Binomial Distribution

1. The experiment consists of a sequence of $n$ smaller experiments called trials, where $n$ is fixed in advance of the experiment;
2. Each trial can result in one of the same two possible outcomes (dichotomous trials), which we denote by success ($S$) and failure ($F$);
3. The trials are independent, so that the outcome on any particular trial does not influence the outcome on any other trial;
4. The probability of success is constant from trial; we denote this probability by $p$.

Definition

An experiment for which Conditions 1 — 4 are satisfied is called a binomial experiment.
Binomial Distribution

Examples:
1. If we toss a coin 10 times, then this is a binomial experiment with $n = 10$, $S = $ Head, and $F = $ Tail.
2. If we draw a card from a deck of well-shuffled cards with replacement, do this 5 times and record whether the outcome is ♠ or not, then this is also a binomial experiment. In this case, $n = 5$, $S = $ ♠ and $F = $ not ♠.
3. Again we draw a card from a deck of well-shuffled cards but without replacement, do this 5 times and record whether the outcome is ♠ or not. However this time it is NO LONGER a binomial experiment.

$$P(♠\text{ on second} \mid ♠\text{ on first}) = \frac{12}{51} = 0.235 \neq 0.25 = P(♠\text{ on second})$$

We do not have independence here!
Binomial Distribution

Examples:
4. This time we draw a card from 100 decks of well-shuffled cards without replacement, do this 5 times and record whether the outcome is ♠ or not. Is it a binomial experiment?

\[
P(♠ \text{ on second draw} \mid ♠ \text{ on first draw}) = \frac{1299}{5199} = 0.2499 \approx 0.25
\]

\[
P(♠ \text{ on sixth draw} \mid ♠ \text{ on first five draw}) = \frac{1295}{5195} = 0.2492 \approx 0.25
\]

\[
P(♠ \text{ on tenth draw} \mid \text{not } ♠ \text{ on first nine draw}) = \frac{1300}{5191} = 0.2504 \approx 0.25
\]

... 

Although we still do not have independence, the conditional probabilities differ so slightly that we can regard these trials as independent with \(P(♠) = 0.25\).
Rule

Consider sampling without replacement from a dichotomous population of size $N$. If the sample size (number of trials) $n$ is at most 5% of the population size, the experiment can be analyzed as though it were exactly a binomial experiment.

e.g. for the previous example, the population size is $N = 5200$ and the sample size is $n = 5$. We have $\frac{n}{N} \approx 0.1\%$. So we can apply the above rule.
Binomial Distribution

Definition

The **binomial random variable** $X$ associated with a binomial experiment consisting of $n$ trials is defined as

$$X = \text{the number of S's among the } n \text{ trials}$$

Possible values for $X$ in an $n$-trial experiment are $x = 0, 1, 2, \ldots, n$.

Notation

We use $X \sim Bin(n, p)$ to indicate that $X$ is a binomial rv based on $n$ trials with success probability $p$.

We use $b(x; n, p)$ to denote the pmf of $X$, and $B(x; n, p)$ to denote the cdf of $X$, where

$$B(x; n, p) = P(X \leq x) = \sum_{y=0}^{x} b(y; n, p)$$
Example:
Assume we toss a coin 3 times and the probability for getting a head for each toss is $p$. Let $X$ be the binomial random variable associated with this experiment. We tabulate all the possible outcomes, corresponding $X$ values and probabilities in the following table:

<table>
<thead>
<tr>
<th>Outcome</th>
<th>$X$</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>HHH</td>
<td>3</td>
<td>$p^3$</td>
</tr>
<tr>
<td>HHT</td>
<td>2</td>
<td>$p^2 \cdot (1 - p)$</td>
</tr>
<tr>
<td>HTH</td>
<td>2</td>
<td>$p^2 \cdot (1 - p)$</td>
</tr>
<tr>
<td>HTT</td>
<td>1</td>
<td>$p \cdot (1 - p)^2$</td>
</tr>
<tr>
<td>TTT</td>
<td>0</td>
<td>$(1 - p)^3$</td>
</tr>
<tr>
<td>TTH</td>
<td>1</td>
<td>$(1 - p)^2 \cdot p$</td>
</tr>
<tr>
<td>THT</td>
<td>1</td>
<td>$(1 - p)^2 \cdot p$</td>
</tr>
<tr>
<td>THH</td>
<td>2</td>
<td>$(1 - p) \cdot p^2$</td>
</tr>
</tbody>
</table>

e.g. $b(2; 3, p) = P(HHT) + P(HTH) + P(THH) = 3p^2(1 - p)$. 

Liang Zhang (UofU)
Binomial Distribution

More generally, for the binomial pmf \( b(x; n, p) \), we have

\[
b(x; n, p) = \left\{ \begin{array}{c} \text{number of sequences of length } n \text{ consisting of } x \text{'s} \\ \text{probability of any particular such sequence} \end{array} \right\}.
\]

\[
\left\{ \begin{array}{c} \text{number of sequences of length } n \text{ consisting of } x \text{'s} \\ \text{probability of any particular such sequence} \end{array} \right\} = \binom{n}{x} \quad \text{and} \quad \binom{n}{x} \cdot p^x(1 - p)^{n-x}
\]

**Theorem**

\[
b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1 - p)^{n-x} & x = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}
\]
Example: (Problem 55)
Twenty percent of all telephones of a certain type are submitted for service while under warranty. Of these, 75% can be repaired, whereas the other 25% must be replaced with new units. If a company purchases ten of these telephones, what is the probability that exactly two will end up being replaced under warranty?
Let \( X \) = number of telephones which need replace, \( S \) = a telephone need repair and with replacement. Then
\[
p = P(\text{repair and replace}) = P(\text{replace} \mid \text{repair}) \cdot P(\text{repair}) = 0.25 \cdot 0.2 = 0.05
\]
Now,
\[
P(X = 2) = b(2; 10, 0.05) = \binom{10}{2} 0.05^2 (1 - 0.05)^{10-2} = 0.0746
\]
Binomial Distribution

Binomial Tables

Table A.1 Cumulative Binomial Probabilities (Page 664)

\[ B(x; n, p) = \sum_{y=0}^{x} b(x; n, p) \ldots \]

b. \( n = 10 \)

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.01</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>0</td>
<td>.904</td>
<td>.599</td>
<td>.349</td>
</tr>
<tr>
<td>1</td>
<td>.996</td>
<td>.914</td>
<td>.736</td>
</tr>
<tr>
<td>2</td>
<td>1.000</td>
<td>.988</td>
<td>.930</td>
</tr>
<tr>
<td>3</td>
<td>1.000</td>
<td>.999</td>
<td>.987</td>
</tr>
</tbody>
</table>

Then for \( b(2; 10, 0.05) \), we have

\[ b(2; 10, 0.05) = B(2; 10, 0.05) - B(1; 10, 0.05) = .988 - .914 = .074 \]
Binomial Distribution

Mean and Variance

Theorem

If $X \sim Bin(n, p)$, then $E(X) = np$, $V(X) = np(1 - p) = npq$, and
$
\sigma_X = \sqrt{npq} \text{ (where } q = 1 - p).$

The idea is that $X = nY$, where $Y$ is a Bernoulli random variable
with probability $p$ for one outcome, i.e.

$$
Y = \begin{cases} 
1, & \text{with probability } p \\
0, & \text{with probability } 1 - p 
\end{cases}
$$

$E(Y) = p$ and $V(Y) = (1 - p)^2p + (-p)^2(1 - p) = p(1 - p)$. Therefore $E(X) = np$ and $V(X) = np(1 - p) = npq$. 

Liang Zhang (UofU)  
Applied Statistics I  
June 19, 2008  
10 / 11
Example: (Problem 60)
A toll bridge charges $1.00 for passenger cars and $2.50 for other vehicles. Suppose that during daytime hours, 60% of all vehicles are passenger cars. If 25 vehicles cross the bridge during a particular daytime period, what is the resulting expected toll revenue? What is the variance?

Let $X =$ the number of passenger cars and $Y =$ revenue. Then

$$Y = 1.00X + 2.50(25 - X) = 62.5 - 1.50X.$$ 

$$E(Y) = E(62.5 - 1.5X) = 62.5 - 1.5E(X) = 62.5 - 1.5 \cdot (25 \cdot 0.6) = 40$$

$$V(Y) = V(62.5 - 1.5X) = (-1.5)^2V(X) = 2.25 \cdot (25 \cdot 0.6 \cdot 0.4) = 13.5$$