April 6, 2006

# Chapter 2. Discrete Models

The main purpose of this chapter is to introduce probabilistic ideas and terminology. We shall illustrate these ideas using discrete experiments where the total number of possible outcomes of the experiment is either finite or countably infinite. We say that a set of outcomes is countably infinite if the elements of the set can be arranged similarly to the counting numbers: $1, 2, 3, 4, \ldots$. For example, consider an experiment where note the number of radioactive particles emitted from a radioactive substance in an hour. Then the possible values for this recorded number are $1, 2, \ldots,$ with no upper bound. This is a discrete type experiment. On the other hand, if we measure the weight of a engine prototype, the measured weight can vary on a continuum of numbers, say from 300 to 350 pounds. A weight measurement is an example of a continuous variable. We discuss continuous models in the next chapter.

**2.1 Probability - The Foundation of Statistics.**

Every aspect of life entails uncertainty which yields variability and statistical models incorporate this uncertainty. In this chapter, we provide an introduction to probability theory which is a branch of mathematics providing a language describing uncertainty and allows for solving problems related to uncertainty. Just because the outcome of an experiment is uncertain does not mean we throw up our hands and give up. Usually, certain outcomes are more likely than others. Some outcomes are very rare. Probability helps us to quantify these uncertainties.

To begin, consider one of the most simple examples possible: flip a fair coin. The result of the flip is either heads (H) or tails (T), both with a 50-50 chance since the coin is fair. Another example of an experiment with two possible outcomes is buying a lottery ticket and noting whether or not you win. This example is similar to the coin flip example in that there are only two possible outcomes. However, we do not expect that the chance of winning is the same as the chance of losing. Another example that we are all familiar with is rolling a fair die, which is done in numerous board games so as to introduce the element of chance into the game. Here, we have 6 equally likely outcomes (instead of two with the coin).

The examples mentioned so far are fairly straightforward, but everyday life is full of examples where the probability computations can become quite complex. For instance, if you are playing poker, what is the probability of getting a full house?

To begin getting a handle on such problems, we need to introduce some terminology.

**Definition.** The *Sample Space $S$* is the set of all possible outcomes for a situation of interest.

**Definition.** An *Event* is a subset of the sample space, typically denoted by capital letters such as $A$ or $B$.

In the die tossing example $S = \{1, 2, 3, 4, 5, 6\}$ and we may be interested in an event such as $A$, the event an even number turns up: $A = \{2, 4, 6\}$. Our goal is to assign a number to events that tell us how likely the event is to occur – these numbers are called probabilities.

*Probabilities are defined as numbers between zero and one (inclusive) that indicate how likely an event is to occur. Probabilities near one indicate that the event is very likely to occur and probabilities near zero indicate that the even is unlikely to occur. A probability of* 0.5 *is a 50-50 chance.*

One way to assign a probability to an event $A \subset S$ is to consider performing the experiment over and over and look at the relative frequency at which the event $A$ occurs:

$$P(A) = \lim_{n\to\infty} \frac{\#\text{ of times } A \text{ occurs}}{n},$$

where $n$ is the number of times the experiment is performed. In ancient times, people realized that if a woman is going to have a baby, the probability that she will have a girl is 1/2 (i.e. $P(\text{girl}) = 1/2$) without knowing anything about x and y chromosomes. They realized that about half the children born are girls and the other half are boys. Of course, just because there are only two possible outcomes (e.g. boy or girl), does not mean they are equally likely. Consider another experiment where you need to run a pump for an experiment. When you turn the pump on, it either powers on or does nothing. Most of the time, say 95% of the time, it will power on and one can assign a probability of 0.95 to the event that the pump will power on. Because there are only two possibilities (powers on or does nothing), the probability the pump does not power on would be 0.05.

In other experiments, all the outcomes are equally likely. For instance, when tossing a fair die, the sample space $S$ contains 6 equally likely outcomes since the die is fair. The event $A$ of rolling an even number contains three of these outcomes, so $P(A) = 3/6 = 0.5$. If $B$ is the event you roll a multiple of 3, then $B = \{3, 6\}$ and $P(B) = 2/6 = 1/3$. Sometimes, probabilities can be computed by simply knowing the setup of the experiment. Suppose we receive a shipment of 100 engines and we know that 10 are defective. Consider an experiment of selecting one of the engines at random and let $A$ be the event that the selected engine is defective. Then it makes sense to assign a probability of 10/100 for the $P(A)$.

**FACT:** For any event $A$,

$$0 \le P(A) \le 1$$

and $P(S) = 1$, that is, the probability of the entire sample space is always one. The probability of the empty set is zero since the empty set contains no outcomes of the experiment.

## 2.2 Some Basics of Probability.

Given two events $A$ and $B$ in a sample space, we can form their *union*: $A \cup B$ which is the event that either $A$ **or** $B$ (or both) occur. The intersection of two events $A$ and $B$ is denoted by $A \cap B$ and is defined as the event that $A$ **and** $B$ occurs. Note that the key words are *and* for intersection and *or* for union. If you roll a fair die, and $A$ is the event that you roll an even number and $B$ is the event that you roll an odd number, then $A \cup B = S = \{1, 2, 3, 4, 5, 6\}$ and $A \cap B$ is the empty set. The probability of $A \cap B$ is zero because you cannot simultaneous roll a die and get both an even and an odd number. On the other hand, $P(A \cup B) = 1$ because the event you roll either an odd or an even number has to occur.

In this example, $B$ is known as the *complement* of $A$, denoted $\bar{A}$, which is defined as the event that $A$ does not occur.

Here are some rules for probability:

1. **The Law of Complements.** If $A$ is an event, then the **complement** of $A$, denoted by $\bar{A}$, is the event that $A$ does not occur. The probability of $\bar{A}$ is given by
$$P(\bar{A}) = 1 - P(A).$$

2. **The Additive Law of Probability.** Given two events $A$ and $B$,
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$
Extensions to the additive law exist for three or more events. For example,
$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

3. **The Law of Total Probability.**
$$P(A) = P(A \cap B) + P(A \cap \bar{B}).$$

**Definition.** Two events $A$ and $B$ are *mutually exclusive* if they cannot both simultaneously occur, i.e. their intersection is empty: $P(A \cap B) = 0$.

### 2.3 Conditional Probability.

A very important concept in probability is the concept of conditional probability. Often in practice the outcome of an experiment is uncertain, but we may have some additional information that helps shed some light on the outcome.

To illustrate, let $A$ be the event that a randomly selected cup of soup will be overfilled during the production process and suppose $P(A) = 0.1$. Further, suppose there are two filling tanks. Let $B$ be the event that the cup of soup was filled by the first tank. If both tanks fill an equal number of tanks, then it makes sense to set $P(B) = 0.5$. Suppose that $P(A \cap B) = 0.08$. That is, the probability that a randomly selected cup is overfull and was filled by the first tank is 0.08.

Consider the following question: Given that the cup was filled by the first tank, what is the probability it will be overfilled? That is

$$P(A \text{ given } B).$$

In probability, we call this the *conditional probability of A given B* and we write

$$P(A|B).$$

If we know the cup was filled by the first machine, then we know the event $B$ has occurred, and therefore, we can reduce our sample space from $S$ to $B$. The definition of conditional probability is then

**Conditional Probability Formula:** $P(A|B) = \dfrac{P(A \cap B)}{P(B)}.$

In the cup of soup example,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.08}{0.5} = 0.16.$$

That is, given that the cup was filled by the first tank, the probability it is overfilled is 0.16. We can turn the problem around: given that the cup is overfull, what is the probability that it was filled by the first tank? That is, find

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.08}{0.1} = 0.80.$$

We know that 50% of the cups are filled by the first tank. However, if we know that a cup has been overfilled, then there is an 80% chance it was filled by the first tank. Note that in general $P(A|B) \neq P(B|A)$.

### 2.4 Independence.

Another very important concept in statistics is that of independence. Note that in the previous example the event that a cup is overfilled is not *independent* of the event that the cup was filled by the first tank. If knowing that the event $B$ has occurred effects the likelihood of whether or not event $A$ occurs, then the two events are *dependent*. Conversely, if knowing that event $B$ has occurred does not effect the chance of event $A$ occurring, then we say the events $A$ and $B$ are **independent** :

> If $P(A|B) = P(A)$, then events $A$ and $B$ are *independent*. Otherwise they are *dependent*. It is easy to show that if $P(A|B) = P(A)$, then it follows that $P(B|A) = P(B)$.

A convenient way to think of independence mathematically is as follows: If events $A$ and $B$ are independent, then $P(A|B) = P(A)$. However, by the definition of conditional probability, $P(A) = P(A|B) = P(A \cap B)/P(B)$. Multiplying both sides by $P(B)$ gives

Events $A$ and $B$ are independent if and only if $P(A \cap B) = P(A)P(B)$.

The notion of independence allows us to solve lots of problems. Consider a jet that has two engines that operate independently. The probability of engine failure for an engine is 0.01. What is the probability that both engines fail? Let $F_1$ be the event the first engine fails and $F_2$ be the event the second engine fails. Then the probability of both engines failing is $P(F_1 \cap F_2)$. Since the engines operate independently,

$$P(F_1 \cap F_2) = P(F_1)P(F_2) = (0.01)(0.01) = 0.0001.$$

If the jet can fly as long as one of the engines is operating, what is the probability that the jet does not crash due to engine failure? That is, find $P(\bar{F}_1 \cup \bar{F}_2)$:

$$\begin{aligned} P(\bar{F}_1 \cup \bar{F}_2) &= P(\bar{F}_1) + P(\bar{F}_2) - P(\bar{F}_1 \cap \bar{F}_2) \ \text{(additive rule)} \\ &= (1 - P(F_1)) + (1 - P(F_2)) - P(\bar{F}_1)P(\bar{F}_2) \\ &= (1 - 0.01) + (1 - 0.01) - (1 - 0.01)^2 \\ &= 0.9999 \end{aligned}$$

The definition of independence extends to more than two events: events $A_1, A_2, \ldots, A_k$ are mutually independent if and only if for any subcollection of the events the probability of their intersection is equal to the product of their probabilities. In particular, it follows that $P(A_1 \cap A_2 \cap \cdots \cap A_k) = P(A_1)P(A_2)\cdots P(A_k)$.

**Example.** Suppose there is a $1/100 = 0.01$ chance you would win a weekly lottery if you buy a single ticket. If you play every week for a year (52 weeks), what is the probability you win at least once?

Let $W_j$ be the event you win on the $j$ week, $j = 1, 2, \ldots, 52$. Then $P(W_j) = 0.01$. Then

$$\begin{aligned} P(\text{Win at least once}) &= 1 - P(\text{Lose every week}) \\ &= 1 - P(\bar{W}_1 \cap \bar{W}_2 \cap \cdots \cap \bar{W}_{52}) \\ &= 1 - P(\bar{W}_1)P(\bar{W}_2)\cdots P(\bar{W}_{52}) \\ &= 1 - (1 - 0.01)^{52} \\ &= 1 - 0.59296645 \\ &= 0.40703355. \end{aligned}$$

So, there is about a 41% chance of winning at least once during the year.

### 2.5 Random Variables and Distributions.

Statistics is the science of data and data usually consists of numbers corresponding to recorded variables. When an experiment is conducted that produces sample points, we often assign numbers to these sample points via a *random variable*:

**Definition.** A *random variable* $Y$ is a function that assigns a number to each outcome in a sample space.

We can denote random variables by other letters besides $Y$ and typically use letters towards the end of the alphabet (e.g. $X$, $Y$, $Z$).

Some simple examples:

- Sample 10 engines from a large shipment and let $Y$ equal the number of defective engines out of the 10. Then the random variable $Y$ can assume possible values of $0, 1, \ldots, 10$. Thus $Y$ varies between these eleven values and the value that $Y$ assumes is random, depending on which engines you happen to choose by chance for inspection. This random variable is an example of a *discrete* random variable because it can only assume a finite number of values.
- In the cup-a-soup example, let $Y$ equal the weight of soup in a randomly chosen cup of soup. In this case, $Y$ is a *continuous* random variable because it can (theoretically) assume any value in a continuum of values. In such cases, it does not make sense to assign a non-zero probability to any specific value that $Y$ can assume because there are an uncountably infinite number of possible values. Note that for even though a random variable is continuous, we are only able to record their values on a discrete scale (say to the nearest 10th of a pound for example).
- Inspect cups of soup coming off the assembly line and let $X$ equal the number of cups inspected until one is overfilled. The possible values that $X$ can assume are $1, 2, 3, \ldots,$ with no upper limit. Even though there is no upper limit to the number of values $X$ can assume, $X$ is nonetheless a discrete random variable because the values $X$ can take can be put into a one-to-one correspondence with the natural numbers. Random variable that can take arbitrary values in a continuum cannot be put into a one-to-one correspondence with the natural numbers.

In each of these examples, its fairly easy to determine what values the random variable can assume. What we also need to know is how likely it is that the random variables assumes these values. That is, we need to know its *probability distribution.*

**Definition.** The *Cumulative Distribution Function* (CDF) of a random variable $Y$, denoted by $F(y)$ is defined by

$$F(y) = P(Y \leq y)$$

for all real numbers $y$.

We will see examples of CDF's in the next two sections.

### 2.6 Discrete Random Variables.

**Definition.** A *Discrete Random Variable* is a random variable that can assume at most a countably infinite number of values.

**Definition.** The *Probability Function* $p(y)$ for a discrete random variable is defined by

$$p(y) = P(Y = y).$$

Therefore, $0 \leq p(y) \leq 1$ since probabilities must lie between zero and one. Also, if we sum up all the values of $p(y)$ over all $y$ values, we must get one.

The CDF of a discrete random variable is

$$F(y) = \sum_{t \leq y} p(t).$$

**Example.** Roll a fair die and let $Y$ equal the face value that comes up. Then we can express the probability function of $Y$ conveniently in tabular form:

| $y$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $p(y)$ | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |
| $F(y)$ | 1/6 | 2/6 | 3/6 | 4/6 | 5/6 | 1 |

Note that in this example $F(y) = 0$ for $y < 1$ and $F(y) = 1$ for $y \geq 6$. Note also that the cdf is defined for all real numbers. For instance, in this example, $F(2.344) = P(Y \leq 2.344) = P(Y \leq 2) = 2/6$.

### 2.7 Expected Values for Discrete Distributions.

Suppose we perform an experiment and record a random variable. If we repeat the experiment over and over, what is the long-run average value of the random variable? This is called the *expected value* or the *mean* of the random variable and is denoted $E[Y]$ or by $\mu$. The expected value is defined by

$$\mu = E[Y] = \sum_y yp(y)$$

which is simply a weighted average of the values assumed by $Y$. Note that $\mu$ is considered a *parameter* of the distribution.

**Example.** Suppose that a randomly selected compressor can have either $0, 1, 2$ or 3 leaks in its seal. Let $Y$ equal the number of leaks and suppose $Y$ has the following probability function:

| $y$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $p(y)$ | 0.85 | 0.10 | 0.04 | 0.01 |

Note that the $p(y)$ values sum to one. The expected value of $Y$ is computed as

$$E[Y] = \sum_y yp(y) = 0(0.85) + 1(0.10) + 2(0.04) + 3(0.01) = .21.$$

Thus, if you sampled hundreds of compressors, you expect to see about 0.21 leaks on average.

We saw in the cup-a-soup example that the variation in the process was also very important. In probability we can formally define the variance of a random variable which is a measure of how "spread out" its values are. The variance is denoted by the Greek letter $\sigma^2$ ("sigma-squared"). In order to measure variability, a natural approach is to examine how far a measured variable $Y$ differs from the average value $\mu$. Now, $Y$ varies according to its probability distribution, so to get an over measure of variation, we can look at average deviations: $E(Y-\mu)$, but this quantity is always zero because the positive deviations from $\mu$ always cancel out the negative deviations from $\mu$. Instead, we compute the average *squared* deviations from $\mu$ and call this the *variance*:

**Definition.** The *population variance* of a discrete random variable $Y$ is defined as the average squared distance of $Y$ from its mean:

$$\sigma^2 = \text{var}(Y) = E[(Y-\mu)^2] = \sum_y (y-\mu)^2 p(y).$$

**Definition.** The *Standard Deviation* of a random variable is the (positive) square root of the variance:

$$\text{standard deviation} = \sigma = \sqrt{\sigma^2}.$$

Note that the standard deviation is in the same units as the original measurements.

Thus, $\sigma^2$ is another parameter of the distribution (as well as $\mu$).

The definition of the variance is the expected value of a function of $Y$, namely $(Y-\mu)^2$. More generally, we can compute the expected value of any function of a random variable $g(Y)$ using the following formula:

$$E[g(Y)] = \sum_y g(y)p(y).$$

In the case of the variance of $Y$, we simply compute the average value of $g(Y) = (Y-\mu)^2$.

It is also useful to know that the expectation operator is linear. That is,

$$E[a + bY] = a + bE[Y]$$

for any two constants $a$ and $b$. This allows us to provide a convenient formula for the variance of a random variable:

$$\begin{aligned}
\sigma^2 &= E[(Y-\mu)^2] \\
&= E[Y^2 - 2\mu Y + \mu^2] \\
&= E[Y^2] - 2\mu E[Y] + \mu^2 \\
&= E[Y^2] - 2\mu^2 + \mu^2 \\
&= E[Y^2] - \mu^2.
\end{aligned}$$

Therefore, a convenient formula for the variance of $Y$ is

$$\sigma^2 = \text{var}(Y) = E[Y^2] - \mu^2.$$

**Compressor Example continued.**

| $y$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $y^2$ | 0 | 1 | 4 | 9 |
| $p(y)$ | 0.85 | 0.10 | 0.04 | 0.01 |
| $yp(y)$ | 0 | 0.10 | 0.08 | 0.03 |
| $y^2p(y)$ | 0 | 0.10 | 0.16 | 0.09 |

From the table we compute that $E[Y^2] = 0 + 0.10 + 0.16 + 0.09 = 0.35$ and therefore the variance of $Y$ is $\sigma^2 = E[Y^2] - \mu^2 = 0.35 - (0.21)^2 = 0.3059$. The standard deviation $\sigma = \sqrt{\sigma^2} = \sqrt{0.3059} = 0.5531$.

Now that some of the basics of probability have been introduced, we present the most important discrete probability model – the binomial distribution.

## 2.8 The Binomial Distribution.

Consider any experiment that consists of repeated trials, like taking several foul shots in basketball. In each trial, the result is either a success (make a basket) or a failure (you miss the basket). The *binomial distribution* results when we count the number of *successes* (S) in $n$ independent and identical trials where the outcome of each trial is either a success (S) or a failure (F). We let $p$ denote the probability of a success (in which case the probability of a failure is $1 - p$, which we denote by $q$). If we let $Y$ equal the number of successes, then we say that $Y$ is a *binomial random variable.*

The simplest example of a binomial experiment consists of flipping a coin $n$ times and recording the number of tails. Suppose $n = 10$ and the coin is fair ($p = 0.5$). Then we will see how to compute probabilities such as $P(Y = 5)$, that is, what is the probability of getting exactly 5 tails in 10 flips?

Let us consider a more interesting example which we will use to derive the probability function for a binomial random variable and also illustrate some of the basic properties of *hypothesis testing.*

**Example.** Consider the cup-a-soup example. Suppose that a cup is ok for shipment if the weight dispensed into is between 237-239. The probability that a cup is within this specification is 0.80. A change is made to the production process to see if the proportion of cups that fall in this specified range is increased. In order to test if the change has improved matters, $n = 10$ cups are sampled. For each cup we record whether it is a success (S), (i.e., if its weight is between 237 - 239.), or a failure (F) (i.e., if its weight falls outside this range). Let $Y$ denote the number of good cups out of the $n = 10$ trials. Then $Y$ is a binomial random variable (assuming the 10 trials are independent and identical). The question of interest concerns the probability of success, $p$, under the changed process. In particular, is $p > 0.80$: did the change improve matters?

Let us assume for the sake of argument that $p$ has not changed and that $p = 0.80$ still. Let us compute the probability function $p(y)$ of $Y$. Since $Y$ is the number of success out of $n = 10$ trials, $Y$ can assume the values $0, 1, \ldots, 10$.

One step at a time, we shall compute $p(0) = P(Y = 0)$ first.

$$\begin{aligned} p(0) &= P(Y = 0) \\ &= P(FFFFFFFFFF) \\ &= P(F)P(F)P(F)P(F)P(F)P(F)P(F)P(F)P(F)P(F) \text{ by independence} \\ &= 0.20^{10} \\ &= 0.0000001. \end{aligned}$$

Thus, it is not very likely that all 10 cups would fail to meet specifications if $p = 0.80$. Next, consider $p(1) = P(Y = 1)$. One way this can happen is if we get an outcome such as $SFFFFFFFFF$. Again, by independence, the probability of this outcome is

$$\begin{aligned} (0.8)(0.2)(0.2)(0.2)(0.2)(0.2)(0.2)(0.2)(0.2)(0.2) &= (0.8)(0.2)^9 \\ &= p^y q^{n-y} \end{aligned}$$

where $y = 1$. However, the outcome $SFFFFFFFFF$ is just one of many ways in which $Y = 1$. Another possible outcome giving $Y = 1$ is $FSFFFFFFFF$. There are 10 possible ways to get exactly one success – we can put the single $S$ in any one of the 10 slots. Thus,

$$p(1) = 10(0.8)^1(0.2)^9 = 0.0000041.$$

How about $p(2)$? One possible outcome giving $Y = 2$ is $SSFFFFFFFF$. Another is $SFSFFFFFFF$ and there are many more. We need a clever way of counting all the possibilities. One way to think of the problem is as follows: label the two success $S_1$ and $S_2$. We have $n = 10$ slots and we need to choose two of them to place the successes into. One such possibility is:

$$\_\ \_\ \_\ \_\ \underline{S_2}\ \_\ \_\ \underline{S_1}\ \_\ \_.$$

We have $n = 10$ choices of slots to place $S_1$ into leaving $n - 1 = 9$ slots to place $S_2$ into. Thus the total number of possible ways of placing $S_1$ and $S_2$ into the 10 slots is $10 \cdot 9 = 90$. In order to derive a general formula, we will adopt the *factorial* notation:

$$n! = n(n-1)(n-2)\cdots 2 \cdot 1 \qquad (\text{“}n \text{ factorial”}).$$

(Recall that $0! = 1$.) Thus the total number of ways of placing $S_1$ and $S_2$ into the $n$ slots is given by

$$90 = 10 \cdot 9 = \frac{10!}{(10-2)!} = \frac{n!}{(n-y)!}$$

where $y = 2$. Note that we have artificially labeled the two successes as $S_1$ and $S_2$. The 90 possibilities we just counted distinguishes the order in which we placed the two successes. However, we are not interested in the order; the labeling was artificial. To get the correct number of possibilities we need to divide the 90 by 2 because there are two ways to rearrange $S_1$ and $S_2$ by simply having them change places. Thus, the total number of ways of choosing $y = 2$ slots out of the $n = 10$ possible slots to place successes in is

$$90/2 = 45 = \frac{10 \cdot 9}{2} = \frac{n!}{(n-y)!y!}.$$

The same logic can be applied when $y = 3$ successes. We can label the three successes $S_1, S_2$ and $S_3$. There are $720 = 10 \cdot 9 \cdot 8 = 10!/(10-3)!$ ways of arranging these three successes into the 10 slots. Once again, we are not interested in distinguishing the three successes. Thus, the 720 possibilities is too large by a factor of $3 \cdot 2 \cdot 1 = 6$ ways of rearranging the three successes. The total number of possibilities is then $n!/((n-y)!y!)$ which is the same formula we derived when $y = 2$. This formula is the general formula for all values of $y = 0, 1, \ldots, n$. This expression for counting the number of *combinations of* $n$ *objects taken* $y$ *at a time* is given by the *binomial coefficient* which is denoted by $\binom{n}{y}$:

$$\binom{n}{y} = \frac{n!}{y!(n-y)!} \tag{1}$$

The binomial coefficient $\binom{n}{y}$ counts the number of ways of choosing $y$ items from a collection of $n$ items.

Using the binomial coefficient, we can now give the formula for the binomial probability function on $n$ trials and success probability $p$: the probability of exactly $y$ successes out of $n$ trials is given by

$$p(y) = \binom{n}{y} p^y q^{n-y}, \; y = 0, 1, \ldots, n, \tag{2}$$

and zero otherwise. Continuing our previous computation,

$$p(2) = \binom{10}{2} (0.8)^2 (0.2)^{10-2} = 0.00007373.$$

Applying the formula for $y = 3, 4, \ldots, 10$, will give the remaining probabilities. Figure 1 shows the probability function for the binomial distribution. The left panel of Figure 1 shows the binomial probability function for $n = 10$ which is skewed to the left. The right panel of Figure 1 shows the binomial probability function for $n = 100$ which looks symmetric and bell-shaped.

The mean of a binomial random variable is

$$\mu = np \tag{3}$$

which can be found by computing $\sum_{y=1}^{n} y \binom{n}{y} p^y q^{n-y}$.

**Caution:** This formula for the mean only works for binomial random variables.

The formula is quite intuitive. Suppose you are an 80% free-throw shooter in basketball and you take $n = 10$ shots. How many would you expect to make? The answer

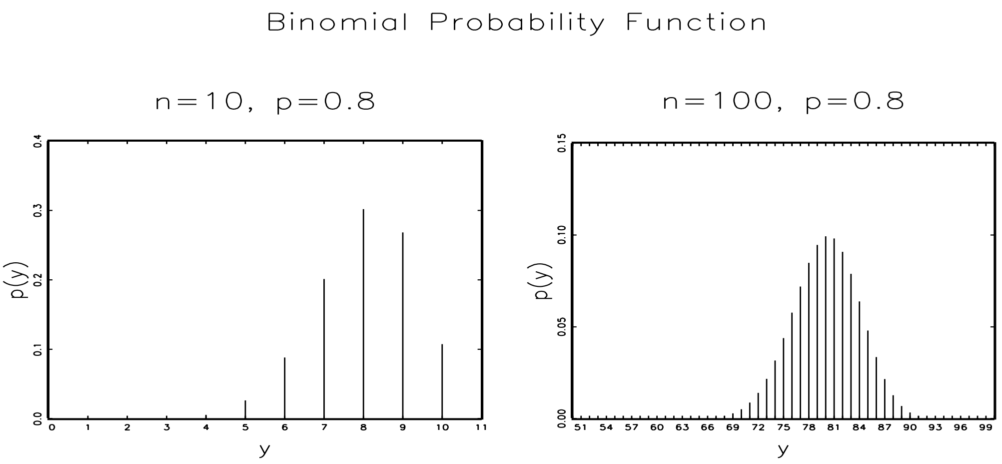


Figure 1: **Binomial Probability Function** Left Panel: $n = 10, p = 0.80$, Right Panel: $n = 100, p = 0.80$.

is 80% of 10, or 8. Here's an interesting question, if you are an 80% shooter, are you more likely to make 6 of 10 shots or make all 10 shots? Just plug the numbers into (2) to find the answer (most people's intuition is wrong on this one).

The variance of a binomial random variable is

$$\sigma^2 = npq \quad \text{(Binomial only)}. \tag{4}$$

The binomial coefficient is useful for counting the number of outcomes of experiments when the total number of outcomes is very large. Here are a couple common examples.

**Example.** When dealing $y = 5$ cards from an ordinary deck of $n = 52$ cards for a poker hand, there are

$$\binom{52}{5} = \frac{52!}{(52-5)!5!} = 2,598,960$$

possible poker hands. To compute the probability of a royal straight flush in poker (i.e. 10, Jack, Queen, King and Ace all of the same suit), note that there are only 4 possible royal straight flushes for the four different suits (hearts, diamonds, spades, clubs). Thus,

$$P(\text{Royal Straight Flush}) = \frac{\#\text{ of possible Royal Straight Flushes}}{\text{Total }\#\text{ of poker hands}} = \frac{4}{\binom{52}{5}} = \frac{4}{2,598,960} = .0000015390772.$$

**Example** (*Super Lotto*) Suppose you buy a super lottery ticket where you choose 6 numbers from the set of numbers $1, 2, \ldots, 47$. Then there are

$$\binom{47}{6} = 10,737,573$$

possible combinations. If you buy one ticket, your chances of winning are 1 out of $10,737,573$ ... not very good odds.

One of the main statistical inference procedures is hypothesis testing. The basic ideas of hypothesis testing are now introduced using the binomial distribution. The concepts covered here carry over to other statistical models.

**The R Software**. In the next section we illustrate hypothesis testing with a binomial example. The R statistical software (R Development Core Team 2003) will be used to do some of the tedious probability computations. The R software is freely available on the internet from the following web site:

`http://cran.r-project.org/`

From this web site you can download the executable (.exe) file onto your computer and double-click the executable icon to install R. The R-code in the notes below can be cut and pasted into the R window to obtain the results.

## 2.9 An Introduction to Hypothesis Testing.

We shall use the following example to introduce the concept of hypothesis testing using a binomial distribution. This example requires computing cumulative binomial probabilities. The R-software will be used to do the probability computations.

**Example:** 20% of the electrodes produced by a machine are defective and cannot be used resulting in a waste of time and money. The company is considering purchasing a new but expensive replacement machine in the hope that the proportion of defective electrodes will decrease. Before purchasing the machine, the company decides to test it first by producing $n = 100$ electrodes with the new machine. Based on the test run producing $n = 100$ electrodes, a decision needs to be made: buy the new machine or stick with the old machine. How should the decision be made?

The decision can be made using *hypothesis testing*. Out of the $n = 100$ electrodes produced by the new machine, let $Y$ denote the number of electrodes that are defective. From the previous sections of this chapter, we would expect $Y$ to follow a binomial distribution with $n = 100$ trials and success probability $p$. The success probability $p$ in this problem is an example of a *parameter* and the problem is that we do not know the value of $p$. If the new machine is no better than the old machine, then $p \geq 0.20$ and there is no sense in buying the expensive new machine. If, on the other hand, $p < 0.20$, then the defect rate for the new machine is less than that of the old machine and it may make sense to replace the old machine by the new machine. Suppose the defect rate for the new machine is the same as the old machine (i.e. $p = 0.20$). Then we would expect the number of defective electrodes (out of $n = 100$) to be around $\mu = np = 100(0.20) = 20$ plus or minus a standard deviation or two. However, if the number of defective electrodes is considerably less than 20, then we would conclude the new machine is better than the old machine.

The logic behind hypothesis testing is as follows. We assume for the sake of argument that the new machine is no better than the old machine (the status quo) and we call this the **null hypothesis** and denote it by $H_0$. In terms of the defect rate parameter $p$ for the new machine, the null hypothesis $H_0$ can be written

$$H_0 : p = 0.20.$$

The null hypothesis is always stated in terms of a model parameter, in this case $p$, the defect rate of the new machine. We also set up an **alternative hypothesis** also in terms of the model parameter, denoted $H_a$, which states the research hypothesis: is the new machine better than the old machine? In terms of the defect rate $p$ for the new machine, the alternative hypothesis $H_a$ is

$$H_a : p < 0.20.$$

The idea now is to run the experiment (i.e. produce $n = 100$ electrodes with the new machine) and see if the data from the experiment allow us to reject the null hypothesis $H_0$ and accept the alternative hypothesis $H_a$ that the new machine is better than the old machine.

In order to make the decision based on the data, we plug the data into a **test statistic**. Test statistics can be quite complicated in practice, but for this example we shall use a very simple test statistic: let $Y =$ the number of defective electrodes. We shall let $Y$ be the test statistic.

If the number of defective electrodes $Y$ is small, we will reject the null hypothesis $H_0$ and accept the alternative hypothesis $H_a$ that the new machine has a lower defect rate. The question is: how small does $Y$, the number of defective electrodes, have to be in order to reject $H_0$ and conclude the new machine is better than the old machine? In order to make this decision, we need a cut-off value for $Y$ so that if $Y$ is less than this cut-off value we reject the null hypothesis. Whenever we make a decision there are two types of errors possible (described below). The cut-off value is determined by minimizing the chance of committing one of these errors. Here are the definitions for the two types of errors when making a decision:

**Definition.** A *Type I* error occurs if the null hypothesis is rejected when it is true.

**Definition.** A *Type II* error occurs if the null hypothesis is accepted when it is false.

In the context of the electrode example, a type I error occurs if we conclude the new machine works better than the old machine (reject $H_0 : p = 0.20$ and conclude $H_a : p < 0.20$) when in fact the new machine is no better than the old machine.

A type I error here would be very bad because an expensive new machine will be purchased that is no better than the old machine. A type II error would be to claim the new machine performs the same as the old machine (accept $H_0$) when in fact the new machine has a smaller defect rate. A type II error in this context is also bad, but committing it means the company would just continue producing electrodes with the old machine. Often hypothesis tests are set up in such a way that a type I error is the more serious error.

To help understand the logic behind hypothesis testing, consider an analogy with a courtroom trial. The defendant on trial is either guilty or not guilty. Evidence is heard to decide whether to convict or not convict the defendant. To begin, the defendant is assumed to be innocent and then the data is examined to determine if we can "reject the hypothesis" of innocence and convict. Thus, we can set this up as a hypothesis test:

**Null Hypothesis** $H_0$ : Innocent

versus the

**Alternative Hypothesis** $H_a$ : Guilty.

In statistics, the evidence is in the data and we use the data to determine if the null hypothesis should be rejected or not. In a court trial there are two possible decisions (convict or not convict) and also two possible errors: type I and type II. In the trial analogy, a type I error is to reject the assumption of innocence and convict the defendant when in fact the defendant is innocent. Convicting an innocent person is considered a very bad thing, and thus we generally need to be convinced beyond a reasonable doubt that the defendant is guilty. A type II error in the context of a court case is to let a guilty person go free.

Note that in a court of law, failing to convict the defendant does not necessarily mean that the defendant is innocent. Failing to convict the defendant could mean that there was not enough evidence. In the statistical framework, failing to reject the null hypothesis could result either because the null hypothesis is true *or* because there is not enough data (i.e. evidence) to conclude the null hypothesis is false. Therefore, in practice if the null hypothesis is not rejected, one will typically refrain from claiming the null hypothesis is true because this could cause a type II error. Instead, one can say there is insufficient evidence to reject the null hypothesis.

The type II error problem highlights the importance of designing experiments appropriately. We want to avoid conducting a costly experiment or survey where we collect evidence (data) and find out afterwards we cannot reject the null hypothesis simply due to a lack of evidence. Lack of evidence could be due to insufficient sample size or a poor experimental design or sampling design. Great care must go into the data collection process.

Recall that in the electrode example, we need to determine a cut-off value for $Y$ in order to make a decision. This cut-off value will be chosen to make the probability of a type I error small since a type I error is considered more serious than a type II error. The probability of committing a type I error is called the *significance level* and denote it by the Greek letter $\alpha$ ("alpha").

**Definition.** The *significance level* of a test, denoted by $\alpha$, is the probability of committing a type I error, i.e. rejecting $H_0$ when it is true.

Typical values for the significance level $\alpha$ are 0.01, 0.05, or 0.10 depending on how much protection one wants against committing a type I error. The value $\alpha = 0.05$ is used most frequently. Because the binomial distribution is discrete, it is usually not possible to set the significance level at exactly some fixed value like $\alpha = 0.05$ as we shall see. Let $c$ be the cut-off value for $Y$ so that we will reject $H_0$ if we observe a value $Y = y \leq c$. Let us choose $c$ so that the significance level is $\alpha = 0.05$ (or as close as possible to 0.05).

**Definition:** The **critical region** (or rejection region) of a test is the set of values of the test statistic that will lead to a decision to reject the null hypothesis.

In the electrode example, the critical region will be of the form $y \leq c$ where $y$ is the observed number of defective electrodes. If we choose $c = 13$, then

$$\begin{aligned}
\alpha &= P(\text{type I error}) \\
&= P(\text{Rejecting } H_0 \text{ when } H_0 \text{ is true}) \\
&= P(Y \leq c \text{ when } p = 0.20) \\
&= P(Y \leq 13 \text{ when } p = 0.20) \\
&= 0.0469.
\end{aligned}$$

This probability can be found using the R-software by typing:

```
n=100
p=0.20
sum(dbinom(0:13,n,p))
```

The "dbinom" function in R gives binomial probabilities based on $n$ trials and success probability $p$. The "0:13" asks R to compute the probabilities for number of successes from 0 to 13 and the "sum" tells R to add up all these probabilities which will give $P(Y \leq 13)$.

From the above probability computation we see that if the number of defective electrodes out of $n = 100$ produced by the new machine is less than or equal to 13, then we will reject $H_0$ conclude that the defect rate $p$ of the new machine is less than 0.20. The probability of making a type I error in this case is only 0.0469. Stated another way, if the defect rate $p$ for the new machine is the same as the old machine ($p = 0.20$) then observing 13 or fewer defects with the new machine is very unlikely. Figure 2 shows a picture of the binomial distribution when $p = 0.20$ along with the critical region. Note that the probabilities $p(y)$ in this figure are essentially zero once you get more than three standard deviations away from the mean of $\mu = 20$.

Suppose the test run with the new machine is run and out of the $n = 100$ electrodes produced, we observe $y = 10$ defective electrodes. Since the value $y = 10$ falls in the critical region ($y = 10 \leq 13$), we would reject $H_0$ and conclude that the defect

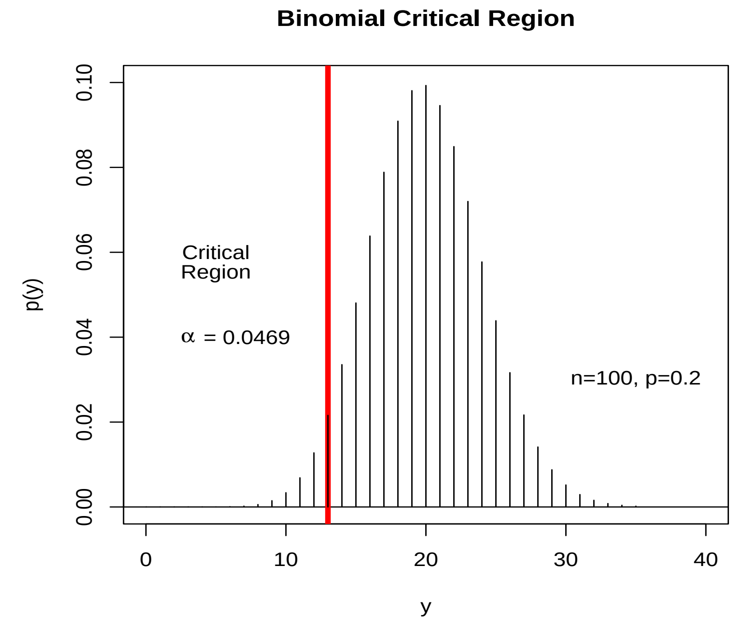


Figure 2: **Null Distribution** for the binomial distribution with $n = 100, p = 0.20$ with cut-off for the critical region.

rate $p$ for the new machine is less than 0.20 with a significance level $\alpha \approx 0.05$. It is important to state the significance level $\alpha$ in your conclusion because this specifies the strength of the statistical evidence against the null hypothesis. In this example we are claiming that the new machine is better than the old machine. This could be an incorrect claim (i.e. a type I error) but the probability of making that error is only $\alpha \approx 0.05$.

In the electrode example, the hypothesis test was an example of a *one-sided* test. That is, we decided to reject $H_0$ for only small values of $Y$. In other examples where one wants to determine if the parameter *differs* from some hypothetical value, then we would have a *two-sided* test where we would reject the null hypothesis for either very large or very small values of the test statistic.

## 2.10 $p$-values.

In the previous section where hypothesis testing was described, a small probability of a type I error ($\alpha = 0.05$) was specified which determined the cut-off value for the critical region. Another common approach to testing a hypothesis is to report the strength of the evidence against $H_0$. In the electrode example, observing $y = 10$ defective electrodes with the new machine would lead to the rejection of the null hypothesis using a significance level $\alpha \approx 0.05$ because $y = 10$ is in the critical region. In this section, we ask: *How likely is it to observe 10 or fewer defects with the new machine if the defect rate is the same as the old machine's defect rate?* This probability is known as a $p$-value. Formally, for this example, the $p$-value is computed as:

$$p\text{-value} \quad = \quad P(Y \leq 10) \text{ (assuming } p = 0.20)$$

$$= \quad 0.0057.$$

This probability was found using the R-software by typing

```
n=100
p=0.20
sum(dbinom(0:10,n,p))
```

If the defect rate for the new machine is the same as the old machine ($p = 0.20$), then the probability of observing 10 or fewer defective electrodes out of $n = 100$ is extremely unlikely (the probability is 0.0057). Reporting this $p$-value is more informative than performing a test at a fixed significance level $\alpha$ because the $p$-value tells you exactly the strength of the evidence against $H_0$.

Here is a general definition of a $p$-value:

**Definition.** The ***p*-value** of a statistical test is the probability of observing an outcome as extreme or more extreme (away from $H_0$) than what was actually observed when the null hypothesis is true.

Because $p$-values are probabilities, they range in value between 0 and 1. $p$-values near zero are evidence against the null hypothesis. For instance, in the electrode example above, the $p$-value was 0.0057 is very small and provides strong evidence against $H_0$. Small $p$-values tell us that an observed outcome is very unlikely if the null hypothesis is true. A rough rule of thumb is that if the $p$-value is less than 0.01, one has very strong evidence against $H_0$. If $p$-value $< 0.05$, then one strong evidence against $H_0$. If $0.05 < p - \text{value} < 0.10$, then the evidence against $H_0$ is only moderate. Generally, $p$-values $> 0.10$ are not considered as evidence against $H_0$. Of course, there is some grey area in interpreting $p$-values.

Recall that there are two types of errors in hypothesis testing: type I and type II. In the context of the electrode problem, a type I error is to conclude that the defect rate for the new machine is lower than that of the old machine when in fact it is not lower. A type II error is claim the defect rate for the new machine is the same as the old machine when in fact the new machine has a lower defect rate. As mentioned above, it is important to plan experiments and surveys so that you have enough data (evidence) to reject the null hypothesis when the null hypothesis is false. In statistical terminology, one wants to plan experiments so that the hypothesis test has high *power*.

**Definition.** The *Power* of a statistical test is the probability of rejecting the null hypothesis when the null hypothesis is false.

In the courtroom analogy, low power is similar to little evidence. A guilty defendant may not be convicted if there is a lack of evidence. In statistics, a false null hypothesis will not be rejected if there is not enough data. Power computations tend to be a little complicated and we will not provide one here. However, we can illustrate the problem with poor power using the electrode example again. Suppose in the electrode example a test run with the new machine was run that produced only $n = 10$ electrodes instead of $n = 100$. If only $y = 1$ defective electrode is observed from the $n = 10$ test run, then the proportion of defections is 1/10 or 10% which is the same proportion in the above example (10 out of 100 or 10%). In the large test run ($n = 100$), observing ten defective electrodes provided very strong evidence against the null hypothesis $H_0$. However, if the smaller test run is made ($n = 10$), the $p$-value of the test is $P(Y \leq 1) = 0.3758$ which is not a small probability. In other words, if the null hypothesis is true (i.e. the defect rate is $p = 0.20$), then it is not unusual that the number of defective electrodes produced out of ten is less than or equal to one. With such a large $p$-value, we cannot conclude the new machine is better than the old machine (i.e. we cannot reject $H_0$). The new machine may indeed be better than the old machine, but we cannot make that determination based on a test run of only $n = 10$ electrodes.

When designing an experiment or survey an important consideration then is that your test will have adequate power to detect differences from the null hypothesis. Required sample sizes needed for an experiment are determined by specifying ahead of time the desired power. For instance, requiring a power of 90% is quite common. Higher power requires a greater sample size. There are many software packages available for doing sample size computations. For more complicated models, computer simulations may be needed to determine an adequate sample size to guarantee a high power.

**Two-Tailed Test**

In the electrode example, we rejected the null hypothesis if the number of defective electrodes $Y$ produced by the new machine was small. That is we rejected $H_0$ for $Y \leq c$, where $c$ is a designated cut-off value. In many applications we may set up a hypothesis test to reject a null hypothesis if the test statistic is either too large or too small: in these cases, the test is known as a *two-tailed test.* The following example will help illustrate a two-tailed test.

**Example.** 30% of air tanks begin to leak when the pressure in the tank exceeds a specific threshold. The company manufacturing the tanks begins using a new valve produced by a different supplier. Fifty tanks are tested with the new valve to determine if the proportion of tanks that leak has changed. Let $p$ denote the proportion of tanks that will leak with the new valve when the pressure exceeds a the specific threshold. The null hypothesis of the test is

$$H_0 : p = 0.30,$$

which says that the proportion of tanks that leak with the new valve is the same as with the old valve. We want to determine if the proportion of tanks that leak with the new valve has changed, so the alternative hypothesis is

$$H_a : p \neq 0.30.$$

If the observed proportion of tanks out of the $n = 50$ tested is either much bigger or much smaller than 0.30, then we will reject $H_0$ and accept $H_a$. This is an example of a *two-tailed* test because we will reject $H_0$ if the observed number of leaking tanks falls in either the left or right tail of the binomial distribution. Let $Y$ denote the number of leaking tanks observed from the experiment. The critical region now takes the form: reject $H_0$ if $Y < c_1$ or $Y > c_2$. The question again comes down to finding cut-off values $c_1$ and $c_2$ in order to make a decision to reject $H_0$ or not.

Let us choose a significance level $\alpha = 0.05$. Because we have a two-tailed alternative, we can to split the 0.05 probability in two for the two tails of the binomial distribution: 0.025 for the left tail (small values of $Y$) and 0.025 for the right tail (large values of $Y$). Because $n$ is fairly large and $p = 0.3$ is not too close to zero or one, the binomial distribution for $n = 50$ and $p = 0.3$ will be fairly symmetric and we can use the empirical rule to get a rough idea of the cut-off values for the critical region. If $H_0$ is true, from (3), the mean number of tanks with leaks will be $np = 50(0.3) = 15$ with standard deviation $\sigma = \sqrt{np(1-p)} = 3.24$. (which follows from (4)). Approximately 95% of the probability will lie between $\mu \pm 2\sigma = 15 \pm 6.48$ which gives values of 8.52 and 21.48. Let $y$ denote the observed number of leaking tanks out of fifty. Let us choose cut-off values for our two-tailed critical region as:

$$\text{Reject } H_0 \text{ if } y \leq 8 \text{ or } y \geq 22.$$

The exact significance level for this test can be computed in R using the following command:

```
sum(dbinom(0:8,n,p)) + sum(dbinom(22:50,n,p))
```

which gives a value of

$$\alpha = 0.0433.$$

Thus, if the observed number of leaking tanks is less than or equal to 8 or greater than or equal to 22, then we will reject $H_0$ and conclude that the proportion of leaking tanks with the new valve has changed and is no longer 0.30. The probability of making a type I error (i.e. claiming the proportion of leaking tanks differs from 0.3 when in fact it does not) is only 0.0433. That is, with this test, it is not likely that a type I error will be made. A picture of the two-tailed critical region for this example is shown in Figure 3.

## 2.11 Some Other Discrete Distributions

In this final section, we introduce a few other well-known discrete probability distributions.

### Poisson Distribution

A binomial random variable is a discrete random variable that can assume a finite number of values, namely $0, 1, 2, \ldots, n$. Another type of discrete random variable that can take the values $0, 1, 2, \ldots,$ is the Poisson distribution. Consider an engineer who's job is to troubleshoot problems for customers that have purchased the company's product. Let the random variable $Y$ denote the number of calls that arrive per hour.

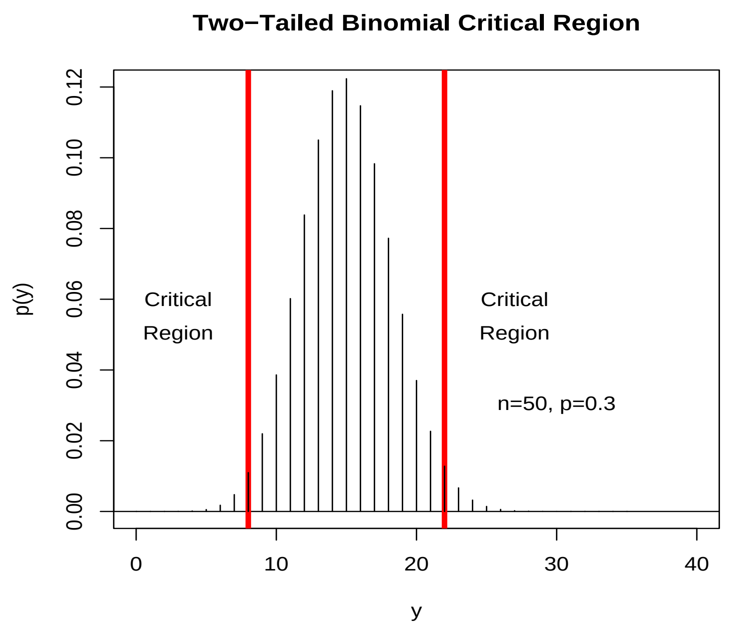


Figure 3: **Two-tailed Critical Region** for the binomial distribution with $n = 50, p = 0.30$.

A Poisson distribution often provides a reasonable model for data generated by such a process. The Poisson distribution is parameterized by a rate parameter $\lambda > 0$ and the probability function for the Poisson distribution is

$$p(y) = e^{-\lambda}\lambda^y/y!, \text{ for } y = 0, 1, 2, \ldots, \qquad (5)$$

and zero otherwise. The expected value of a Poisson random variable is $\lambda$. The Poisson distribution has an interesting property where the variance is equal to the expected value, i.e. $\text{var}(Y) = \lambda$.

The Poisson distribution is quite useful in practice for a couple of reasons. One reason is that the Poisson distribution provides a good approximation to the binomial distribution when the number of trials $n$ is large and the success probability $p$ is small. In such cases, the binomial distribution is well approximated by a Poisson distribution with mean $\lambda = np$.

**Example.** Suppose a typesetting company observes that the probability there is a typographical error on a given page is $p = 0.01$. In a manuscript of $n = 200$ pages, let $Y$ equal the number of pages with typographical errors. Assuming errors occur independently from page to page, $Y$ has a binomial distribution. The probability that there are no typos in the manuscript is

$$P(Y = 0) = \binom{200}{0}(0.01)^0(0.99)^{200} = 0.13397967.$$

Now, if we use a Poisson approximation with rate parameter $\lambda = np = 200(0.01) = 2$, then we can approximate the probability using (5) to get

$$e^{-2}2^0/0! = 0.13533528,$$

which is very close to the exact value.

Another reason the Poisson distribution arises is due to the *Poisson Process*. Consider a physical process where a particular type of event occurs (such as a defect in a product or the emission of a radioactive particle). Let $Y(t)$ denote the number of such events that occur in a given interval of time $[0, t]$. In many such processes, the probability an event occurs in a short interval of time is proportional to the size of the time interval and the occurrences of events in disjoint time intervals are independent. If the probability of two or more events occurring in a small interval of time is very small. Processes that satisfy these conditions are called (homogeneous) Poisson Processes. One can show that if $Y(t)$ is the number of occurrences of the event in the interval $[0, t]$, then $P(Y(t) = k) = e^{-\lambda t}(\lambda t)^k/k!$, for $k = 0, 1, 2, \ldots,$. That is, $Y(t)$ has a Poisson distribution.

There are several other well-known discrete probability distributions that are very useful in practice and we briefly note a few of them here:

**Hypergeometric Distribution.** The hypergeometric distribution is very similar to the binomial, the difference being that the hypergeometric distribution is appropriate when the population is finite. Suppose you conduct an opinion poll by randomly sampling $n = 1000$ from a population of $N = 1,000,000$. Let $Y$ equal the number of respondents that answer yes to your poll question. If we had *sampled with replacement* (i.e. put a million names in a hat, pick one, record the outcome, put the name back in the hat, mix it up and pick again), then $Y$ would have a binomial distribution. Of course, it would be silly to sample with replacement. Opinion polls *sample without replacement* so that the same person will not be polled twice (or more!). Since the population is finite ($N = 1,000,000$), $Y$ turns out to have a hypergeometric distribution with probability function

$$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$$

for $y = 1, 2, \ldots, n$; subject to $y \leq r$ and $n - y \leq N - r$. Here, $r$ equals the number of "successes" in the population of size $N$. The difference between the hypergeometric and the binomial distributions in the opinion poll example is negligible. However, when the population size is small then the differences between the two distributions can be substantial and the hypergeometric distribution should be used.

**Example.** Suppose that in a shipment of $N = 100$ computers, $r = 10$ have defective hard drives. If you randomly select 5 computers, what is the probability that $y = 3$ of them will be defective?

This is an example of a hypergeometric distribution problem. Let

$$A = \text{the event that 3 of the 5 selected computers is defective.}$$

Then

$$P(A) = \frac{\text{\# of ways A can occur}}{\text{Total \# of outcomes}}.$$

Since we are selecting $n = 5$ computers from a set of $N = 100$ computers, the denominator is $\binom{100}{5}$. As for the numerator, if we select three defective computers, they were selected from the $r = 10$ defective computers in the shipment and the number of ways that can occur is $\binom{10}{3}$. However, we are not done yet – if we selected $y = 3$ defective computers, then we must of selected $n - y = 5 - 3 = 2$ non-defective computers from the $N - r = 100 - 10 = 90$ non-defective computers in the shipment. The number of ways that can occur is $\binom{90}{2}$. Thus,

$$P(A) = \frac{\binom{10}{3}\binom{90}{2}}{\binom{100}{5}} = 0.0063835281.$$

**Geometric Distribution.** Suppose you monitor a production process until you find a defective item. If we let $Y$ denote the number of items monitored until a defective is found, then $Y$ has a geometric distribution, assuming the trials are independent and the probability an item is defective does not change throughout the process. The probability function for the geometric distribution is

$$p(y) = (1 - p)^{y-1}p \quad \text{for } y = 1, 2, \ldots.$$

*Question: Can you derive this probability function based on the description given above (see Problem 4(d)?*

## Problems

1. A company has two pumps, either of which can be used to pump water. The probability the older pump malfunctions is 0.5 and the probability that the newer pump malfunctions is 0.3.

   a) What is the probability that both pumps fail?

   b) What is the probability that at least one of the pumps does not malfunction?

   c) What assumption is necessary about how the two pumps work in order to answer parts (a) and (b)?

2. A gear box is selected at random from a collection of gear boxes that were manufactured over the last week at a factory. The factory operates with three shifts (day, early evening, late night). Let $A$ be the event the gear box was manufactured during the day shift, let $B$ be the event it was manufactured during the early evening shift and let $C$ denote the event that it was manufactured during the late night shift. Suppose $P(A) = 0.4$ and $P(B) = 0.3$. Find the following:

   a) $P(C)$

   b) $P(A \cap B)$. What is the term used to describe the relation between events $A$ and $B$?

   c) $P(A \cup B)$.

   d) $P(A|B)$.

   e) Are events $A$ and $B$ independent?

3. A company purchases parts for a product. 80% of the parts are from a Japanese company and 20% of the parts are from a German company. 5% of the Japanese parts are defective and 3% of the German parts are defective. A part is selected at random. Let $D$ be the event the part is defective, let $G$ be the event the part is from the German company and let $J$ be the event the part is from the Japanese company. Find the following:

   a) $P(D|G)$ and $P(D|J)$.

   b) $P(D \cap G)$ and $P(D \cap J)$

   c) $P(D)$

   d) $P(G|D)$. In plain English, what does this probability tell us?

4. The number of defects $Y$ in a paint job on newly manufactured cars has the following distribution:

| $y$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $f(y)$ | .6 | .3 | .07 | ? |

a) What is the probability that a car will have 3 defects?

b) What is the probability that a car will have less than two defects?

c) What is the average number of defects?

d) Consider an experiment where cars are observed coming off the production line until a car with a paint defect is encountered. Let $X$ equal the number of cars observed until a defect is found. What is the probability $P(X = 1)$, $P(X = 2)$, $P(X = 3)$, and $P(X = k)$ for an arbitrary value $k = 1, 2, 3, \ldots$? What is the name given to the probability distribution for $X$?

5. The fiberglass side of an aircraft has two flaws of sizes 1.1 inches and 1.7 inches in diameter. The probability of detecting the flaws using non-destructive inspection is 0.3 for the 1.1 inch flaw and 0.4 for the 1.7 inch flaw. An inspector inspects the side of the aircraft. Let $Y$ denote the number of flaws found. Assume the event of detecting one of the flaws is independent of whether or not the other flaw is detected.

   a) Find the probability distribution for $Y$.

   b) Find $P(Y > 0)$

   c) Find $P(Y \geq 0)$

   d) What is the expected value of $Y$?

   e) Suppose it costs \$200 to fix each of the flaws that are found. What is the expected cost?

6. Let $Y$ denote a binomial random variable on $n = 5$ trials with success probability $p = 0.2$. Find

   a) $P(Y = 4)$

   b) $P(Y \geq 4)$

   c) Use part (b) to compute $P(Y \leq 3)$.

   d) $E[Y]$

   e) the standard deviation $\sigma$ of $Y$.

7. A basketball player is an 80% freethrow shooter. If she takes $n = 10$ shots, what is more likely: making all 10 shots or making 6 of the 10 shots? Assume that the shots are independently of each other.

8. A company manufactures smooth-top ovens at two plants, a small plant and a large plant. In a given week, the small plant manufactures 100 ovens and the large plant manufactures 1000 ovens. The probability that a given oven will have an electrical system failure at some point during its lifetime is $p = 0.5$. Let $Y_1$ and $Y_2$ denote the number of ovens produced at the small and large plant respectively in a given week that will eventually develop electrical problems. What is more likely – more than 60 of the ovens at the small plant will eventually

have electrical problems or more than 600 ovens at the large plant will eventually have problems? It may seem at first glance that both events are equally likely. Do the following parts to try and answer the question.

a) Find $E[Y_1]$ and $E[Y_2]$, that is, find the expected number of ovens with eventual electrical problems at each plant.

b) Find $\sigma_1$ and $\sigma_2$, the standard deviations of $Y_1$ and $Y_2$ at each plant.

c) To answer the question, we could compute $P(Y_1 \geq 60)$ and $P(Y_2 \geq 600)$. However, a direct computation of these probabilities is tedious (for example, using (2), $P(Y_2 \geq 600) = p(600) + p(601) + \cdots + p(1000)$). Instead, use the empirical rule.

   How many standard deviations is 60 from the mean of $Y_1$?

   How many standard deviations is 600 from the mean of $Y_2$?

d) Apply the empirical rule to get an estimate of $P(Y_1 \geq 60)$.

e) Apply the empirical rule to get an estimate of $P(Y_2 \geq 600)$.

9. The random variable $Y$ represents the number of imperfections in the tread of a new automobile tire. Suppose $Y$ has the following probability function $p(y)$:

| $y$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $p(y)$ | 0.7 | 0.2 | 0.05 | 0.03 | 0.02 |

a) Find the probability of more than one imperfection.

b) What is the average number of imperfections?

c) What is the standard deviation $\sigma$ of $Y$?

d) Suppose two tires are produced independently of each other. What is the probability that both tires each have more than one imperfection?

10. In the previous problem, a tire is considered suitable for sale if it has no imperfections. Suppose the morning shift at the plant produces $n = 100$ tires.

a) What is the probability $p$ that a given tire will have no imperfections?

b) What is the probability all 100 tires from the morning shift are suitable for sale?

c) What is the probability that exactly 95 of the tires from the morning shift are suitable for sale?

d) What is the expected number of tires from the morning shift that are suitable for sale?

11. This problem is a continuation of problems 9 and 10. The defect rate on the tires is considered to be too high. In order to address this problem, the manufacturing process is changed in the hope of increasing the proportion of tires with zero imperfections. $n = 100$ tires from a morning shift are produced under the new conditions to see if the new conditions will lead to a higher proportion of tires suitable for sale. The plant manager wants to test if the change has improved the process. As before, a tire is suitable for sale only if it has no imperfections in its tread.

    a) State the appropriate null and alternative hypotheses in the context of this problem. Be sure to define the parameter used in the statement of $H_0$ and $H_a$.
    b) In the context of this problem, describe a type I error.
    c) In the context of this problem, describe a type II error.
    d) Suppose the plant manager decides to adopt the new (and expensive) change in the production process if the number of tires suitable for production out of the 100 is 74 or greater. You advise the plant manager that this may not be a wise decision. Using the empirical rule, compute the approximate significance level $\alpha$ of the hypothesis test that rejects the null hypothesis if 74 or more good tires are produced. That is, what is probability of committing a type I error if we reject $H_0$ when the number of successes $Y$ is greater than or equal to 74?

12. Survey results indicate that 47% of automobile drivers use their seat belts. In order to obtain a higher rate of seat belt use, a law was passed to require drivers to wear their seat belt. In order to determine if the new law has increased seat belt usage, a random sample of $n = 50$ drivers was observed and it was noted whether or not each of the drivers were using their seat belts. Let $p$ denote the proportion of drivers in the population that use their seat belts since the law was passed. Let $Y$ equal the number of drivers (out of the $n = 50$ observed) that were wearing their seat belts.

    a) If the goal of the new law is to increase the proportion of drivers that use their seat belts, set up an appropriate null and alternative hypothesis **in terms of** $p$ to test if the new law is working.
    b) In plain English, explain what a type I error is in the context of this problem.
    c) In plain English, explain what a type II error is in the context of this problem.
    d) If $H_0$ is true, what is the expected value and standard deviation of $Y$?
    e) If $p = 0.47$, what is the probability of observing exactly 30 drivers wearing their seat belts out of the $n = 50$ drivers observed?
    f) Using the empirical rule, what is the approximate probability that out of the 50 drivers, more than 34 of them were wearing their seat belts?

13. An engineering consultant is sent to solve problems for clients. Previous experience indicates that the consultant is able to successfully solve 75% of the problems.

    a) In a given day, suppose the consultant is sent out to $n = 8$ jobs. What is the expected number of successful jobs?

    b) What is the probability the consultant is successful in all 8 jobs?

    c) What is the probability the consultant is successful in exactly 5 of the 8 jobs?

14. This is a continuation of problem 13. Suppose the consultant attends a training class in the hopes of being able to solve a higher proportion of the service call problems. Let $p$ denote the proportion of calls the consultant can successfully solve after taking the training course (recall that the proportion of successful jobs before the training course was 0.75). We want to test if the training course is successful. Answer the following parts:

    a) State the null and alternative hypotheses for this problem in terms of $p$.

    b) In plain English, what does it mean to commit a type I error in the context of this problem?

    c) In plain English, what does it mean to commit a type II error in the context of this problem?

    d) Suppose the consultant's work is logged for a month after the training class. During this period she had $n = 300$ service jobs. Let $Y$ denote the number of successful jobs out of these 300 jobs. If the training course did not improve her ability to solve problems, what is the expectation and standard deviation of $Y$?

    e) Suppose the consultant successfully solved 240 of the 300 jobs during this month. How likely is it that the consultant would have 240 or more successes out of $n = 300$ trials if the training course did not help (i.e. if $p = .75$)? Use the empirical rule to approximate this probability.

## References

R Development Core Team (2003), "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria.