Math 302.102 Fall 2010
The Normal Approximation to the Binomial Distribution

Suppose that X is a Bernoulli(p) random variable which means that

$$\mathbf{P}\{X=1\}=p \quad \text{and} \quad \mathbf{P}\{X=0\}=1-p$$

for $0 \leq p \leq 1$. Furthermore,

$$\mathbb{E}(X)=1\cdot\mathbf{P}\{X=1\}+0\cdot\mathbf{P}\{X=0\}=p$$

and

$$\mathbb{E}(X^2)=1^2\cdot\mathbf{P}\{X=1\}+0^2\cdot\mathbf{P}\{X=0\}=p$$

so that

$$\mathrm{Var}(X)=\mathbb{E}(X^2)-[\mathbb{E}(X)]^2=p-p^2=p(1-p).$$

Suppose now that $X_1, X_2, \ldots, X_n$ are iid Bernoulli(p) random variables. If we let

$$Y=X_1+X_2+\cdots+X_n,$$

then

$$\mathbb{E}(Y)=\mathbb{E}(X_1+X_2+\cdots+X_n)=\mathbb{E}(X_1)+\cdots+\mathbb{E}(X_n)=p+p+\cdots+p=np$$

and

$$\begin{aligned}\mathrm{Var}(Y)=\mathrm{Var}(X_1+X_2+\cdots+X_n)&=\mathrm{Var}(X_1)+\cdots+\mathrm{Var}(X_n)\\&=p(1-p)+p(1-p)+\cdots+p(1-p)\\&=np(1-p).\end{aligned}$$

Thus, if

$$\overline{X}=\frac{X_1+X_2+\cdots+X_n}{n}=\frac{Y}{n},$$

then

$$\mathbb{E}(\overline{X})=\frac{\mathbb{E}(Y)}{n}=\frac{np}{n}=p$$

and

$$\mathrm{Var}(\overline{X})=\frac{\mathrm{Var}(Y)}{n^2}=\frac{np(1-p)}{n^2}=\frac{p(1-p)}{n}.$$

By the central limit theorem, we know that the limiting distribution of

$$\frac{\overline{X}-\mathbb{E}(\overline{X})}{\sqrt{\mathrm{Var}(\overline{X})}}=\frac{\overline{X}-p}{\sqrt{p(1-p)/n}}=\frac{Y-np}{\sqrt{np(1-p)}}$$

is $\mathcal{N}(0,1)$. Now here is the key. The distribution of Y is actually known. It is Binomial with parameters n and p. What this means is that if n is reasonably large, then we have a way of approximating binomial probabilities.

We know that if $Y \sim \text{Binomial}(n, p)$, then

$$\mathbf{P}\{Y = k\} = \binom{n}{k} p^k (1-p)^{n-k}.$$

But if n is large, this can be impossible to compute. We saw that Stirling's formula is one way to approximate the probability. However, even that result does not help us when we need to add up a bunch of binomial probabilities. For example,

$$\mathbf{P}\{Y \leq k\} = \sum_{j=0}^{k} \mathbf{P}\{Y = j\} = \sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j}.$$

However, the central limit theorem can help. Notice that

$$\begin{aligned}
\mathbf{P}\{Y = k\} &= \mathbf{P}\{k - 1/2 < Y \leq k + 1/2\} \\
&= \mathbf{P}\left\{ \frac{k - 1/2 - np}{\sqrt{np(1-p)}} < \frac{Y - np}{\sqrt{np(1-p)}} \leq \frac{k + 1/2 - np}{\sqrt{np(1-p)}} \right\} \\
&= \mathbf{P}\left\{ \frac{Y - np}{\sqrt{np(1-p)}} \leq \frac{k + 1/2 - np}{\sqrt{np(1-p)}} \right\} - \mathbf{P}\left\{ \frac{Y - np}{\sqrt{np(1-p)}} \leq \frac{k - 1/2 - np}{\sqrt{np(1-p)}} \right\} \\
&\approx \Phi\left(\frac{k + 1/2 - np}{\sqrt{np(1-p)}} \right) - \Phi\left(\frac{k - 1/2 - np}{\sqrt{np(1-p)}} \right)
\end{aligned}$$

where Φ denotes the normal distribution function which can be evaluated using a table of normal probabilities. This is sometimes called the *normal approximation to the binomial distribution* and the idea of adding and subtracting $1/2$ is called the *continuity correction*.

Remark. When using the normal approximation to the binomial, first write down the exact binomial expression you want to evaluate. Next, add and/or subtract $1/2$ as appropriate at any endpoint with numbers. Then normalize by subtracting np and dividing by $\sqrt{np(1-p)}$. That will then give you an expression for an approximately normal random variable.

Remark. There is no agreed upon convention for "how large" n needs to be in order for this approximation to be "good." Instead of teaching students about proper error estimates, most books just ramble off a rule that says something like the following. "Use the normal approximation to the binomial when both $np > 5$ and $n(1-p) > 5$." Some books say instead: "Use the normal approximation to the binomial when both $np > 10$ and $n(1-p) > 10$."

Remark. When the normal approximation to the binomial does not apply, there is often a Poisson approximation that can be used. We'll discuss this later.

Example. It is well-known that commercial airlines oversell seats on their flights since, for various reasons, passengers routinely do not show up. Suppose that historically 3% of Air Canada passengers do not show up. If Air Canada sells 302 tickets for an airplane that has 300 seats, what is the probability that there will be a seat for every passenger who shows up? You may assume that each passenger behaves independently.