Stat 400, section 6.1c Proportion, Including Point Estimate

notes by Tim Pilachowski

Example F: You flip a fair coin forty times. Define a random variable X = number of heads.

The coin flips are Bernoulli trials which make up a **binomial** experiment. (See Lecture 3.4.)

The number of trials, n, in an experiment is fixed in advance.

There are exactly two events/outcomes for each trial, usually labeled success (S) and failure (F).

Trials are independent from one trial to the next, i.e. the outcome of one trial doesn't affect the next.

$P(S) = p$ must be the same for each trial. $[P(F) = 1 - p = q]$.

The expected value (mean) of a binomial probability distribution is a simple formula:

$$\text{If } X \sim \text{Bin}(n, p), \text{ then } E(X) = np.$$

It is reasonable to expect that a previously-observed proportion p will still hold for any sample of size n.

a) What is the expected value for this experiment?

We also have straightforward formulas for variance and standard deviation of a binomial probability distribution:

$$V(X) = npq = np(1-p) \qquad \sigma_X = \sqrt{npq} = \sqrt{np(1-p)}.$$

b) What are the variance and standard deviation for this experiment?

For a binomial distribution define X = number of successes, $x = 0, 1, 2, \ldots, n$

$$P(X = x) = b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} = \dfrac{n!}{x!(n-x)!} p^x q^{n-x} & x = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

c) You flip a fair coin forty times (sample size $n = 40$). What is the probability of tossing at most 40% heads?

Appendix Table A.1 in your text has cumulative binomial probability distribution tables which provide approximate probabilities for various values of n, p and x. Unfortunately, there isn't a table there for $n = 40$, so we'll have to 1) find another source which has a table for $n = 40$, 2) do the calculations by hand, or 3) find an online utility which will do the calculations for us.

d) You flip a fair coin forty times (sample size $n = 40$). Calculate a normal approximation to the binomial distribution (see Lecture 4.3) to find the probability of tossing at most 40% heads.

Note: Without the continuity correction, because $n = 40$ is relatively small, we would have gotten a different result: $P(X \le 16) = P(Z \le -1.26) = 0.1038$.

If you flip a coin ten times, you would not be surprised if it came up heads only four times or six times rather than exactly 1/2, or five, of the times. If, on the other hand, you flipped a coin ten times and got ten heads, you probably would be surprised. We know intuitively that it is unlikely we will get such an extremely unrepresentative sample. It is not impossible, just very rare. We may even decide the coin must be weighted in some way so that heads are more likely to appear.

Sampling variability is also affected by the number of observations we include. If we flip a coin ten times and get only 3 heads, or 30%, we may not be very surprised. If we flip the same coin 1000 times and only get 300 heads, however, this is more surprising. Again, we may begin to wonder whether the coin is fair. Over a larger number of trials, the percent of heads should be closer to the 50% heads we expect.

Example F revisited – a reinterpretation
You flip a fair coin forty times. Define a random variable $W = 0$ for tails and $W = 1$ for heads.

a) What is the expected value for a single toss?

b) What are the variance and standard deviation of W for a single toss?

c) You flip a fair coin forty times (sample size $n = 40$). What are the expected value and standard error for the sampling distribution (i.e. distribution of $\overline{W}$)?

d) You flip a fair coin forty times (sample size $n = 40$). What is the probability that the average value for W (i.e. the sample mean $\overline{W}$) is less than 0.4?

This Example is a special case, not quite the same as the other Examples in the Central Limit Theorem supplement (section 5.4).

Tossing a coin, for which there are two possible outcomes, is a binomial experiment. The first time around we defined our random variable as X = number of heads (successes) in n tosses of the coin. For the reinterpretation we will define our random variable as a *proportion*, i.e.

$$Y = \frac{\text{number of heads}}{\text{number of tosses}} = \frac{X}{n} = \frac{W_1 + W_2 + \ldots + W_n}{n} = \overline{W} = p \text{ (probability of success).}$$

From the perspective of proportions, the answers to the reinterpreted Example F above would be interpreted as follows.

a) The expected value for a single toss = $E(W) = p = 0.5$.

c) You flip a fair coin forty times (sample size $n = 40$). What are the expected value and standard deviation for Y?

important note for the standard error:

If we generalized the mathematics used above, we would find that the standard deviation for a proportion will always equal $\sqrt{\frac{pq}{n}}$, in this case $\sqrt{\frac{(0.5)(0.5)}{40}} = \sqrt{\frac{1}{160}} \approx 0.0791$.

d) You flip a fair coin forty times (sample size $n = 40$). What is the probability that there will be at most 40% heads?

important note for the z-score calculation:

In terms of proportions, the z-score formula and calculation would be $z = \frac{y-p}{\sqrt{pq/n}} = \frac{0.4-0.5}{\sqrt{0.5*0.5/40}} \approx -1.26$.

"A point estimate of a parameter θ is a single number that can be regarded as a sensible value for θ… The selected statistic is called the point estimator of θ." The symbol $\hat{\theta}$ is used for both the random variable and the calculated value of the point estimate.

Ideally, the point estimator $\hat{\theta}$ is unbiased, i.e. $E(\hat{\theta}) = \theta$. In words, the sampling distribution based on the statistic has an expected value equal to the actual (but unknown) population parameter.

Task: Show that the point estimator $\hat{p}$ (sample proportion) is an unbiased estimator of the population parameter p. That is, show $E(\hat{p}) = p$.

Note that, as with means and variances, a larger sample size will increase the probabilities that our sample statistic $\hat{p}$ will be representative of the population because the larger value in the denominator of the standard deviation of Y (standard error of $\overline{W}$) will mean a smaller variance and standard deviation.

Example G. A news organization polls voters and asks, "Do you intend to vote for incumbent Senator Phillip E. Buster in the upcoming election?" Pollsters record the following numbers.

	Yes	Undecided	No
Male	56	18	47
Female	82	12	35

a) Calculate a point estimate for the proportion of all voters who would vote for Senator Buster.
b) Find the probability (based on this point estimate) that at least 500 out of 1000 voters would favor reelecting Senator Buster. [In the real world, the question would not be phrased this way. Instead, a *confidence interval* (chapter 7) would be calculated.]
c) Calculate point estimates for the proportion of male voters who would vote for Senator Buster vs. the proportion of female voters who would vote for Senator Buster.

Homework exercises

1. You flip a fair coin 100 times. Define random variable Y = proportion of heads.
   a) What is the expected proportion of heads for this experiment?
   b) What is the standard deviation for this experiment?
   c) What is the probability that the proportion of heads is between 0.45 and 0.60?
   d) What is the probability that the proportion of heads is greater than 0.60?

2. A news organization polls voters and asks, "Are you in favor of Proposition 123?" with the results below.

Yes	Undecided	No
72	43	65

a) Calculate a point estimate for the proportion of voters who would vote "Yes", and use it to find the probability that at least 500 out of 1000 voters would favor passing Proposition 123.
b) Include the Undecided voters in your calculation to determine a point estimate for the proportion of voters who *might* vote "Yes", and use it to find the probability that at least 500 out of 1000 voters would favor passing Proposition 123 if *all* of the Undecideds changed to "Yes".