Homework 3

Aleck Zhao

September 23, 2016

1. A population consists of N individuals. Each individual has a certain number of friends. Suppose the count of individuals in the population with a given number of friends is as in the following table:

k	number in population with k friends
0	1
1	1
M	$N-4$
$2M-1$	1
$2M$	1

where $M \geq 3$ is an unknown positive integer and $N \geq 7$. A sample if size 3 is drawn without replacement and the numbers of friends for the i sampled individual is denoted by X_i, for $i = 1, 2, 3$.

(a) Let S denote the 3-tuple of elements X_1, X_2, X_3 that are drawn but written in increasing order. Write down a list of the 15 possible values S can take on.

(b) Make a table giving the PMF of S.

(c) Compute the PMF of $Y = X_1 + X_2 + X_3$. Note that Y is a function of S the set that is drawn.

(d) Compute the PMF of $\bar{X} = (X_1 + X_2 + X_3)/3$ and use this to determine $E[\bar{X}]$.

(e) Compute the population variance σ^2. Compute $\text{Var}(\bar{X})$.

(f) If we use $\bar{X}$ to estimate M, what is the mean square error $E[(\bar{X} - M)^2]$?

(g) Let T denote the *sample median*. Compute the PMF of T and determine $E[T]$.

(h) Find an expression for $\text{Var}(T)$ and simplify it.

(i) If we use T to estimate M, what is the mean squared error $E[(T - M)^2]$?

(j) Suppose we will decide which estimator of M to use (sample mean or sample median) based on which has a smaller MSE. Define the *efficiency* of the sample median *relative* to the sample mean to be
$$\text{eff} = \frac{E[(\bar{X} - M)^2}{E[(T - M)^2]}.$$
Show that this expression can be written as the product of two terms, one which is linear in N and the other which is a ratio of two quadratics in M.

(k) Describe situations (for some integers M and N with $M \geq 3$ and $N \geq 7$) when the sample mean has a smaller MSE than the sample median. If $N > 12$, show that the sample median has a smaller MSE than the sample mean no matter what M is (as long as it is at least 3).

2. Complete the following:

(a) Show that if X_i are iid Bernoulli random variables with success probability p for some $p \in (0, 1)$, then
$$\frac{1}{n} \sum_{i=1}^{n} X_i \to p$$
as $n \to \infty$.

(b) In R, get an approximation to the expected value of the length of the longest run in n flips of a fair coin for $n = 10, 20, 30, \cdots, 250$.

(c) Plot the expected value in (a) vs n and try to fit a curve of the form $y = c \log n$ for some c to the data.

(d) Use your fit in (c) to predict the expected value when $n = 500$. Then approximate the value you get using simulation and compare.

(e) Now, consider a Monte-Carlo approximation of the variance of a random variable. Explain why the expression

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2$$

can be used to approximate $E[X^2]$, and thus why

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 \approx E[X^2] - \mu^2 = \mathrm{Var}(X).$$

(f) From the previous part, explain why, for large n, we can approximate $\mathrm{Var}(X)$ using the sample variance of the values $X_1, \cdots, X_n$

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

(g) Take X to be the length of the longest run in n trials. Estimate $\mathrm{Var}(X)$ for $n = 10, 20, 30, \cdots, 250$, and plot $\mathrm{Var}(X)$ vs n.

3. Consider sampling *with replacement* using a sample of size n from a population of size N where each individual i has two attributes x_i, y_i. Let

$$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)$$

denote the population covariance between x and y where μ_x and μ_y denote the population means. Let (X_i, Y_i), $i = 1, \cdots, n$ denote the (x, y) values for the individuals sampled.

Show that the sample covariance

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

is unbiased for σ_{xy}.

Chapter 7: Survey Sampling

45.

46.

48.

Chapter 4: Expected Values

102.