§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 for $-\infty < a < b < \infty$

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

Normal Approximation To The Binomial Distribution

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

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

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

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

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

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

Examples

4g, pp 194–195

Let S_n be binomial with $n = 40$, $p = 1/2$. Since $np = 20$ and $\sqrt{np(1-p)} = \sqrt{10}$, we have

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

$$= \left(\frac{-.5}{\sqrt{10}} \le \frac{S_n-20}{\sqrt{10}} \le \frac{.5}{\sqrt{10}}\right) = P\left(-.16 \le \frac{S_n-20}{\sqrt{10}} \le .16\right)$$

$$\approx P(-.16 \le N(0,1) \le .16) = \Phi(.16) - \Phi(-.16) = .1272$$

Compare this with $P(S_n = 20) = \binom{40}{20}\left(\frac{1}{2}\right)^{40} = .1254$

HW#7, #23 Let S_n be binomial with $n = 1000$, $p = 1/6$.

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

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

$$P(150 \le S_n \le 200) = P(149.5 \le S_n \le 200.5)$$

$$= P\left(\frac{149.5-u}{v} \le \frac{S_n - u}{v} \le \frac{200.5-u}{v}\right)$$

$$\approx P\left(\frac{149.5-u}{v} \le N(0,1) \le \frac{200.5-u}{v}\right)$$

Verify: $u = 166.67$, $v = 11.79$, $\frac{149.5-u}{v} = -1.46$, $\frac{200.5-u}{v} = 2.87$, and probability involving $N(0,1)$ is $\Phi(2.87) + \Phi(1.46) - 1$
$= .9258$

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 $np^2 = 40(1/2)^2 = 10$ and $1000(1/6)^2 = 27.8$).

For fixed p and large n we expect NA to be accurate (in examples above $1/\sqrt{n} = .16$ and $.03$; actual error in 4g is $.0018$)