Lecture 14. Central Limit Theorem

11/2/2011

Suppose we have a random sample of n random variables, $X_1, ..., X_n$, drawn from a distribution with finite mean μ and finite variance σ^2 and we are interested in estimating the population mean EX_1. We want to use $\bar{X}_n$ to estimate . We already learned from the law of large numbers that

$$\bar{X}_n \to_p EX_1 \text{ as } n \to \infty. \tag{1}$$

This tells us: as the sample size gets larger and larger, the distribution of the sample mean gets more and more concentrated around the population mean. This piece of information is useful, but too crude. It does not tell us much about how precise $\bar{X}_n$ approximates EX_1 for any finite n. We would like to know more about the distribution of $\bar{X}_n$.

Suppose that $X_1 \sim N(\mu, \sigma^2)$. Then we know exactly what the distribution of $\bar{X}_n$ is: $\bar{X}_n$ is a linear combination of i.i.d. normal random variables, and thus is normally distributed. We have computed the mean and variance of $\bar{X}_n$ in previous lectures: $E\bar{X}_n = \mu$ and $Var(\bar{X}_n) = \sigma^2/n$. Therefore

$$\bar{X}_n \sim N(\mu, \sigma^2/n). \tag{2}$$

With this information, we can compute probabilities like:

$$\Pr(\bar{X} < \mu + k\sigma), \quad \text{or} \quad \Pr(\bar{X}_n > \mu + c\sigma/\sqrt{n}) \tag{3}$$

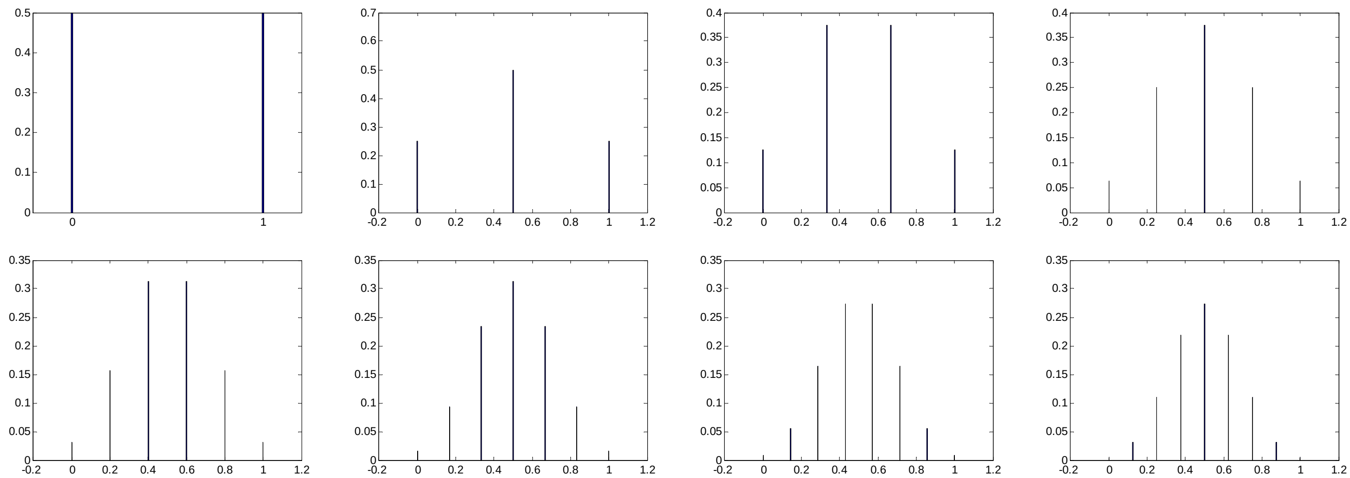
Pretty much all we hope to learn about the distribution of $\bar{X}_n$ is immediately known if the random sample is from a normal distribution.

What if $X_1, ..., X_n$ are not from a normal distribution? Other parametric family of distributions are not as nice as the normal family. Sample averages typically do not belong the same parametric family as the underlying distribution of the sample. It typically is difficult to calculate the exact distribution of $\bar{X}_n$ even if we know the distribution of $(X_1, ..., X_n)$. More disturbingly, the underlying distribution of a random sample is rarely known in practice. However, the seemingly hopeless situation is turned around by the Central Limit Theorem

(CLT).

Before introducing the CLT, let's actually compute the distribution of $\bar{X}_n$ in a non-normal example.

Example. Suppose $(X_1, ..., X_n) \sim i.i.d.\ Bern(0.5)$. Then $n\bar{X}_n \sim B(n, 0.5)$.



Central Limit Theorem: Let $X_1, ..., X_n$ be iid random variables with mean μ and variance $\sigma^2 \in (0, \infty)$. Then,

$$\Pr(\sqrt{n}(\bar{X} - \mu)/\sigma \leq x) \to \Pr(Z \leq x) \text{ as } n \to \infty \text{ for all } x \in R.$$

or alternatively,

$$F_{\sqrt{n}(\bar{X}-\mu)/\sigma}(x) \to \Phi(x) \text{ as } n \to \infty \text{ for all } x \in R,$$

where $Z \sim N(0, 1)$ and $\Phi(x)$ is the cdf of the standard normal distribution.

The CLT tells us that the distribution of the standardized sample mean: $\sqrt{n}(\bar{X}_n - \mu)/\sigma$, converges to the standard normal distribution as the sample size gets large. In other words, the distribution of $\bar{X}_n$ in a small (shrinking) neighborhood of μ gets closer and closer to a normal distribution. Now, we not only know that the sample mean gets closer and closer to the population mean, but also know that the probability $\Pr(\bar{X}_n > \mu + c\sigma/\sqrt{n})$ gets closer and closer to

$$\Pr(Z > c) = 1 - \Phi(c),$$

as n gets larger and larger.

Remark about the CLT: (1) The scaling $\sqrt{n}$ in front of $(\bar{X} - \mu)$ is important. If there is no scaling, all we get is

$$(\bar{X}_n - \mu)/\sigma^2 \to_p 0.$$

The $\sqrt{n}$ stretches the distribution of $\bar{X}_n$ around its probability limit. The larger n is, the more concentrated $\bar{X}_n$ is around its probability limit, the more streching we need. It just so turns out that $\sqrt{n}$ is the appropriate speed at which the stretching should increase. To see why $\sqrt{n}$ is appropriate, recall that $Var(\bar{X}_n)$ is σ^2/n. Then for any scaling factor κ_n, $Var(\kappa_n(\bar{X}_n-\mu)/\sigma) = \kappa_n^2/n$. If κ_n is of order smaller than $\sqrt{n}$, then $Var(\kappa_n(\bar{X}_n-\mu)/\sigma) \to 0$ and we have not stretched enough and the image is still a point mass. If κ_n is of order larger than $\sqrt{n}$, then $Var(\kappa_n(\bar{X}_n-\mu)/\sigma) \to \infty$ and we have stretched too much and we cannot have a whole picture of the distribution of $\bar{X}_n$ around its probability limit. Notice that only the order is important. Suppose that $\kappa_n = 2\sqrt{n}$, then the CLT still works and gives us:

$$\Pr(\kappa_n(\bar{X}_n - \mu)/\sigma \leq x) \to \Pr(2Z \leq x).$$

Notice that the center μ is also very important. We already know that $\bar{X}_n$ concentrates around its probability limit for large n. If we stretch at a point different from the probability limit, we will not find anything interesting.

(2) Why does the limiting distribution need to be normal? The question can be broken into two parts. First, why should the distribution of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ converge to one single distribution, regardless of from what distribution the random sample is drawn? To this question, I can only answer that it's the beauty of Mother Nature. Second, why should the limiting distribution be $N(0,1)$? This is because of the special property of normal distributions: $\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim N(0,1)$ if $(X_1, ..., X_n) \sim i.i.d.\ N(\mu, \sigma^2)$ for every n. Thus, if CLT applies to all random samples from distributions with finite mean and variances, it has to be applicable to normal samples. If it applies to normal samples, the limiting distribution has to be $N(0,1)$.

Approximation using CLT

The way we typically use the CLT result is to approximate the distribution of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ by that of a standard normal. Note that if $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ is exactly a $N(0,1)$ random variable, then X_n is exactly a $N(\mu, \sigma^2/n)$ random variable for any n. Consequently, if $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ is approximately a $N(0,1)$ random variable, then it makes sense to use $N(\mu, \sigma^2/n)$ as an approximating distribution for X_n .

This approximation is not good unless n is sufficient large. How large is sufficiently large is a good question. It depends on the underlying distribution from which the random sample $(X_1, ..., X_n)$ is drawn. Typically, if the underlying distribution does not have thick tails or extreme skewness, n does not need to be very large.