Project 5 - Sampling Distributions

Statistics 401: Fall 2006

Due: Monday, October 16

Your write-up must be typed. Please number your answers as the questions were numbered. Your grade will be determined by how well you answer the questions, your justification for your answers, and by the professionalism and clarity of our write-up.

1. Let X be the number of SPAM emails received per day per employee at a large software engineering company. Suppose that the distribution of X is:

X	0	1	2
$P(X = x)$	0.60	0.30	0.10

   Compute μ_x and σ_x. See the Chapter 8 Handout for definitions of μ and σ^2.

2. Consider problem 8.10(a) (so don't do (b)) on page 340 of your textbook.

   (a) Compute μ and σ for the data. Use the fact that $P(x) = \frac{1}{N} = \frac{1}{4}$ in your computations. Page 118 of your Course Notes show how the formulas for μ and σ simplify.

   (b) To compute the sampling distribution of $\overline{X}$ for all samples of size 2, note that there are 6 ways to choose two books from 4 books.

   (c) **From the sampling distributions of** $\overline{X}$, compute $\mu_{\overline{x}}$, $\sigma^2_{\overline{x}}$, and $\sigma_{\overline{x}}$.

   Include the following two tables in your report which summarize your results from (b) and (c):

Sample	$\bar{x}$
1. ___ ___	_____
2. ___ ___	_____
3. ___ ___	_____
4. ___ ___	_____
5. ___ ___	_____
6. ___ ___	_____

Sampling Distribution of $\overline{X}$

Value of $\bar{x}$	$P(\bar{x})$
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
$\mu_{\bar{X}}$:	_____
$\sigma^2_{\bar{X}}$:	_____
$\sigma_{\bar{X}}$	_____

3. Do problem 8.16 on page 350.

4. Do problem 8.18 on page 350. In addition to (a) and (b) in the textbook, also answer the following:

   (c) What is the approximate distribution of $\overline{X}$ when the sample size is $n = 50$? Explain why your answer is correct.

   (a) What is the probability that the average wait time for 50 individuals is longer than 25 seconds?

5. Do problem 8.32 on page 357.

6. Do problem 8.34 on page 357.

7. Visit the website

   `http://www.maths.soton.ac.uk/teaching/units/ma1c6/links/samplingapplet/samplingapplet.html`

   and read the introductory remarks. At the bottom of the page is an applet which you will use to answer the following questions.

   (a) Repeatedly take samples of size 1 from each of the normal, exponential and uniform distributions. The applet displays histograms of the data for each case. You do NOT need to copy these graphs into your report. Create a table, and in the table give statistics which describe the center and spread, and also describe the shape of the histogram for each of these three cases.

   (b) Repeatedly take samples of size 5 from each of the normal, exponential and uniform distributions. The applet displays histograms of the sampling distribution of $\overline{X}$ when $n = 5$ for each of these three cases. You do NOT need to copy these graphs into your report.

i. Create a table, and in the table give statistics which describe the center and spread of the sampling distribution and also describe the shape of the histogram for each of these three cases when $n = 5$. How does this compare with your description of the these statistics in #7a? Do the values you get agree with the theoretical values?

ii. Which of the three histograms looks most normal? What about the data suggests that this must be the case?

iii. Which of these three histograms looks least normal? What about the data suggests that this ought to be the case?

(c) Repeatedly take samples of size 30 from each of the normal, exponential and uniform distributions. The applet displays histograms of the sampling distribution of $\overline{X}$ when $n = 30$ for each of these three cases. You do NOT need to copy these graphs into your report. The histograms of the sampling distribution for $\overline{X}$ for each of the three cases should look relatively normal. Explain why this ought to always be true.

8. Consider rolling a six-sided die. For each roll, we'll keep track of X, the side which faces up.

(a) Give the distribution of this variable X. You can either give a table of values and probabilities or give a graph.

(b) Calculate μ_x and $sigma_x$.

9. Visit the website

`http://www.stat.sc.edu/~west/javahtml/CLT.html`

and read the introductory remarks. At the bottom of the page is an applet which you will use to answer the remaining questions.

(a) In the applet, roll 1 die a few thousand times. Do the center and shape of the resulting histogram agree or disagree with your distribution from #8a?

(b) We stated the Central Limit Theorem in class as: for a large SRS, $\overline{X} \dot{\sim} N(\mu, \frac{\sigma}{\sqrt{n}})$. Stated another way, CLT asserts that for a large SRS

$$\sum X_i \dot{\sim} N(n\mu, \sqrt{n}\sigma).$$

This fact can be used to answer the following questions:

i. In the applet, roll 2 dice a few thousand times. Where is the histogram centered? Explain why the center is where you say it is. Is the spread of the distribution increasing or decreasing? Why?

ii. Roll 5 dice a few thousand times. Where is the histogram centered? Explain why the center is where you say it is. Is the spread of the distribution increasing or decreasing? Why?

iii. Compare the histograms from 2 die rolls and 5 die rolls. Is one more normal than the other? Why or why not?