STA 261s2005 Assignment 7

Do this assignment in preparation for the quiz on Wednesday, March 9th. The questions are practice for the quiz, and are not to be handed in.

1. Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with expected value μ and variance σ^2.
   (a) Find $Cov(X_4, (X_4 - \overline{X}))$. Why is this a key to establishing the independence of $\overline{X}$ and S^2?
   (b) Given that $\overline{X}$ and S^2 are independent, prove that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.
   (c) Using Question 1b, derive an exact $(1-\alpha)100\%$ confidence interval for σ^2.
   (d) Using Question 1b and the independence of $\overline{X}$ and S^2, derive an exact $(1-\alpha)100\%$ confidence interval for μ.

2. Let $X_1, \ldots, X_{n_1}$ be a random sample from a $N(\mu_1, \sigma^2)$ distribution, and let $Y_1, \ldots, Y_{n_2}$ be a random sample from a $N(\mu_2, \sigma^2)$ distribution. These are *independent* random samples, meaning that the X and Y values are independent. Using the facts that $\overline{X}$ and S^2 are independent and $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ for each sample,
   (a) Derive an exact $(1-\alpha)100\%$ confidence interval for $\mu_1 - \mu_2$.
   (b) Derive an exact $(1-\alpha)100\%$ confidence interval for $\frac{\sigma_1^2}{\sigma_2^2}$.

3. Let $X_1, \ldots, X_{n_1}$ be a random sample from a distribution (not necessarily normal) with expected value μ_1 and variance σ_1^2, and let $Y_1, \ldots, Y_{n_2}$ be a random sample from a distribution (not necessarily normal) with expected value μ_2 and variance σ_2^2. The random samples are independent of each other. The Central Limit Theorem tells us that for large n_1, the distribution of $\overline{X}_{n_1}$ is approximately $N(\mu_1, \frac{\sigma_1^2}{n_1})$. Similarly, the distribution of $\overline{Y}_{n_2}$ is approximately $N(\mu_2, \frac{\sigma_2^2}{n_2})$. So, what should the approximate distribution of $\overline{X}_{n_1} - \overline{Y}_{n_2}$ be?

4. Using your answer to Question 3, give an approximate large-sample confidence interval for $\mu_1 - \mu_2$. Compare Theorem 11.4. Now, to make it useable in practice, substitute consistent estimators for σ_1^2 and σ_2^2. If you don't know the distributions (the typical case, unless they are Bernoulli), you can always use S_1^2 and S_2^2.

5. Read Sections 11.1 to 11.3 and do Exercises 11.6 and 11.9.

6. Read Sections 11.4 and 11.5.
   (a) The following is a substitute for Exercise 11.1, which has a hint that may not be too helpful. Let $X_1, \ldots, X_{n_1}$ be a random sample from a Bernoulli distribution with parameter θ.
      i. Starting with the *unmodified* Central Limit Theorem, show that

$$P\left\{\left(\frac{\sqrt{n}(\overline{X}_n - \theta)}{\sqrt{\theta(1-\theta)}}\right)^2 < z_{\alpha/2}^2\right\} \approx 1 - \alpha.$$

ii. Using that, show

$$P\left\{(n+z_{\alpha/2}^2)\theta^2 - (2n\overline{X}_n + z_{\alpha/2}^2)\theta + n\overline{X}_n^2 < 0\right\} \approx 1-\alpha.$$

iii. Clearly, the function $g(\theta) = (n+z_{\alpha/2}^2)\theta^2 - (2n\overline{X}_n + z_{\alpha/2}^2)\theta + n\overline{X}_n^2$ is the equation of a parabola. It's a *random parabola*; what a concept. Every time you select a new sample, you get a different parabola. Show that they all open upward.

iv. This means that the points where the parabola intersects the θ axis are the endpoints of the confidence interval we want, because between those two points, the parabola is below zero, and that's the event that has probability $1-\alpha$. Find the two points.

v. Simplify the expression $B^2 - 4AC$ under the radical in the quadratic formula, obtaining the confidence interval

$$\frac{2n\overline{X}_n + z_{\alpha/2}^2}{2(n+z_{\alpha/2}^2)} \pm \frac{\sqrt{4nz_{\alpha/2}^2\overline{X}_n(1-\overline{X}_n) + z_{\alpha/2}^4}}{2(n+z_{\alpha/2}^2)}.$$

Compare this to the expression in Exercise 11.1.

Well, it looks a bit messy (though it's easy to program) and it was a lot of work, but the goal was worthwhile. We were able to isolate θ in the middle of a random interval without having to resort to replacing the variance (a function of θ) by an estimate.

(b) Here is a rational way to decide on your sample size before the data are collected. Do Exercise 11.12, but start with your calculations from Question 6(a)i rather than Theorem 11.7 (there's not much difference, but it's a little cleaner). Hint: You need to show $\theta(1-\theta) \leq \frac{1}{4}$.

(c) Do Exercise 11.13. Hint: Consider three cases, depending on whether the interval from θ' to θ'' is to the left of $\frac{1}{2}$, to the right of $\frac{1}{2}$, or includes $\frac{1}{2}$. The answer in the textbook is not quite right, because they missed the case where the interval includes $\frac{1}{2}$.

(d) Do Exercise 11.15; Your answer to Question 3 should help. Also, find a second (higher) probable upper bound that does not depend on $\widehat{\theta}_1$ or $\widehat{\theta}_2$.

(e) Before proceeding to Exercise 11.15, consider this. For fixed total sample size $n = n_1 + n_2$, what choice of n_1 will *minimize* the conservative upper bound from the last question? Minimizing the (probable) margin of error is a good thing; it makes estimation more precise.

(f) Now that you have justified the choice of equal sample sizes, do Exercise 11.16.

7. Read Sections 11.6 and 11.7. Do exercises 11.25, 11.29, 11.31 (assume normality), 11.34 (Answ: CI for $\mu_1 - \mu_2$ is -16.3 to +1.5), 11.35, 11.38 (Answ: 0.68 ± 0.053), 11.45, 11.46 (Answ: $n = 2017$), 11.53, 11.58 (Answ: 0.165 to 2.752).