THE CENTRAL LIMIT THEOREM

Throughout the discussion below , let $X_1, X_2, \dots$ be i.i.d. rv's, each with finite expected value μ and finite nonzero standard deviation σ. Given n, define $\overline{X}$ to be the average $(X_1 + \cdots + X_n)/n$, and define S_n to be the sum $X_1 + \cdots + X_n$. Then

$$\begin{aligned} E(S_n) &= n\mu \qquad\qquad \text{and} \\ V(S_n) &= n\sigma^2 \\ \text{st.dev.}(S_n) &= \sigma\sqrt{n}. \end{aligned}$$

(The equation for the variance of S_n holds because the X_i are independent, so the variance of the sum of the X_i is the sum of the variances.) Now, $\overline{X} = (1/n)S_n$, so

$$\begin{aligned} E(\overline{X}) &= \mu & &= (1/n)\ E(S_n) \\ V(\overline{X}) &= \frac{\sigma^2}{n} & &= (1/n)^2\ V(S_n) \\ \text{st.dev.}(\overline{X}) &= \frac{\sigma}{\sqrt{n}} & &= (1/n)\ \text{st.dev.}(S_n) \end{aligned}$$

REMARKS

1. We can think of the i.i.d. condition as meaning that the X_i are repeated experiments, or alternately random samples, from some given probability distribution.

2. Note: the expected value of the sample average is the same as the expected value of each X_i. This is common sense. We can think of $\overline{X}$ as an estimate of the true population average μ.

3. Note, as n gets bigger, the spread (standard deviation) of $\overline{X}$ gets smaller. This is common sense: a bigger sample should give a more reliable estimate of the true population average.

4. Notation: let $\mathcal{N}(\mu, \sigma)$ denote the normal distribution with mean μ and standard deviation σ.

THEOREM (Central Limit Theorem) Suppose that $X_1, X_2, \dots$ is a sequence of i.i.d. rv's, each with finite expected value μ and finite nonzero standard deviation σ. Let Z be the standardized version of $\overline{X}$, i.e.

$$Z = \frac{\overline{X} - \mu}{(\sigma/\sqrt{n})} \ .$$

Then as $n \to \infty$, $Z \longrightarrow \mathcal{N}(0, 1)$.

****

REMARKS

1. Note the CLT has an extra assumption (finite variance) which the LLN does not have. The CLT gives you more information when it is applicable.

2. The CLT is an incredible law of nature. Under modest assumptions, the process of independent repetition has a universal effect on the averaging process, depending only on the mean and standard deviation of the underlying population.

3. There are different ways to describe exactly what $Z \longrightarrow \mathcal{N}(0,1)$ means. We say for this: Z converges in distribution to $\mathcal{N}(0,1)$ as $n \to \infty$. This means just what it meant when we discussed the normal approximation to Binom(n,p) as $n \to \infty$. Informally, it means that if n is large enough, then we have (for all numbers $a < b$)

$$\text{Prob}\Big(a < \frac{(\overline{X} - \mu)}{(\sigma/\sqrt{n})} < b\Big) \approx \text{Prob}\,(a < Z < b)$$

where $\approx$ means "approximately equals". As n goes to infinity, the approximation gets as good as we want.
Equivalently we could express the approximation by

$$\text{Prob}\big(\ a(\sigma/\sqrt{n})\ < (\overline{X} - \mu)\ < b(\sigma/\sqrt{n})\ \big) \approx \text{Prob}\big(a < Z < b\big)$$

or by

$$\text{Prob}\big(\ \mu + a(\sigma/\sqrt{n})\ < \overline{X}\ < \mu + b(\sigma/\sqrt{n})\big) \approx \text{Prob}\big(a < Z < b\big)\ .$$

We might also use the notation $\overline{X} \longrightarrow \mathcal{N}(\mu, \sigma/\sqrt{n})$ to describe this situation.

RULE OF THUMB
How large should n be for the CLT approximation to be good enough? Really, that depends on the particular r.v. X and on how good "good enough" has to be. Our rule of thumb will be that, unless we have explicit information to the contrary, $n \geq 30$ is large enough for "good enough".

EXAMPLE
Let us go through those inequalities in an example, with $a = -2$ and $b = 2$.
We will take the r.v. X corresponding to flipping a fair coin.
So, $X = 0$ with probability 1/2, and $X = 1$ with probability 1/2.
For this X, $\mu = .5$ and $\sigma = \sqrt{(.5)(1 - .5)} = .5$. Let $n = 10{,}000$.
Then the standard deviation for $\overline{X}$ is $(\sigma/\sqrt{n}) = (.5)/\sqrt{10{,}000} = .005$.
Here are the inequalities above with these numbers put in.

$$\text{Prob}\Big(-2 < \frac{(\overline{X} - .5)}{.005} < 2\Big) \approx \text{Prob}\,(-2 < Z < 2)$$

$$\text{Prob}\big(\ -2(.005)\ < (\overline{X} - .5)\ < 2(.005)\ \big) \approx \text{Prob}\big(-2 < Z < 2\big)$$

$$\text{Prob}\big(\ .5 - 2(.005)\ < \overline{X}\ < .5 + 2(.005)\big) \approx \text{Prob}\big(-2 < Z < 2\big)\ .$$

If we compute out the last line, we get

$$\begin{aligned}\text{Prob}\big(\ .49\ < \overline{X}\ < .51\big)\ \approx\ \text{Prob}\big(-2 < Z < 2\big)\ &= \text{Prob}(Z \le 2) - \text{Prob}(Z \le -2)\\ &= .9772 - .0228 = .9544.\end{aligned}$$

This means: if the experiment is to flip a fair coin 10,000 times:
then in about 95% of those experiments,
the percentage of the flips which equal heads will be between 49% and 51%.

SUMS AND AVERAGES
Let us look again at one of the ways to express the CLT:

$$\text{Prob}\big(\ a(\sigma/\sqrt{n})\ < (\overline{X} - \mu)\ < b(\sigma/\sqrt{n})\ \big)\ \approx\ \text{Prob}\big(a < Z < b\big)\ .$$

Remember, $\overline{X} = (X_1 + \cdots + X_n)/n$. If we multiply each element of the inequality on the left by n, we don't change the truth of the inequality, so we don't change its probability. So we get

$$\text{Prob}\big(a\sigma\sqrt{n}\ \ < (X_1 + \cdots + X_n) - n\mu\ < b\sigma\sqrt{n}\ \big)\ \approx\ \text{Prob}\big(a < Z < b\big)\ .$$

Sometimes it's easier to think in terms of sums.

For example, suppose $X_1, \ldots, X_{10{,}000}$ are i.i.d. random variables corrsponding to 10,000 flips of a fair coin. For each X_i, the mean μ is 1/2; the variance is 1/4; and the standard deviation σ is 1/2. Let $S = X_1 + \cdots + X_{10{,}000}$. The standard deviation of S is $\sqrt{10,000}\,\sigma$, which is $(100)(1/2) = 50$. The mean of S is $5,000$.

Now what is the probability that the number of heads seen is in the interval $[4900, 5100]$? That's the interval of numbers within two standard deviations of the mean, so the probability is about .95.

What's the probability that the number of heads is not in the interval $[4800, 5200]$? That's the probability of being more than four standard deviations away from the mean. That probability is about .000 063 .