Math 302.102 Fall 2010
The Binomial Distribution

Suppose that we are considering repeated trials where in each trial we might observe one of only two possible results. We can label these results as *success* or *failure*. (Or, in applications, they might be on/off, zero/one, yes/no, left/right, male/female, shows improvement/does not show improvement, etc.) Assume further that the result of each trial is independent of the results of any other trial and that the probability of success in any given trial is p (so that the probability of failure on any given trial is $1-p$).

If we repeat the trials a total of n times and let the random variable X denote the *total number of successes* observed, then the probability that $X = k$ is computed as follows. In order for there to be k successes, then there must be $n-k$ failures. Any particular sequence of k successes and $n-k$ failures occurs with probability $p^k(1-p)^{n-k}$. Since we are interested in just the total number of successes k and not a particular ordering, we need to count the number of ways to arrange k successes among n trials. There are

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

ways to do this. Thus,

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

for $k = 0, 1, 2, \ldots, n$.

Example. What is the probability that we observe $k = 3$ successes in $n = 5$ trials? We are going to answer this question by writing out in detail in order to explain the formula above. Let S_j be the event that a success is observed on the jth trial so that $\mathbf{P}\{S_j\} = p$. Let $F_j = S_j^c$ be the event that a failure is observed on the jth trial so that $\mathbf{P}\{F_j\} = 1-p$. Hence,

$$\begin{aligned}
\mathbf{P}\{X = 3\} &= \mathbf{P}\{\text{exactly 3 successes in 5 trials}\} \\
&= \mathbf{P}\{S_1S_2S_3F_4F_5 \text{ or } S_1S_2F_3S_4F_5 \text{ or } S_1S_2F_3F_4S_5 \text{ or } S_1F_2S_3S_4F_5 \text{ or } S_1F_2S_3F_4S_5 \text{ or} \\
&\qquad S_1F_2F_3S_4S_5 \text{ or } F_1S_2S_3S_4F_5 \text{ or } F_1S_2S_3F_4S_5 \text{ or } F_1S_2F_3S_4S_5 \text{ or } F_1F_2S_3S_4S_5\} \\
&= \mathbf{P}\{S_1S_2S_3F_4F_5\} + \mathbf{P}\{S_1S_2F_3S_4F_5\} + \mathbf{P}\{S_1S_2F_3F_4S_5\} + \mathbf{P}\{S_1F_2S_3S_4F_5\} \\
&\qquad + \mathbf{P}\{S_1F_2S_3F_4S_5\} + \mathbf{P}\{S_1F_2F_3S_4S_5\} + \mathbf{P}\{F_1S_2S_3S_4F_5\} \\
&\qquad\qquad + \mathbf{P}\{F_1S_2S_3F_4S_5\} + \mathbf{P}\{F_1S_2F_3S_4S_5\} + \mathbf{P}\{F_1F_2S_3S_4S_5\} \\
&= \mathbf{P}\{S_1\}\mathbf{P}\{S_2\}\mathbf{P}\{S_3\}\mathbf{P}\{F_4\}\mathbf{P}\{F_5\} + \mathbf{P}\{S_1\}\mathbf{P}\{S_2\}\mathbf{P}\{F_3\}\mathbf{P}\{S_4\}\mathbf{P}\{F_5\} \\
&\quad + \mathbf{P}\{S_1\}\mathbf{P}\{S_2\}\mathbf{P}\{F_3\}\mathbf{P}\{F_4\}\mathbf{P}\{S_5\} + \mathbf{P}\{S_1\}\mathbf{P}\{F_2\}\mathbf{P}\{S_3\}\mathbf{P}\{S_4\}\mathbf{P}\{F_5\} \\
&\quad + \mathbf{P}\{S_1\}\mathbf{P}\{F_2\}\mathbf{P}\{S_3\}\mathbf{P}\{F_4\}\mathbf{P}\{S_5\} + \mathbf{P}\{S_1\}\mathbf{P}\{F_2\}\mathbf{P}\{F_3\}\mathbf{P}\{S_4\}\mathbf{P}\{S_5\} \\
&\quad + \mathbf{P}\{F_1\}\mathbf{P}\{S_2\}\mathbf{P}\{S_3\}\mathbf{P}\{S_4\}\mathbf{P}\{F_5\} + \mathbf{P}\{F_1\}\mathbf{P}\{S_2\}\mathbf{P}\{S_3\}\mathbf{P}\{F_4\}\mathbf{P}\{S_5\} \\
&\quad + \mathbf{P}\{F_1\}\mathbf{P}\{S_2\}\mathbf{P}\{F_3\}\mathbf{P}\{S_4\}\mathbf{P}\{S_5\} + \mathbf{P}\{F_1\}\mathbf{P}\{F_2\}\mathbf{P}\{S_3\}\mathbf{P}\{S_4\}\mathbf{P}\{S_5\} \\
&= p^3(1-p)^2 + p^3(1-p)^2 + \cdots + p^3(1-p)^2 \\
&= 10p^3(1-p)^2 \\
&= \binom{5}{3} p^3 (1-p)^2.
\end{aligned}$$

Example. What is the probability that we observe $k = 120$ successes in $n = 200$ trials? Notice that we can simply write down the answer. It is

$$\mathbf{P}\{X = 120\} = \binom{200}{120} p^{120}(1-p)^{200-120} = \frac{200!}{120!80!} p^{120}(1-p)^{80}.$$

However, if we tried to evaluate the coefficient

$$\binom{200}{120} = \frac{200!}{120!80!}$$

with a calculator (or even some computer programs), we are not able to do it because the numbers involved are simply too large! Fortunately, there is an approximation for $n!$ due to Stirling that sometimes helps. It states that

$$n! \approx \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n}$$

when n is *large*.

Problem 1. Suppose that a fair coin was tossed 20 times and that there were 12 heads observed. (You may assume that the results of subsequent tosses were independent.)

(a) What is the probability that the first toss showed heads?

(b) What is the probability that the first two tosses showed heads?

(c) What is the probability that at least two of the first five tosses landed heads?

Problem 2. Suppose that the lifetime X (in years) of a particular television model is exponentially distributed with parameter $\lambda = 1/2$ so that the density function of X is

$$f(x) = \begin{cases} \frac{1}{2}e^{-x/2}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Suppose that 57 televisions of this particular model are selected at random, and assume that television lifetimes are independent. Determine the probability that exactly 41 of the televisions last for less than one year (so that the other 16 last for at least one year).