§ 7. The Weak Law of Large Numbers and the Central Limit Theorem

Both theorems deal with the situation of what happens in the limit for large numbers of repeated, independent, trials.

Let $X_1, X_2, \ldots, X_n, \ldots$ be an infinite sequence of indep. r.v's each with the same mean μ and variance σ^2 (e.g. $X_1, X_2, \ldots, X_n$ i.i.d with mean μ, var σ^2).

Set $S_n = X_1 + X_2 + \cdots + X_n$, $n \geq 1$ (sum)

and $\overline{X}_n = \frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n}$, $n \geq$ (average).

Then

$$E(\bar{X}_n) = E\left(\frac{X_1 + \cdots + X_n}{n}\right)$$

$$= \frac{1}{n} E(S_n)$$

$$= \frac{1}{n} \sum_i E(X_i)$$ Thm 4.14, p. 153

$$= \frac{1}{n} \cdot \sum_i \mu$$

$$= \frac{n\mu}{n} = \mu.$$

$$V(\bar{X}_n) = V\left(\frac{X_1 + \cdots + X_n}{n}\right)$$

$$= V\left(\frac{X_1}{n} + \cdots + \frac{X_n}{n}\right)$$

by independence and Cor 4.3 p. 154

$$= \sum_{i=1}^{n} \frac{1}{n^2} V(X_i)$$

$$= \sum_{i=1}^{n} \frac{1}{n^2} \sigma^2$$

$$= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

The important facts here are that

- $\bar{X}_n$ has the same mean μ as all the X_i's.

- $V(\bar{X}_n) \to 0$ as $n \to \infty$.

If we apply Chebyshev to $\bar{X}_n$, then for any $\varepsilon > 0$

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{V(\bar{X}_n)}{\varepsilon^2}$$

$$= \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0 \text{ as } n \to \infty.$$

We have proved

<u>Thm</u> Weak Law of Large Numbers (WLLN)

If $X_1, X_2, \ldots, X_n, \ldots$ are indep. with the same fte mean μ & fte. var σ^2, then for any $\varepsilon > 0$

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0.$$

(Book Thm 8.2, p. 267)

This is an example of <u>convergence in probability</u>. There are other versions of this result which involve other types of convergence (the Strong Law of Large Numbers).

The conclusion of the WLLN is actually very intuitive. Basically it says that one converges to the average over more and more repeated trials.

e.g. If I keep flipping a coin, I expect that the proportions of heads & tails get closer and closer to 50% the more I flip.

Can use the estimate

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

to see how fast $\bar{X}_n$ converges to the mean μ.

Ex. I flip a fair coin repeatedly.
How many times do I need to flip in order to be at least 95% sure the proportions of heads and tails are within 1% of 50%.

Here each X_i is a Bernoulli $(\frac{1}{2})$ r.v. with mean $\frac{1}{2}$ & variance $\frac{1}{2}(1-\frac{1}{2}) = \frac{1}{4}$.

Here $\varepsilon = .01$

Want

$$P(|\bar{X}_n - .5| \geq .1) \leq \frac{(\frac{1}{4})^2}{n(.01)^2} < .05$$

ie

$$\frac{(\frac{1}{4})^2}{n(.01)^2} \leq .05$$

so

$$n \geq \frac{(\frac{1}{4})^2}{.05(.01)^2} = 12,500 \text{ flips.}$$

One can think of the r.v.'s $\bar{X}_n = \frac{S_n}{n}$ in WLLN as being normalized so that the means $E(\bar{X}_n) = \mu$ are constant.

Now suppose we consider the r.v.'s.

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

Then $E(Z_n) = \frac{1}{\sigma/\sqrt{n}}\left(E(\bar{X}_n) - \mu\right) = 0$

and

$$V(Z_n) = E\left((Z_n - E(Z_n))^2\right)$$

$$= E(Z_n^2) \quad \text{as} \quad E(Z_n) = 0$$

$$= E\left(\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right)^2\right)$$

$$= \frac{1}{\sigma^2/n} E\left(\left(\overline{X}_n - \mu\right)^2\right)$$

$$= \frac{1}{\sigma^2/n} \cdot V(\overline{X}_n) \qquad \text{by def}^{\text{n}} \text{ of variance}$$

$$= \frac{\sigma^2/n}{\sigma^2/n} = 1.$$

Thus the r.v.'s Z_n can be thought of as a way of normalizing the sums S_n so that they all have mean 0 and variance 1.

Not only are these the mean and variance of the std. normal, but the sequence of r.v.'s converges to a standard normal in the following sense

<u>Thm</u> Central Limit Theorem.

If $X_1, X_2, \ldots, X_n, \ldots$ are i.i.d r.v's with the same fte mean μ & the same fte variance σ^2, then for any $z \in \mathbb{R}$

$$\lim_{n \to \infty} P(Z_n \leq z) = \Phi(z)$$

or

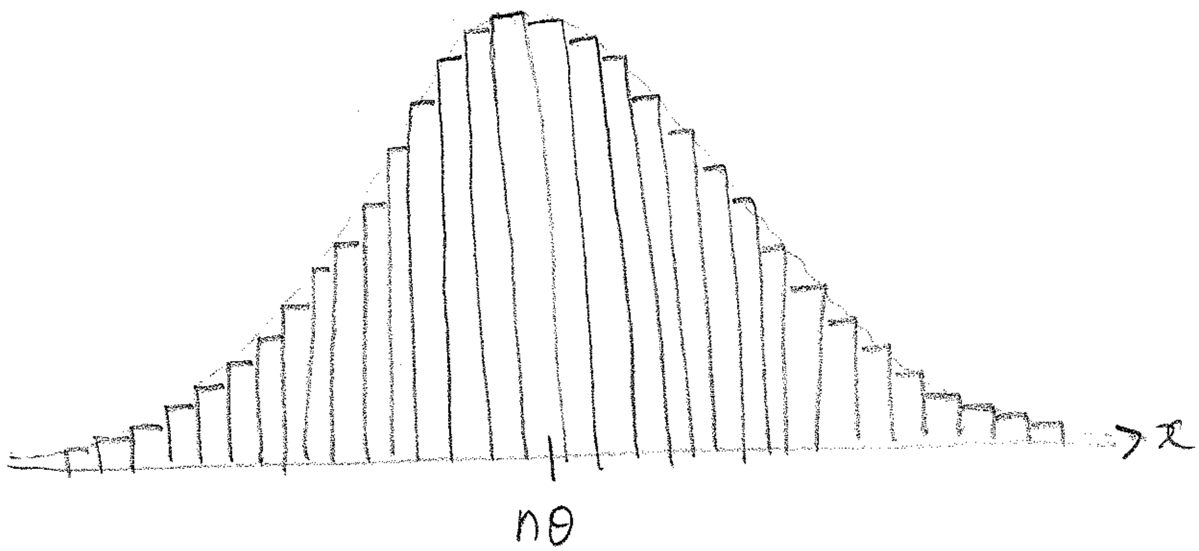
$$\lim_{n \to \infty} P\left(\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \leq z \right) = \Phi(z)$$

where $\Phi(z)$ is the cdf of the standard normal.

The convergence in the statement of this theorem is an example of <u>convergence in distribution</u> (also called <u>weak convergence</u>).

<u>Fact</u> A binomial (n, θ) r.v. with n large is approx a normal $(n\theta, n\theta(1-\theta))$ r.v. in the sense of distributions.

Picture



Example Suppose X is a binomial (n, θ) r.v. Then as we saw earlier

$$X = X_1 + \cdots + X_n$$ where $X_1, \ldots, X_n$ are i.i.d. Bernoulli(θ) r.v.'s.

From earlier (Thm 5.2 P. 168), X has mean $\mu = n\theta$, & variance $\sigma^2 = n\theta(1-\theta)$.

Then if n is large, by CLT

$$Z = \frac{X - n\theta}{\sqrt{n\theta(1-\theta)}}$$

is approx a normal $(0,1)$ in distr.

This then means that

$$X = \sqrt{n\theta(1-\theta)}\, Z + n\theta$$

is approx a normal $(n\theta, n\theta(1-\theta))$ in distr.

Ex. Let X be a binomial $(30, \frac{1}{6})$ r.v.

Use the CLT to find the approx. prob. that $X = 3$.

X has mean $30 \times \frac{1}{6} = 5$

and variance. $30 \times \frac{1}{6} \times (1 - \frac{1}{6}) = \frac{25}{6}$

By CLT X is approx a normal $(5, \frac{25}{6})$

and $\dfrac{X-5}{\sqrt{25/6}}$ is therefore approx a std. normal.

Then

$$P(X=3) = P(2.5 \leq X \leq 3.5)$$

(X is discrete).

$$= P\left(\frac{2.5-5}{\sqrt{25/6}} \leq \frac{X-5}{\sqrt{25/6}} \leq \frac{3.5-5}{\sqrt{25/6}}\right)$$

$$= P(-2.69 \leq Z \leq -.73)$$ where Z is a std. normal

$$= \Phi(-.73) - \Phi(-2.69)$$

$$\approx .2291.$$