution: First note that, Y/n is an unbiased estimator for p and that the sample is large.

$$\begin{aligned} E\left[\frac{Y}{n}\right] &= \frac{E[Y]}{n} \\ &= \frac{np}{n} \\ &= p \end{aligned}$$

Now note that the variance of Y/n is given by,

$$\begin{aligned} Var\left(\frac{Y}{n}\right) &= \frac{Var(Y)}{n^2} \\ &= \frac{np(1-p)}{n^2} \\ &= \frac{p(1-p)}{n} \end{aligned}$$

Therefore because by the central limit theorem Y/n is normally distributed, a confidence interval for p can be derived as follows,

$$\begin{aligned} P\left(-z_{\frac{\alpha}{2}} < \frac{Y/n - p}{\sqrt{\frac{Y/n(1-Y/n)}{n}}} < z_{\frac{\alpha}{2}}\right) &= 1-\alpha \\ P\left(-\sqrt{\frac{Y/n(1-Y/n)}{n}}z_{\frac{\alpha}{2}} < Y/n - p < \sqrt{\frac{Y/n(1-Y/n)}{n}}z_{\frac{\alpha}{2}}\right) &= 1-\alpha \\ P\left(-Y/n - z_{\frac{\alpha}{2}}\sqrt{\frac{Y/n(1-Y/n)}{n}} < p < z_{\frac{\alpha}{2}}\sqrt{\frac{Y/n(1-Y/n)}{n}} - Y/n\right) &= 1-\alpha \\ P\left(Y/n + z_{\frac{\alpha}{2}}\sqrt{\frac{Y/n(1-Y/n)}{n}} > p > Y/n - z_{\frac{\alpha}{2}}\sqrt{\frac{Y/n(1-Y/n)}{n}}\right) &= 1-\alpha \end{aligned}$$

Using observed values, it follows that the 90% confidence interval for p is given by,

$$\left[75/300 - 1.645\sqrt{\frac{75/300(1-75/300)}{300}},\ 75/300 + 1.645\sqrt{\frac{75/300(1-75/300)}{300}}\right]$$

$$[0.208875, 0.291125]$$

Problem 4.2.17: *A random variable X has a Poisson distribution with parameter μ. A sample of 200 observations from this distribution has a mean equal to 3.4. Construct an approximate 90% confidence interval for μ.*

Solution: First note that $E[X] = \mu$ and that the sample is large. Therefore because the sample mean of a poison distribution, $\overline{X}$, is both an unbiased estimator of μ and distributed $N(\mu, \mu/n)$, a confidence interval for μ can be onstructed as follows:

$$P\left(-z_{\frac{\alpha}{2}} < \frac{\overline{X}-\mu}{\sqrt{\mu/n}} < z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(-z_{\frac{\alpha}{2}}\sqrt{\frac{\mu}{n}} < \overline{X}-\mu < z_{\frac{\alpha}{2}}\sqrt{\frac{\mu}{n}}\right) = 1-\alpha$$

$$P\left(-\overline{X}-z_{\frac{\alpha}{2}}\sqrt{\frac{\mu}{n}} < -\mu < -\overline{X}+z_{\frac{\alpha}{2}}\sqrt{\frac{\mu}{n}}\right) = 1-\alpha$$

$$P\left(\overline{X}+z_{\frac{\alpha}{2}}\sqrt{\frac{\mu}{n}} > \mu > \overline{X}-z_{\frac{\alpha}{2}}\sqrt{\frac{\mu}{n}}\right) = 1-\alpha$$

Using the observed values, it follows that the 90% confidence interval for μ is given by,

$$\left[\overline{X}-z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X}}{n}},\ \overline{X}+z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X}}{n}}\right]$$

$$\left[3.4-1.645\sqrt{\frac{3.4}{200}},\ 3.4+1.645\sqrt{\frac{3.4}{200}}\right]$$

$$[3.18552,\ 3.61448]$$

Problem 4.2.18: *Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ where both parameters μ and σ^2 are unknown. A confidence interval for σ^2 can be found as follows. We know that $(n-1)S^2/\sigma^2$ is a random variable with a $\chi^2(n-1)$ distribution. Thus we can find constants a and b so that $P((n-1)S^2/\sigma^2 < b) = 0.975$ and $P(a < (n-1)S^2/\sigma^2 < b) = 0.95$. Complete sections (a) through (c).*

Solution:

(a) *Show that this second probability statement can be written as*

$$P\left(\frac{(n-1)S^2}{b} < \sigma^2 < \frac{(n-1)S^2}{a}\right) = 0.95$$

This equality follows from simple algebraic manipulation as shown:

$$\begin{aligned} 0.95 &= P\left(a < \frac{(n-1)S^2}{\sigma^2} < b\right) && (2) \\ &= P\left(\frac{a}{(n-1)S^2} < \frac{1}{\sigma^2} < \frac{b}{(n-1)S^2}\right) \\ &= P\left(\frac{(n-1)S^2}{a} > \sigma^2 > \frac{(n-1)S^2}{b}\right) && (3) \end{aligned}$$

(b) *If $n = 9$ and $s^2 = 7.93$, find a 95% confidence interval for σ^2*

With the observed values all that remains in order to use (3) as a confidence interval for σ^2 is to identify a and b. Because $(n-1)S^2/\sigma^2$ is distributed $\chi^2(n-1)$, (2) can be used to identify values of a and b as follows:

For a random variable, $Y \sim \chi^2(8)$:

$$P(2.180 < Y < 17.535) = 0.95$$

It follows from (2) 2.180 and 17.535 can then be used as values of a and b respectively. Now all that remains is to substitute in all observed values in addition to a and b into (3) as follows,

$$\left[\frac{(n-1)S^2}{b}, \ \frac{(n-1)S^2}{a}\right]$$

$$\left[\frac{(8)7.93}{17.535}, \ \frac{(8)7.93}{2.180}\right]$$

$$[3.6179, \ 29.1009]$$

(c) *If μ is known how would you modify the preceding procedure in order to find a confidence interval for σ^2?*

If μ was known, it would be possible to use the fact that the sum of n squared normally distributed random variables has a Chi-Squared distribution with n degrees of freedom to construct a narrower confidence interval as follows:

Normalizing the observations yields: (this is the step that can not be done without knowing μ)

$$\frac{X_i - \mu}{\sigma} \sim N(0,1)$$

It follows then that,

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n)$$

From here it is possible to similarly derive a confidence interval for σ^2 as follows,

$$\begin{aligned} 0.95 &= P\left(a < \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 < b\right) \qquad (4) \\ &= P\left(\frac{a}{\sum_{i=1}^{n}(X_i-\mu)^2} < \frac{1}{n\sigma^2} < \frac{b}{\sum_{i=1}^{n}(X_i-\mu)^2}\right) \\ &= P\left(\frac{\sum i = 1^n (X_i-\mu)^2}{na} > \sigma^2 > \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{nb}\right) \end{aligned}$$

It is now possible to use (4) to find values for a and b similarly to how (2) was used above. From here all that remains to be done is to substitute the observed values.

Problem 4.2.21: *let two independent random samples, each of size 10, from two normal distributions $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ yield, $\bar{x} = 4.8$, $s_1^2 = 8.64$, $\bar{y} = 5.6$, $s_2^2 = 7.88$. Find a 95% confidence interval for $\mu_1 - \mu_2$.*

Solution: First note that $E[\overline{X} - \overline{Y}] = \mu_1 - \mu_2$,

$$\begin{aligned} E[\overline{X} - \overline{Y}] &= E[\overline{X}] - E[\overline{Y}] \\ &= \frac{\sum_{i=1}^{10} E[X_i]}{10} - \frac{\sum_{i=1}^{10} E[Y_i]}{10} \\ &= \frac{10\mu_1}{10} - \frac{10\mu_2}{10} \\ &= \mu_1 - \mu_2 \end{aligned}$$

Because the sample is small, it is possible to use the fact that $\overline{X} - \overline{Y}$ follows a t-distribution with 18 degrees of freedom to construct a confidence interval as follows.

First note that the variance of $\overline{X} - \overline{Y}$ is given by

$$\begin{aligned} Var(\,\overline{X} - \overline{Y}\,) &= Var(\,\overline{X}\,) + Var(\,\overline{Y}\,) \\ &= \frac{\sigma^2}{10} + \frac{\sigma^2}{10} \\ &= \frac{\sigma^2}{5} \end{aligned}$$

Now because it is assumed that both samples share the same variance, it is possible to use a pooled estimator, S_p^2, of σ^2 as follows

$$\begin{aligned} S_p^2 &= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \\ &= \frac{(9)(8.64) + (9)(7.88)}{10+10-2} \\ &= 8.26 \end{aligned}$$

Next utilizing the fact that $\overline{X} - \overline{Y}$ follows a t distribution,

$$\begin{aligned} 0.95 &= P\left(-t_{\frac{\alpha}{2}} < \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sigma\sqrt{1/5}} < t_{\frac{\alpha}{2}}\right) \\ &= P\left(-t_{\frac{\alpha}{2}}\sigma\sqrt{1/5} < \overline{X} - \overline{Y} - (\mu_1 - \mu_2) < t_{\frac{\alpha}{2}}\sigma\sqrt{1/5}\right) \\ &= P\left(-(\overline{X} - \overline{Y}) - t_{\frac{\alpha}{2}}\sigma\sqrt{1/5} < -(\mu_1 - \mu_2) < -(\overline{X} - \overline{Y}) + t_{\frac{\alpha}{2}}\sigma\sqrt{1/5}\right) \\ &= P\left((\overline{X} - \overline{Y}) + t_{\frac{\alpha}{2}}\sigma\sqrt{1/5} > \mu_1 - \mu_2 > (\overline{X} - \overline{Y}) - t_{\frac{\alpha}{2}}\sigma\sqrt{1/5}\right) \end{aligned}$$

Finally using the observed values yields the following confidence interval,

$$\left[(\bar{x} - \bar{y}) - t_{\frac{\alpha}{2}}S_p\sqrt{1/5},\ (\bar{x} - \bar{y}) + t_{\frac{\alpha}{2}}S_p\sqrt{1/5}\right]$$

$$\left[(4.8 - 5.6) - 2.101\sqrt{8.26}\sqrt{1/5},\ (4.8 - 5.6) + 2.101\sqrt{8.26}\sqrt{1/5}\right]$$

$$[-3.500, 1.900]$$

Problem 4.2.22: *Let two independent random variables, Y_1 and Y_2, with binomial distributions that have parameters $n_1 = n_2 = 100$, p_1 and p_2, respectively, be observed to be equal to $y_1 = 50$ and $y_2 = 40$ Determine an approximate 90% confidence interval for $p_1 - p_2$.*

Solution: First note that, $(Y_1 - Y_2)/n$ is an unbiased estimator for $p_1 - p_2$,

$$\begin{aligned} E\left[\frac{Y_1 - Y_2}{n}\right] &= \frac{E[Y_1 - Y_2]}{n} \\ &= \frac{np_1 - np_2}{n} \\ &= p_1 - p_2 \end{aligned}$$

Because the sample is large and $(Y_1 - Y_2)/n$ is a form of sample mean, it is possible to use the central limit theorem to construct a confidence interval as follows,

$$\begin{aligned} 0.90 &= P\left(-z_{\frac{\alpha}{2}} < \frac{\frac{(Y_1-Y_2)}{n} - (p_1 - p_2)}{\sqrt{\frac{(Y_1/n)(1-Y_1/n)+(Y_2/n)(1-Y_2/n)}{n}}} < z_{\frac{\alpha}{2}}\right) \\ &= P\left(-z_{\frac{\alpha}{2}}\sqrt{\frac{(Y_1/n)(1-Y_1/n)+(Y_2/n)(1-Y_2/n)}{n}} < \frac{(Y_1 - Y_2)}{n} - (p_1 - p_2)\right. \\ &\qquad \left. < z_{\frac{\alpha}{2}}\sqrt{\frac{(Y_1/n)(1-Y_1/n)+(Y_2/n)(1-Y_2/n)}{n}}\right) \\ &= P\left(-\frac{(Y_1 - Y_2)}{n} - z_{\frac{\alpha}{2}}\sqrt{\frac{(Y_1/n)(1-Y_1/n)+(Y_2/n)(1-Y_2/n)}{n}} < -(p_1 - p_2)\right. \\ &\qquad \left. < -\frac{(Y_1 - Y_2)}{n} + z_{\frac{\alpha}{2}}\sqrt{\frac{(Y_1/n)(1-Y_1/n)+(Y_2/n)(1-Y_2/n)}{n}}\right) \\ &= P\left(\frac{(Y_1 - Y_2)}{n} + z_{\frac{\alpha}{2}}\sqrt{\frac{(Y_1/n)(1-Y_1/n)+(Y_2/n)(1-Y_2/n)}{n}} > (p_1 - p_2)\right. \\ &\qquad \left. > \frac{(Y_1 - Y_2)}{n} - z_{\frac{\alpha}{2}}\sqrt{\frac{(Y_1/n)(1-Y_1/n)+(Y_2/n)(1-Y_2/n)}{n}}\right) \end{aligned}$$

All that remains is to substitute observed values as follows,

$$\left[\frac{(Y_1 - Y_2)}{n} - z_{\frac{\alpha}{2}}\sqrt{\frac{(Y_1)(1-Y_1)+(Y_2)(1-Y_2)}{n}},\right.$$

$$\left.\frac{(Y_1 - Y_2)}{n} + z_{\frac{\alpha}{2}}\sqrt{\frac{(Y_1/n)(1-Y_1/n)+(Y_2/n)(1-Y_2/n)}{n}}\right]$$

$$\left[\frac{(50-40)}{100} - 1.645\sqrt{\frac{(50/100)(1-50/100)+(40/100)(1-40/100)}{100}},\right.$$

$$\left.\frac{(50-40)}{100} + 1.645\sqrt{\frac{(50/100)(1-50/100)+(40/100)(1-40/100)}{100}}\right]$$

$$[-0.01515,\ 0.21515]$$

Problem 4.2.27: *Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be i.i.d samples from the respective distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ where all four parameters are unknown. Construct a confidence interval for σ_1^2/σ_2^2 using the following statistic by answering (a) through (c)*

$$F = \frac{\frac{(m-1)S_2^2}{\sigma_2^2}/(m-1)}{\frac{(n-1)S_1^2}{\sigma_1^2}/(n-1)} = \left(\frac{\sigma_1^2}{\sigma_2^2}\right)\left(\frac{S_2^2}{S_1^2}\right)$$

Solution:

(a) *What kind of distribution does F follow?*

Because F is a ratio of two chi-squared random variables devided by their respective degrees of freedom, F follows the F distribution with parameters $r_1 = m-1$ and $r_2 = n-1$

(b) *From the correct table, a and b can be found such that $P(F < b) = 0.975$ and $P(a < F < b) = 0.95$*

Because $F \sim F(m-1, n-1)$ all that remains is to find the two points from the F-table. The second, b, is given by $F_{0.025}(m-1, n-1)$ which denotes the point on the distribution $F(m-1, n-1)$ such that, $P(F > F_{0.025}(m-1, n-1)) = 0.025$ The first point, a, is given by $F_{0.975}(m-1, n-1)$ which denotes the point on the distribution such that $P(F > F_{0.975}(m-1, n-1)) = 0.975$. A helpful identity in locating these two points is:

$$\frac{1}{F_{0.025}(n-1,\ m-1)} = F_{0.975}(m-1,\ n-1)$$

(c) *Rewrite* $P(a < F < b) = 0.95$ *as follows:*

$$P\left(a\frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < b\frac{S_1^2}{S_2^2}\right)$$

This follows from simple algebraic manipulation,

$$\begin{aligned}
0.95 &= P(a < F < b) \\
&= P\left(a < \frac{\frac{(m-1)S_2^2}{\sigma_2^2}/(m-1)}{\frac{(n-1)S_1^2}{\sigma_1^2}/(n-1)} < b\right) \\
&= P\left(a < \left(\frac{\sigma_1^2}{\sigma_2^2}\right)\left(\frac{S_2^2}{S_1^2}\right) < b\right) \\
&= P\left(a\frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < b\frac{S_1^2}{S_2^2}\right)
\end{aligned}$$

Which is the desired confidence interval for σ_1^2/σ_2^2