The Normal Approximation of the Binomial Distribution

Joyce Tipping

Wednesday, November 19, 2008

Why?

When n gets large, the binomial probabilities become difficult to calculate.

Very Quick Review

Random Variable: A function that assigns a probability to each possible outcome.

Very Quick Review

Random Variable: A function that assigns a probability to each possible outcome.

Expected Value: The long-run expected outcome of a random variable.

Very Quick Review

Random Variable: A function that assigns a probability to each possible outcome.

Expected Value: The long-run expected outcome of a random variable.

Variance: A measurement of the spread of a random variable.

The Bernoulli

A Bernoulli experiment has the following properties:

1. One trial
2. Two outcomes: Success or Failure
3. Known probability of success

The Binomial

A binomial experiment has the following properties:

1. n trials, where n is known
2. The trials are independent
3. Each trial has two outcomes: Success or Failure
4. Probability of success is known and fixed

The Binomial

A binomial experiment can be thought of as a sum of independent Bernoulli trials.

The Central Limit Theorem

Let $X_1, \ldots, X_n$ be a random sample from a distribution with mean μ and variance $\sigma^2 < \infty$. Then,

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

approaches the standard normal, $Z \sim N(0, 1)$ as $n \to \infty$.

The Central Limit Theorem

$$
\begin{aligned}
& \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \\
= \quad & \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\frac{\sigma}{\sqrt{n}}} \\
= \quad & \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma}
\end{aligned}
$$

The Central Limit Theorem

Let $X_1, \ldots, X_n$ be a random sample from a distribution with mean μ and variance $\sigma^2 < \infty$. Then,

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma}$$

approaches the standard normal, $Z \sim N(0, 1)$, as $n \to \infty$.

The Approximation

Let $X_i \sim Ber(p)$.
$\mu_x = p$
$\sigma_x^2 = pq \longrightarrow \sigma_x = \sqrt{pq}$

The Approximation

Let $X_i \sim Ber(p)$.

$\mu_X = p$

$\sigma_X^2 = pq \longrightarrow \sigma_X = \sqrt{pq}$

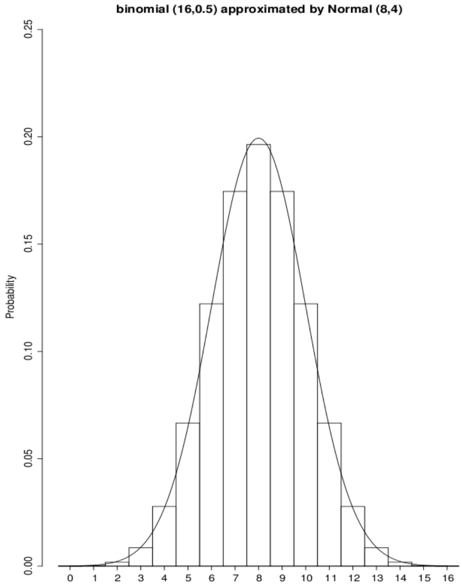
Then, we have $\sum_{i=1}^{n} X_i \sim bin(n, p)$.

Therefore, by the CLT, $bin(n, p)$ is approximately normal with

$$\mu = n \cdot \mu_X = np$$
$$\sigma = \sqrt{n} \cdot \sigma_X = \sqrt{npq}$$

whenever n is large.

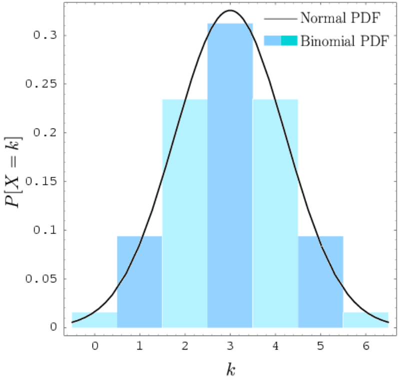
Example


binomial (16,0.5) approximated by Normal (8,4)
Probability
0.00
0.05
0.10
0.15
0.20
0.25
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

When It Works

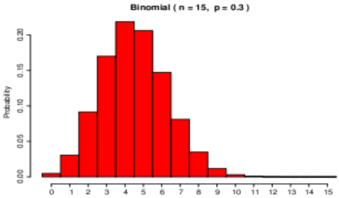
The normal approximation works well when the binomial is symmetric:

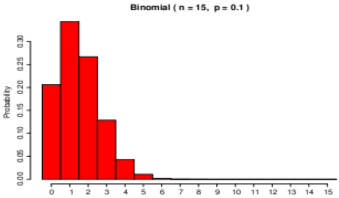
1. p is close to 0.5 OR
2. n is very large



Coming Up Soon ...

When p is not close to 0.5 and n is not large, the binomial is skewed.





In this case, the skew-normal is a better approximation.