CLARK UNIVERSITY

The sample mean and the central limit theorem
Math 218, Mathematical Statistics

D Joyce, Spring 2016

The distribution of the sample mean. You saw last semester that the sample mean is approximately normally distributed according to the central limit theorem. First we'll review that, then see how it can be used to approximate the binomial distribution.

Review of the Central Limit Theorem. Consider a *random sample*

$$\mathbf{X} = (X_1, X_1, \ldots, X_n),$$

that is, an array of n independent random variables with the same distribution (each X_i is called a *trial*), and suppose that this distribution has mean $\mu = \mu_X$ and variance $\sigma^2 = \sigma^2_X$. The sample sum is

$$S_n = X_1 + X_2 + \cdots + X_n,$$

and it has mean $\mu_{S_n} = n\mu$ and variance $\sigma^2_{S_n} = n\sigma^2$. The sample mean is

$$\overline{X} = \frac{1}{n} S_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n),$$

and it has mean $\mu_{\overline{X}} = \mu$ and variance $\sigma^2_{\overline{X}} = \sigma^2/n$.

The normalized sample mean $\overline{X}^*$ is the same as the normalized sample sum S_n^*

$$\overline{X}^* = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} = S_n^*.$$

The central limit theorem says that this normalized sample mean approaches a standard normal distribution as $n \to \infty$. That means that the unnormalized sample mean $\overline{X}$ is approximately normal with mean $\mu_{\overline{X}} = \mu$ and variance $\sigma^2_{\overline{X}} = \sigma^2/n$, and that the unnormalized sample sum S_n is approximately normal with mean $\mu_{S_n} = n\mu$ and variance $\sigma^2_{S_n} = n\sigma^2$.

Normal approximation to the binomial distribution. Suppose the distribution for the random sample is a Bernoulli distribution with parameter p, and, as usual, let $q = 1 - p$. In other words we have n independent Bernoulli trials $X_1, X_2, \ldots, X_n$, each X_i having a probability p of success, that is, $P(X_i{=}1) = p$, and probability q of failure, that is, $P(X_i{=}0) = q$.

Recall that the mean of the Bernoulli distribution is $\mu = p$ and its variance is $\sigma^2 = pq$.

When each of the independent random variables $X_1, X_2, \ldots, X_n$ is a Bernoulli distribution, then the sum $S_n = \sum_{i=1}^n X_i$ is a binomial distribution with parameters n and p. This binomial distribution has mean $\mu_{S_n} = np$ and variance $\sigma^2_{S_n} = np(1-p) = npq$.

By the central limit theorem, the normalized standardized sum is approximately normal. That means that

$$S_n^* = \frac{S_n - np}{\sqrt{npq}}$$

is approximately a standard normal distribution.

We can use that fact to quickly find probabilities related to this distribution using tables for standard normal distributions. For instance, suppose we want to know the probability of getting no more than 8 heads when flipping a fair coin $n = 20$ times. That means we want to know the probability $P(S_n \leq 8)$. But S_n is approximately normal with mean $\mu_{S_n} = np = 10$ and variance $\sigma^2_{S_n} = npq = 5$. Now, 8 is 2 less than the mean of 10, and 2 is $2/\sqrt{5} = 0.8944$ standard deviations, therefore

$$\begin{aligned} P(S_n \leq 8) &\approx P(Z \leq \frac{8-10}{\sqrt{5}}) \\ &= \Phi(-0.8944) = 0.1867. \end{aligned}$$

(That last figure comes from the standard normal curve areas $\Phi(z)$ on table A.3, page 673.)

But we can do better if we use the so-called continuity correction and use 8.5 instead of 8, to get

$$\begin{aligned} P(S_n \leq 8.5) &\approx P(Z \leq \frac{8.5-10}{\sqrt{5}}) \\ &= \Phi(-0.6708) = 0.2514. \end{aligned}$$

That last figure is very close the exact value of 0.2517.

Math 218 Home Page at
`http://math.clarku.edu/~djoyce/ma218/`