§5.4.1 The Normal Approximation to the Binomial Distribution

De Moivre-Laplace Central Limit Theorem

Let $S_n$ be a binomial random variable with parameters $(n, p)$. Thus $S_n = \sum_{i=1}^{n} X_i$, where $X_i$ are $n$ independent Bernoulli-$p$ random variables. Then

$$\lim_{n \to \infty} \text{P} \left( \frac{S_n - np}{\sqrt{np(1-p)}} \leq b \right) = \text{P}(\alpha \leq N(0,1) \leq b) = \Phi(b) - \Phi(\alpha).$$

Normal Approximation to the Binomial Distribution

Let $S_n$ be a binomial random variable with parameters $(n, p)$. Let $\alpha$ and $\beta$ be nonnegative integers satisfying $0 \leq \alpha \leq \beta \leq n$. Then

$$\text{P}(\alpha \leq S_n \leq \beta) = \text{P}(\alpha - \frac{1}{2} \leq S_n \leq \beta + \frac{1}{2}) \quad \text{continuity correction}$$

$$= \text{P} \left( \frac{\alpha - \frac{1}{2} - np}{\sqrt{np(1-p)}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{\beta + \frac{1}{2} - np}{\sqrt{np(1-p)}} \right)$$

$$\approx \text{P} \left( \frac{\alpha - \frac{1}{2} - np}{\sqrt{np(1-p)}} \leq N(0,1) \leq \frac{\beta + \frac{1}{2} - np}{\sqrt{np(1-p)}} \right) = \Phi(\delta) - \Phi(\gamma),$$

where $\delta = \frac{\beta + \frac{1}{2} - np}{\sqrt{np(1-p)}}$ and $\gamma = \frac{\alpha - \frac{1}{2} - np}{\sqrt{np(1-p)}}$.

For $-\infty < \gamma < \delta < \infty$,

$$|\text{P} \left( \frac{S_n - np}{\sqrt{np(1-p)}} \leq \delta \right) - \text{P}(\delta \leq N(0,1) \leq \gamma)| \leq \frac{C}{\sqrt{n}},$$

where $C < \infty$ is a constant independent of $n$. This is the Berry-Esseen Theorem.
Example 2

Let $S_n$ be binomial with $n = 90$, $p = \frac{1}{2}$. Since $np = 20$ and $\sqrt{np(1-p)} = 10$, we have

$$P(S_n = 20) = \frac{P(19.5 < S_n < 20.5)}{\sqrt{10}} = \frac{P\left(\frac{19.5-20}{10} < \frac{S_n-20}{\sqrt{10}} < \frac{20.5-20}{10}\right)}{\sqrt{10}}$$

$$= P\left(-1.16 < \frac{S_n-20}{\sqrt{10}} < 1.16\right) \\
\approx P\left(-1.16 < N(0,1) < 1.16\right) = .8772$$

Compare this with $P(S_n = 20) = \left(\frac{10}{20}\right)^{10} = .1854$

HW #2 #23. Let $S_n$ be binomial with $n = 1000$, $p = \frac{1}{6}$.

Thus $np = 1000/6$, $\sqrt{np(1-p)} = \sqrt{1000 \cdot \frac{1}{6} \cdot \frac{5}{6}}$.

Write these numbers as $U = np$, $V = \sqrt{np(1-p)}$. We have

$$P(150 < S_n < 200) = P(149.5 < S_n < 200.5)$$

$$= P\left(\frac{149.5-U}{V} < \frac{S_n-U}{V} < \frac{200.5-U}{V}\right)$$

$$\approx P\left(\frac{149.5-U}{V} < N(0,1) < \frac{200.5-U}{V}\right)$$

Verify: $U = 166.67$, $V = 11.79$, $\frac{149.5-U}{V} = -1.46$, $\frac{200.5-U}{V} = 2.87$,

and probability involving $N(0,1) \approx \Phi(2.87) + \Phi(-1.46) - 1 = .9258$

Compare Poisson approximation (PA) and normal approximation (NA).

The error in PA is bounded by $np^2$.

The error in NA is bounded by $\text{const}/\sqrt{n}$.

So for fixed $p$ and large $n$ we expect PA not to be accurate (in examples above $np^2 = 90(1/2)^2 = 10$ and $1000(1/6)^2 = 27.8$).

For fixed $p$ and large $n$ we expect NA to be accurate (in examples above $\sqrt{n} = 10$ and $\Phi(10.35)$ actual error in 49 is .0018).