8.3. The Central Limit Theorem

- Recall the ~~CLT for B~~ particular case of CLT — the normal approximation to Binomial (5.4) p. 60

De Moivre-Laplace CLT: $\boxed{\text{Binom}(n,p) \approx N(np,\ np(1-p))}$ if $n \to \infty$ and $p = \text{const}$.

In other words, if $X \sim \text{Binom}(n,p)$, then the Z-score

$$Z_n = \frac{X - np}{\sqrt{np(1-p)}} \approx N(0,1);$$

satisfies $\forall a \in \mathbb{R}$;

$$F_{Z_n}(a) \longrightarrow \Phi(a) \quad \text{as } n\to\infty,\ p \text{ fixed}.$$

F_{Z_n}: cdf of Z; Φ: cdf of N(0,1).

- Recall that if $X \sim \text{Binom}(n,p)$, we can represent X as a sum of indep. r.v's:

$$X = X_1 + \dots + X_n, \qquad X_i \sim \text{Bernoulli}(p);$$

$$X_i = \begin{cases} 1, & \text{with prob. } p \\ 0, & \text{with prob. } 1-p \end{cases}$$

- General CLT holds for sums of indep. r.v's X_i with general distr, not just Bernoulli.

Thus let X_i be indep. r.v's with means μ, variances σ^2,

$$S_n = X_1 + \dots + X_n \qquad \text{Hence} \quad E[S_n] = \mu n, \quad Var(S_n) = \sigma^2 n.$$

General CLT: $\boxed{S_n \approx N(\mu n, \sigma^2 n)}$

if $n \to \infty$; μ, σ fixed.

Central Limit Theorem Let X_i be indep r.v's, $E(X_i) = \mu$, $Var(X_i) = \sigma^2$.

Consider $S_n = X_1 + \dots + X_n$.

Then the Z-score

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

satisfies $\forall a \in \mathbb{R}$:

$$F_{Z_n}(a) \to \Phi(a) \quad \text{as } n \to \infty.$$

Remarks 1) CLT implies $\forall a \le b \in \mathbb{R}$:

$$P\{a \le Z \le b\} \longrightarrow \Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx$$

$$\overset{\shortparallel}{F(b) - F(a)}$$

~~2) Universality.~~ 2) Implies De Moivre - Laplace CLT for $X_i \sim \text{Bernoulli}(p)$.
$(\mu = p,\ \sigma^2 = p(1-p))$

3) Universality.

Proof of CLT

Using mgf's: • Without loss of generality, assume that $\mu = 0$, $\sigma = 1$. (Why? Consider $\frac{X_i - \mu}{\sigma}$ instead of X_i).

• Compute mgf of $Z_n = \frac{S_n}{\sqrt{n}} = \frac{X_1}{\sqrt{n}} + \dots + \frac{X_n}{\sqrt{n}}$

$$M_{Z_n}(t) = M_{X_1/\sqrt{n}}(t) \cdots M_{X_n/\sqrt{n}}(t) = \left(M_{X_1/\sqrt{n}}(t)\right)^n.$$

• Asymptotic analysis of

$$M_{X_1/\sqrt{n}}(t) = E\left[\exp\left(\frac{tX_1}{\sqrt{n}}\right)\right]$$

Taylor's expansion $e^x = 1 + x + \frac{x^2}{2} + o(x^2)$.

$$\Rightarrow \exp\left(\frac{tX_1}{\sqrt{n}}\right) = 1 + \frac{tX_1}{\sqrt{n}} + \frac{t^2X_1^2}{2n} + o\left(\frac{t^2X_1^2}{2n}\right)$$

$$E\left[\exp\left(\frac{tX_1}{\sqrt{n}}\right)\right] = 1 + \frac{t}{\sqrt{n}} \underbrace{E[X_1]}_{=0} + \frac{t^2}{2n} \underbrace{E[X_1^2]}_{=1} + o\left(\frac{\overset{\text{fixed}}{t^2}\ \overset{1}{E[X_1^2]}}{2n}\right)$$

$$= 1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right).$$

• $\Rightarrow M_{Z_n}(t) = \left[1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right]^n = e^{t^2/2 + o(1)} \to e^{t^2/2}$ as $n \to \infty$

• But $e^{t^2/2} = M_{N(0,1)}(t)$, the mgf of $N(0,1)$!

• Thus, in the limit as $n \to \infty$, $Z_n \to N(0,1)$.
(by the uniqueness theorem). Actually, we invoke a limiting version of uniqueness thm (if $M_{Z_n}(t) \underset{\forall t}{\to} M_Z(t)$ then $F_{Z_n}(t) \to F_Z(t)$ $\forall t$, for continuous Z)
This is "Levy's Continuity Lemma" (Lemma 3.1 in the book) QED.

Ex | Suppose that, whenever invited to a party, a person attends alone with prob. 1/3, attends with a guest with prob. 1/3, and does not attend with prob. 1/3.
A company invited all 300 of its employees and their guests to a party.
What is the prob. that at least 320 will attend?

X_i = # of people employee i will bring, including him/her self.

X_i are independent r.v's with pmf

$$P\{X_i=0\} = P\{X_i=1\} = P\{X_i=2\} = 1/3.$$

Thus $\mu = E[X_i] = 1$, $\sigma^2 = Var(X_i) = 2/3$ (check!)

# of people attending the party is

$$S_n = X_1 + \cdots + X_n, \qquad n = 300.$$

$$P\{S_n \geq 320\} = P\left\{ \frac{S_n - n\mu}{\sigma\sqrt{n}} \geq \underbrace{\frac{320 - n\mu}{\sigma\sqrt{n}}}_{\frac{320-300\cdot 1}{\sqrt{\frac{2}{3}\cdot 300}} = 1.41} \right\} \overset{CLT}{=} 1 - \Phi(1.41) = 1 - 0.921 = 0.079 \approx 8\%$$

Remark: $S_n \approx N(\mu n, \sigma^2 n) = N(300, 200)$.