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

Theorem: The Central Limit Theorem (one version)

Let X_1, X_2, … X_n be a random sample of size n from a random variable (population) X with mean μ and variance σ^2. The limiting, $n \to \infty$, distribution of

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \qquad \text{is} \qquad N(0,1)$$

provided X has a moment generating function.. In the formula

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

is the sample mean.

Proof.

Let $M_X(t) = \sum_{i=0}^{\infty} \mu_i \frac{t_i}{i!}$, where $\mu_i = E(X^i)$. Then

$$M_{\bar{X}(t)} = M_{\frac{1}{n}\sum X_i}(t) = M_{\sum X_i}\left(\frac{t}{n}\right) = \left[M_X\left(\frac{t}{n}\right)\right]^n$$

$$\Rightarrow \log\, M_{\bar{X}}(t) = n \log M_X\left(\frac{t}{n}\right)$$

The Taylor series for $\log(1+x)$ is $\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots$

$$\begin{aligned}\Rightarrow \quad \log M_X\left(\frac{t}{n}\right) = &\left\{\mu_1 \frac{t}{n} + \mu_2 \frac{t^2}{2!n^2} + \mu_3 \frac{t^3}{3!n^3} + \cdots\right\} \\ &- \frac{1}{2}\left\{\mu_1 \frac{t}{n} + \mu_2 \frac{t^2}{2!n^2} + \mu_3 \frac{t^3}{3!n^3} + \cdots\right\}^2 \\ &+ \frac{1}{3}\left\{\mu_1 \frac{t}{n} + \mu_2 \frac{t^2}{2!n^2} + \mu_3 \frac{t^3}{3!n^3} + \cdots\right\}^3 - \cdots\end{aligned}$$

$$\Rightarrow \quad n \log M_X\left(\frac{t}{n}\right) = \mu_1 t + \frac{\sigma^2}{n} \cdot \frac{t^2}{2} + O\left(\frac{1}{n^2}\right) \qquad \text{as} \qquad n \to \infty$$

$$\{\sigma^2 = \mu_2 - \mu_1^2\}$$

This shows that

$$M_{\bar{X}} \xrightarrow[n\to\infty]{} e^{\mu t + \frac{(\sigma t)^2}{2n}}$$

and this is the moment generating function for the RV $\sim N\left(\mu, \frac{\sigma^2}{n}\right)$

□

Example 1

$X \sim \text{Binom}(n, p)$ is a sum of n Bernoulli random variables Y_i :

$$Y_i = \begin{cases} 1 & ; \quad p \\ 0 & ; \quad 1-p \end{cases}$$

$M_Y(t) = \sum_x e^{ty} p(y) = e^t p + (1-p) \Rightarrow M_X(t) = [e^t p + (1-p)]]^n \quad \Rightarrow$
The moment generating function of binomial.
Let $\mu = np, \quad \sigma^2 = np(1-p), \quad$ and $Z = \dfrac{X-\mu}{\sigma}$

$$\begin{aligned}
\log M_Z(t) &= \log\left\{e^{-\frac{\mu t}{\sigma}}\left[e^{\frac{t}{\sigma}} p + (1-p)\right]^n\right\} \\
&= -\frac{\mu t}{\sigma} + n \log\left[(1-p) + p\left(1 + \frac{t}{\sigma} + \frac{t^2}{2!\sigma^2} + \frac{t^3}{3!\sigma^3} + \ldots\right)\right] \\
&= -\frac{\mu t}{\sigma} + n \log\left[1 + p\left(\frac{t}{\sigma} + \frac{t^2}{2!\sigma^2} + \frac{t^3}{3!\sigma^3} + \cdots\right)\right] \\
&= -\cancel{\frac{\mu t}{\sigma}} + n\left(\cancel{\frac{pt}{\sigma}} + \frac{pt^2}{2!\sigma^2} + \frac{pt^3}{3!\sigma^3} + \cdots\right) - \frac{n}{2}\left(\frac{pt}{\sigma} + \frac{pt^2}{2\sigma^2} + \cdots\right)^2 + \cdots \\
&= \frac{t^2}{2}\left[\frac{np}{\sigma} - \frac{np^2}{\sigma^2}\right] + O\left(\frac{1}{n}\right) \\
&= \frac{t^2}{2}\left[\frac{np(1-p)}{\sigma^2}\right] + O\left(\frac{1}{n}\right) \\
&= \frac{t^2}{2} + O\left(\frac{1}{n}\right)
\end{aligned}$$

Thus $Z \xrightarrow[n\to\infty]{} \sim N(0,1)$

Example 2

Use CLT to approximate the probability that the sum of 100 tosses of a fair die is in the range 340-360.

For a fair die we have $\mu = 3.5, \sigma = 1.708$. The sample averages for the boundaries of the given range are $\bar{x}_1 = 340/100 = 3.4$ and

$\bar{x}_2 = 360/100 = 3.6$. The left boundary sample z-score is

$$z_1 = \frac{x_1 - \mu}{\sigma/\sqrt{n}} = \frac{3.4 - 3.5}{1.708/\sqrt{100}} = -0.59$$

and by symmetry $z_2 = 0.59$. Thus

$$p(340 \le \sum X_i \le 360) \approx p(3.4 \le \bar{X} \le 3.6) =$$

$$p(-0.59 \le z \le 0.59) = 0.7224 - 0.2776 = 0.4448$$

The Weak Law of Large Numbers (WLLN)

The CLT allows us to solve he following problem: we have a random variable (population) with unknown mean, μ; determine μ with a given accuracy from a (large) sample. Here is how:

Let $X_1, X_2, \ldots, X_n$ be a random sample from a random variable X with finite mean and variance: $E(X_i) = E(X) = \mu$; $\quad \text{Var}(X_i) = \text{Var}(X) = \sigma^2$

Let $\bar{X} = \frac{1}{n}\sum x_i$ be the sample mean.

Then $E(\bar{X}) = \mu$ and $\text{Var}(\bar{X}) = \dfrac{\text{Var}(X)}{n}$

Applying Chebyshev's inequality for some k, we have

$$P\left(|\bar{x} - \mu| \le k\frac{\sigma}{\sqrt{n}}\right) \ge 1 - \frac{1}{k^2}, \quad k > 0$$

Let $\varepsilon = k\dfrac{\sigma}{\sqrt{n}}$, i.e. $k = \dfrac{\varepsilon\sqrt{n}}{\sigma}$. We have

$$P(|\bar{x} - \mu| \le \varepsilon) \ge 1 - \frac{\sigma^2}{n\varepsilon^2}$$

so we have The Weak Law of Large Numbers.

Law: The Weak Law of Large Numbers

For arbitrary (small) $\varepsilon > 0$, as $n \to \infty$.

$$P(|\bar{x} - \mu| \le \varepsilon) \xrightarrow[n\to\infty]{} 1.$$

WLLN specifies that for any nonzero margin (ε), no matter how small, with n sufficiently large sample, the probability approaches 1 that the sample average is closer to the expected value then the margin. (This is convergence in probability; different than the pointwise convergence in Calculus).

Example 3

Let X be a Bernoulli random variable:

$$P(X) = \begin{cases} p & ; \quad \text{if} \quad x = 1 \\ 1 - p & ; \quad \text{if} \quad x = 0 \end{cases}$$

Then $E(X) = p, \quad \text{Var}(X) = p(1-p)$

Let $X_1, X_2, \ldots, X_n$ be a random sample of X. Let $\bar{x}$ be the sample mean. Then

$$E(X) = p \qquad \text{Var}(\bar{x}) = \frac{p(1-p)}{n}$$

$$\text{Prob}\left(|\bar{x} - p| \le k\sqrt{\frac{p(1-p)}{n}} \right) \ge 1 - \frac{1}{k^2}$$

with $\varepsilon = k\sqrt{\dfrac{p(1-p)}{n}}$

$$\text{Prob}(|\bar{x} - p| \le \varepsilon) \ge 1 - \frac{p(1-p)}{n \cdot \varepsilon^2}$$

so we can use (large) random samples to estimate the probability for success in Bernoulli (hence in Binomial) with any given precision.