The Mean and Variance of the Sample Mean

Suppose we have n random variables, $X_1, \ldots, X_n$, all *independent*, and all with the *identical distribution* (sometimes called "i.i.d."). Suppose they have mean μ and variance σ^2.

The average of the X_i is also a random variable, defined by

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

What are the mean and variance of $\bar{X}$?

The mean is $(1/n) \sum_{i=1}^{n} \mu = \mu$.

The variance is $(1/n^2) \sum_{i=1}^{n} \sigma^2 = \sigma^2/n$.

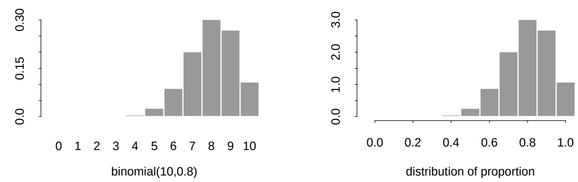
The standard deviation is $\sigma/\sqrt{n}$.

What does this say about how good the sample mean, $\bar{X}$, is as an estimator of μ?

Sampling Distribution for Proportions

If X is the number of successes in n trials, the *proportion* of successes is X/n. We'll call this proportion p, and will regard it as an estimator for π, the actual probability of success.

The distribution for p is just a relabelling of that for X:



This distribution lets us answer questions such as:

> If we ask a sample of 10 people whether they like cornflakes, what is the probability that the proportion who say they do will be less than 2/3 if the proportion in the population is 0.8?

The Mean and Variance of the Sample Proportion

Since $p = X/n$, the mean of p is

$$\mu_X/n = (n\pi)/n = \pi$$

So p is an *unbiased* estimator for π.

The variance of p is

$$\sigma_X^2/n^2 = n\pi(1-\pi)/n^2 = \pi(1-\pi)/n$$

The standard deviation of p is therefore

$$\frac{\sqrt{\pi(1-\pi)}}{\sqrt{n}}$$

This is all a special case of the mean and variance of a sample mean, since

$$p = \frac{1}{n}\sum_{i=1}^{n} S_i$$

where S_i is 0 or 1, indicating failure or success in the ith trial.

The Central Limit Theorem

The sample mean, $\bar{x}$, from n independent observations has close to a normal distribution when n is large.

Specifically, if the population has

mean μ
standard deviation σ

the sample mean is approximately normal with

mean μ
standard deviation $\sigma/\sqrt{n}$

This is true for *any* distribution for which the standard deviation is finite, but how big n needs to be before $\bar{x}$ is close to normal will depend on the distribution.

Practical import: If n is big, we need only find the mean and standard deviation of $\bar{x}$'s sampling distribution. We can then find other things based on the normal distribution.

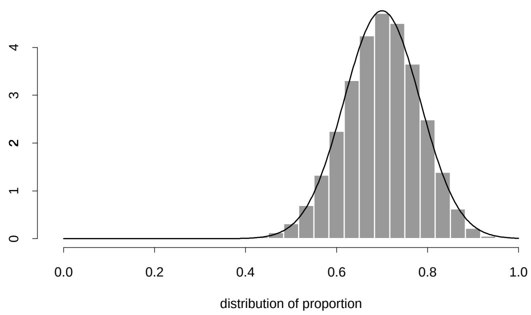
The Central Limit Theorem for the Sample Proportion

Since p can be regarded as a sample mean, the central limit theorem applies. If the proportion in the population is π, the distribution of p for large n is approximately normal, with

mean π
standard deviation $\sqrt{\pi(1-\pi)}/\sqrt{n}$

The approximation is fairly good if $n\pi$ and $n(1-\pi)$ are at least 10.

Here's the approximation for $n = 30$, $\pi = 0.7$:



The Central Limit Theorem for an Exponential Distribution

Here's how $\bar{x}$ approaches a normal distribution when x_i are from the exponential distribution (probability density $f(x) = e^{-x}$, with $x > 0$):