Class Worksheet

April 2 and 4
Math 125 *Kovitz* 2025

The Normal Approximation for Probability Histograms

A probability histogram is a new kind of graph. It represents chance, not data.

A probability histogram represents chance by area.

For example, the histogram for the sum of the draws when rolling a pair of dice is made up of rectangles. The base of each rectangle is centered at a possible value; the area of the rectangle equals the chance of getting that value. The total area of the histogram is 100%.

The probability histogram for the binomial distribution—for example, for the number of heads when a coin is tossed—is close to the normal curve, when the number of binomial trials is large.

The probability histograms of the binomial distributions for a certain p follow the normal curve better and better as the number n of binomial trials goes up.

An empirical histogram, on the other hand, represents experimentally-observed data (where the data are obtained as the result of chance).

As the number of repetitions increases, the empirical histograms converge to the probability histogram.

The Normal Approximation to Binomial Probabilities

If a binomial probability is considered as the sum of repeated draws from a suitable counting box, the normal approximation may be used—provided the number of trials (draws from the box) is suitably large.

The expected value is the product of the number of trials and the average of the counting box. The standard error is the product of the square root of the number of trials and the SD of the counting box (found by the short cut).

Since the sum of the draws is discrete, it is more accurate to correct the endpoints of the intervals by plus or minus one half.

Next convert the endpoints of the given range to standard units using the formula

$$\text{standard units} = \frac{\text{given value (corrected)} - \text{expected value}}{\text{standard error}}.$$

The area under the normal curve between the standard units for the corrected endpoints of the given range will be an approximation for the desired chance.

The Scope of the Normal Approximation

The Central Limit Theorem. When drawing at random with replacement from a box, the probability histogram for the sum will follow the normal curve, even if the contents of the box do not. The histogram must be put into standard units, and the number of draws must be reasonably large.

The closer the histogram of the numbers in the box is to the normal curve, the smaller the number of draws necessary before the normal approximation for the sum of repeated draws from that box takes hold.

This approximation is only valid for sums (including counting problems, which can be considered as sums; for example, binomial chances).

The expected value pins the center of the probability histogram to the horizontal axis, and the standard error fixes its spread.

Problems to think about

Three coins are tossed and the number of heads is recorded. This process was repeated four thousand times. The results included 512 times with no heads, 1463 with one head, 1520 with two heads, and 505 times with three heads.

(a) Draw the probability and the empirical histograms for the number of heads on three coins.

(b) Does the probability histogram follow the normal curve quite closely?

(c) Comment on the similarity of this empirical histogram to the probability histogram.

A coin is tossed 400 times. Estimate the chance of getting—

(a) 217 heads or fewer.

(b) fewer than 218 heads.

A coin is tossed 441 times. Estimate the chance of getting 242 heads and 199 tails.