Topic 1 – Probability Theory  
Module 1.3 – Discrete Random Variables  
Lecture 1.3.3 – Binomial Distribution

Suppose we have a process that can result in one of two outcomes – ‘success’ and ‘failure’, and that the probability of success is constant over repeated and independent trials. Suppose we repeat the process n times. Then the probability of exactly x success in the n trials is

\[ P(x/n \text{ trials}, p) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for } x = 0, 1, 2, \ldots, n \]

where \( \binom{n}{x} \) is read “n choose x” and is the number of ways we can take n things x at a time. The numerical value of \( \binom{n}{x} \) is calculated as \( \binom{n}{x} = \frac{n!}{x!(n-x)!} \), where k! = k(k-1)(k-2)⋯(3)(2)(1) and k! is pronounced “k factorial.”

Thus, 5! = (5)(4)(3)(2)(1) = 120, 3! = (3)(2)(1) = 6, and 2! = (2)(1) = 2, and therefore

\[ \binom{5}{3} = \frac{5!}{3!(5-3)!} = \frac{5!}{3!2!} = \frac{120}{6(2)} = 10. \]

Pascal is credited with a clever technique (known to the ancient Chinese) to generate these combinatorial numbers. The technique results in Pascal’s Triangle, as follows:

\[
\begin{array}{cccccccc}
1 \\
1 & 1 \\
1 & 2 & 1 \\
1 & 3 & 3 & 1 \\
1 & 4 & 6 & 4 & 1 \\
1 & 5 & 10 & 10 & 5 & 1 \\
\end{array}
\]

The borders are 1’s, and any other entry in the table is the sum of the numbers to the upper left and upper right of the number.

You know these numbers as the multipliers in the binomial theorem, which you first encountered as the statement \( (a+b)^2 = a^2 + 2ab + b^2 \). Note that this amounts to \( (a+b)^2 = 1a^2 + 2ab + 1b^2 \), and the multipliers 1, 2, 1 are the entries in the second row of Pascal’s triangle.

Similarly, we have

\[
(a+b)^3 = 1a^3 + 3a^2b + 3ab^2 + 1b^3, \\
(a+b)^4 = 1a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + 1b^4,
\]
and
\[(a + b)^5 = a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5.\]

Note that these multipliers are just the entries in the third, fourth, and fifth rows of Pascal’s triangle, respectively.

The ‘bi’ in binomial does not refer to the power 2 in \((a+b)^2\), but to the fact that there are only two possible outcomes, namely a and b. Thus, we can write the general binomial theorem as follows:

\[(a + b)^n = \sum_{i=0}^{n} \binom{n}{i} a^{n-i}b^i.\]

The binomial probability distribution has two parameters, as follows:

\[
\begin{align*}
E[x] &= np \\
\text{VAR}[x] &= np(1-p).
\end{align*}
\]

Binomial probabilities can be computed in Excel via the BINOMDIST function. See the text discussion on pages 88-89.

EXAMPLE
Suppose we have a binomial distribution with \(n = 10\) and \(p = .3\).
Then the population mean is \(np = (10)(.3) = 3\) and the population variance is \(np(1-p) = (10)(.3)(.7) = 2.1\).

EXAMPLE
A high school student who hadn’t opened his American history book in weeks was dismayed to walk into class and be greeted with a pop quiz. It was in the form of two lists, one naming the 24 Presidents in office during the 19th century in alphabetical order and another list noting their terms in office, but scrambled. The object was to match the Presidents with their terms. The completely clueless student had to guess every time. On average, how many did he guess correctly? (Source: Marilyn vos Savant, “Ask Marilyn,” \textit{Parade}, July 25, 2004.)

The answer is 1, and is obtained by noting that the students answers follow a binomial distribution with \(n = 24\) and \(p = 1/24\). Therefore, the expected number of correct answers is \(E[x] = np = 24(1/24) = 1\).

PROBLEMS:
Ch. 3, #18, 19