STAT 240 Homework 1

Rebecca Barter, Andrew Do and Kellie Ottoboni

February 13, 2015

Question 1. Consider a box that contains 5 “1” tickets and 7 “0” tickets. Consider drawing 6 tickets from this box at random with replacement. Let $X_1, X_2, ..., X_6$ denote the 6 numbers you observe. Let $\bar{X}$ denote the average of the draws.

a) What is $E[\bar{X}]$?

Recall that in class we showed that

$$E(\bar{X}) = \bar{t}$$

where $\bar{t}$ is the population mean. In particular, this implies that

$$E(\bar{X}) = \frac{5}{12}$$

b) What is $SE[\bar{X}]$? (R hint: Be careful whether the function “sd” divides by the square root of n or $n-1$)

Note that since this example corresponds to a simple box model with replacement, we have that

$$SE(\bar{X}) = \sqrt{Var(\bar{X})} = \sqrt{\frac{1}{n}Var(t)}$$

Using R and noting that the `sd()` function in R divides by $N-1$ rather than N, we found that (to 3dp)

$$SE(\bar{X}) = 0.201$$

c) Use R to simulate 100,000 values of $\bar{X}$. Produce a histogram of these values. (R hint: Use the function sample).

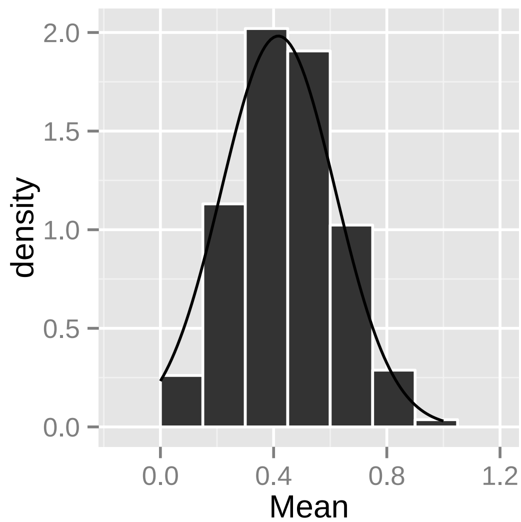

Figure 1: Histogram of 100,000 simulated values of the sample mean when the sample was taken with replacement

d) Let $z_1 = E[\bar{X}] + SE[\bar{X}]$, $z_2 = E[\bar{X}] + 2 \times SE[\bar{X}]$, etc. For $z_1, ..., z_4$ calculate $P(\bar{X} > z_i)$ in three ways:

- **Exactly, using the binomial distribution. (Hint: It will be easier to work with the sample sum than the sample average. R hint: Use function pbinom)**
- **Estimated using the values from part (c)**
- **Using the normal approximation. Use the continuity correction. (R hint: pnorm)**

Do the same for $z_{-4}, ..., z_{-1}$ but calcualte $P(\bar{X} < z_i)$ instead of $P(\bar{X} > z_i)$. Make a table of your results and comment briefly

z	Exact	EmpiricalEst	NormalApprox
-4.00	0.00	0.00	0.00
-3.00	0.00	0.00	0.00
-2.00	0.04	0.04	0.06
-1.00	0.21	0.21	0.28
1.00	0.20	0.20	0.28
2.00	0.05	0.05	0.06
3.00	0.00	0.00	0.00
4.00	0.00	0.00	0.00

Table 1: The exact value, empirical estimation and normal approximation of the probability.

We notice that the Empirical estimation using the results of our simulated value is extremely close the the exact value of the probabilities. On the other hand, the normal approximation is not nearly as accurate. This is likely because our sample size of 6 is very small and the asymptotic assumptions which underly the normal approximation are not yet accurate.

e) Repeat (a)-(d), this time sampling without replacement instead of with replacement. Use the hypergeometric distirbution instead fo the binomial distribuion (R hint: phyper)

Note that since we are now sampling without replacement, we have that

$$E(\bar{X}) = \bar{t} = \frac{5}{12}$$

and

$$SE(\bar{X}) = \sqrt{Var(\bar{X})} = \sqrt{\frac{1}{n}Var(t)\left[\frac{N-n}{N-1}\right]} = 0.149$$

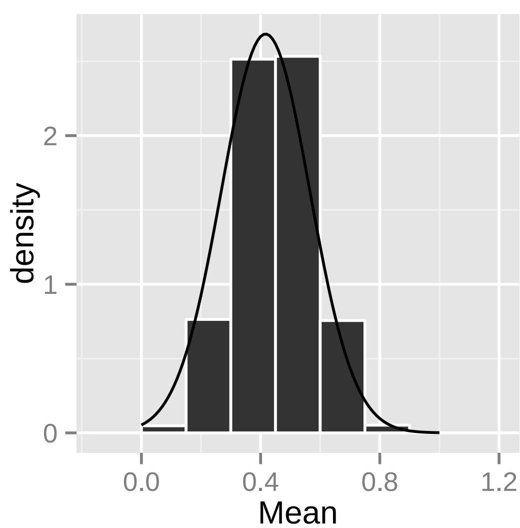

Figure 2: Histogram of 100,000 simulated values of the sample mean when the sample was taken without replacement

z	Exact	EmpiricalEst	NormalApprox
-4.00	0.00	0.00	0.00
-3.00	0.00	0.00	0.01
-2.00	0.01	0.01	0.07
-1.00	0.12	0.12	0.33
1.00	0.12	0.12	0.33
2.00	0.01	0.01	0.07
3.00	0.00	0.00	0.01
4.00	0.00	0.00	0.00

Table 2: The exact value, empirical estimation and normal approximation of the probability.

Small Extra Credit: In class we considered the model

$$Y = a + X$$

where X is standard normal, and ask the question of whether a^2 is estimable. Is it? Justify your answer.

We claim that $Y^2 - 1$ is an unbiased estimator of a^2. The proof is as follows:

$$\begin{aligned}
E(Y^2 - 1) &= E(Y^2 - 1) \\
&= E(a^2 + 2aX + X^2) - E(1) \\
&= E(a^2) + 2aE(X) + E(X^2) - 1 \\
&= a^2 + 0 + 1 - 1 && \text{since } E(X) = 0 \text{ and } X^2 \text{ is distributed as } \chi^2(1)\text{, so } E(X^2) = 1 \\
&= a^2
\end{aligned}$$

Hence a^2 is estimable.