MATHEMATICS E-156, SPRING 2014
MATHEMATICAL FOUNDATIONS OF STATISTICAL SOFTWARE
Module #5 (Sampling Distributions)

Last modified: February 26, 2014

Reading from Chihara and Hesterberg

- Chapter 4 (sampling distributions)
- Appendix A.5 and A.6 (normal distribution)
- Appendix A.7 (moment generating function)

Optional Reading from Haigh

- Pages 111-112 discuss the normal distribution and the central limit theorem.

Proof of the Week

- For random variable X, define the moment generating function

  $$M(t) = E[e^{tX}].$$

  Prove that

  - The nth derivative of $M(t)$, evaluated at $t = 0$, is equal to the nth moment $E[X^n]$.
  - If X_1 and X_2 are independent random variables with moment generating functions $M_1(t)$ and $M_2(t)$, then

    $$M_{X_1+X_2}(t) = M_1(t)M_2(t)$$

  (This is done on pp. 370-372)

R scripts

- 5A-Sampling distributions.R
  Topic 1 – The sampling distribution for a Bernoulli distribution
  Topic 2 – sampling from an arbitrary small population

- 5B-SamplingDistKnown.R
  Topic 1 – the sampling distribution of an exponential distribution is a gamma distribution
  Topic 2 – the sampling distribution of sums from a Poisson distribution is another Poisson distribution
  Topic 3 – if you take the mean of samples from a normal distribution, the result is also normal

- 5C-MaxMinMedian.R
  Topic 1 – new sampling statistics: maximum, minimum, median
  Topic 2 – max, min, median for an exponential distribtion, lambda = 1
  Topic 3 – max and min for a normal distribution

- 5D-CLTpValues.R
  Topic 1 – using the central limit theorem to estimate probabilities
  Topic 2 – using the CLT for discrete sampling distributions with a large number of samples
  Topic 3 – continuity correction

- 5P-Proof 5.R
  Topic 1 – moment generating function for discrete distributions
  Topic 2 – summing discrete random variables in R
  Topic 3 – making the mgf for a continuous random variable

Mathematical notes

1. Normal distribution and central limit theorem

   (a) Let
   $$I = \int_{-\infty}^{\infty} e^{-x^2/2} dx = \int_{-\infty}^{\infty} e^{-y^2/2} dy.$$
   Being rather casual about infinite limits of integration, prove that $I^2 = 2\pi$.

   (b) X has the standard normal distribution $N(0, 1)$ if its density function is
   $$\rho(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \text{ and from part (a), } \int_{-\infty}^{\infty} \rho(x) dx = 1.$$
   Prove that $\mathrm{Var}[X] = 1$.

(c) Y has the normal distribution $N(\mu, \sigma)$ if its density function is

$$\rho(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{(y-\mu)^2/(2\sigma^2)}.$$

Prove that $E[Y] = \mu$ and $\mathrm{Var}[Y] = \sigma$.

(d) Let $\Phi(z)$ denote the distribution function for the standard normal distribution. Suppose that $X_1, X_2, \cdots X_n$ be independent, identically distributed random variables with finite expectation μ and finite variance σ^2. The central limit theorem states that

$$\lim_{n\to\infty} P(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}) = \Phi(z).$$

What is the distribution of $\overline{X}$?

What is the distribution of $X_1 + X_2 + \cdots + X_n$?

2. Density function for the maximum

   Suppose that $X_1, X_2, \cdots X_n$ be independent, identically distributed continuous random variables with density function f and distribution function F.

   Let $X_{max} = \max \{X_1, X_2, \cdots X_n\}$.

   Prove that the density function for X_{max} is $f_{max}(x) = nF(x)^{n-1}f(x)$.

   Specialize to the case where the X_i are random variables from Unif$[0, \beta]$

3. Proof of the week

   For random variable X, define the moment generating function

   $$M(t) = E[e^{tX}].$$

   Prove that

   - The nth derivative of $M(t)$, evaluated at $t = 0$, is equal to the nth moment $E[X^n]$.
   - If X_1 and X_2 are independent random variables with moment generating functions $M_1(t)$ and $M_2(t)$, then

   $$M_{X_1+X_2}(t) = M_1(t)M_2(t)$$

   (This is done on pp. 370-372)

Section problems

1. Page 92, exercise 8. Different members of the class can simulate the result by inventing a variety of distributions with mean 6 and variance 10.

2. Page 93, exercise 13. This is a twin of exercise 14, one of the homework problems.

3. Page 94, exercise 15. The sampling distribution is chi-square (see theorem B.15). For $n = 2, 4$, and 5, overlay a graph of the chi-square distribution with the appropriate number of degrees of freedom on a histogram of your sampling distribution.

Homework assignment This assignment should be submitted as a single R script. Include enough comments so that it is clear what you are doing and where each problem begins. You can upload it to the dropbox on the Class 5 page of the Web site.

It is OK to paste and edit lines from the scripts on the course Web site. It is not OK to paste lines from your classmates' solutions!

1. Page 93, exercise 10.
2. Page 93, exercise 14. This is a twin of exercise 13, one of the section problems.
3. Page 94, exercise 16.
4. Page 95, exercise 24(b). You will have to work part (a) to do the required comparison. Include a brief explanation of how you came up with the theoretical expectation of X_{min}.
5. Plot the moment-generating function for two dice, in two different ways.
   (a) Use `outer()` to get the probability mass function for two dice, then write an R function for the moment generating function of this probability mass function and plot a graph of it. (See script 5P for the "vectorize" trick.)
   (b) Write an R function for the moment generating function of the probability mass function for a single die and plot a graph of its square.