Probability and Hypothesis Testing

1.1 PROBABILITY AND INFERENCE

The area of *descriptive statistics* is concerned with meaningful and efficient ways of presenting data. When it comes to *inferential statistics*, though, our goal is to make some statement about a characteristic of a population based on what we know about a sample drawn from that population. Generally speaking, there are two kinds of statements one can make. One type concerns *parameter estimation*, and the other *hypothesis testing*.

Parameter Estimation

In parameter estimation, one is interested in determining the magnitude of some population characteristic. Consider, for example an economist who wishes to estimate the average monthly amount of money spent on food by unmarried college students. Rather than testing all college students, he/she can test a sample of college students, and then apply the techniques of inferential statistics to estimate the population parameter. The conclusion of such a study would be something like:

> The **probability** is 0.95 that the population mean falls within the interval of £130-£150.

Hypothesis Testing

In the hypothesis testing situation, an experimenter wishes to test the hypothesis that some treatment has the effect of changing a population parameter. For example, an educational psychologist believes that a new method of teaching mathematics is superior to the usual way of teaching. The hypothesis to be tested is that all students will perform better (i.e., receive higher grades) if the new method is employed. Again, the experimenter does not test everyone in the population. Rather, he/she draws a sample from the population. Half of the subjects are taught with the Old method, and half with the New method. Finally, the experimenter compares the mean test results of the two groups. It is not enough, however, to simply state that the mean is higher for New than Old (assuming that to be the case). After carrying out the appropriate inference test, the experimenter would hope to conclude with a statement like:

> The **probability** that the New-Old mean difference is due to chance (rather than to the different teaching methods) is less than 0.01.

Note that in both parameter estimation and hypothesis testing, the conclusions that are drawn have to do with *probabilities*. Therefore, in order to really understand parameter estimation and hypothesis testing, one has to know a little bit about basic probability.

1.2 RANDOM SAMPLING

Random sampling is important because it allows us to apply the laws of probability to sample data, and to draw inferences about the corresponding populations.

Sampling With Replacement

A sample is random if **each member of the population is equally likely to be selected each time a selection is made**. When N is small, the distinction between *with* and *without* replacement is very important. If one samples with replacement, the probability of a particular element being selected is constant from trial to trial (e.g., 1/10 if $N = 10$). But if one draws without replacement, the probability of being selected goes up substantially as more subjects/elements are drawn.

e.g., if $N = 10$
Trial 1: p(being selected) = 1/10 = .1
Trial 2: p(being selected) = 1/9 = .11111
Trial 3: p(being selected) = 1/8 = .125
etc.

Sampling Without Replacement

When the population N is very large, the distinction between *with* and *without* replacement is less important. Although the probability of a particular subject being selected does go up as more subjects are selected (without replacement), the rise in probability is minuscule when N is large. For all practical purposes then, each member of the population is equally likely to be selected on any trial.

e.g., if $N = 1{,}000{,}000$
Trial 1: p(being selected) = 1 / 1,000,000
Trial 2: p(being selected) = 1 / 999,999
Trial 3: p(being selected) = 1 / 999,998
etc.

1.3 PROBABILITY BASICS

Probability Values

A probability must fall in the range 0.00 - 1.00 . If the probability of event A = 0.00, then A is certain **not** to occur. If the probability of event A = 1.00, then A is certain to occur.

Every event has a *complement*: For example, the complement of event A is the event *not A*, which is usually symbolized as $\overline{A}$. The probability of an event plus the probability of its complement must be equal to 1. That is,

$$p(A) + p(\overline{A}) = 1$$
$$p(A) = 1 - p(\overline{A})$$

This is just the same thing as saying that the event (A) must either happen or not happen.

Computing Probabilities

A priori probability. The *a priori* method of computing probability is also known as the *classical method*. It might help to think of it as the *expected* probability value (e.g., like expected frequencies used in calculating the *chi-squared* statistic).

$$p(A) = \frac{\text{number of events classifiable as } A}{\text{total number of possible events}}$$

A posteriori probability. The *a posteriori* method is sometimes called the *empirical* method. Whereas the *a priori* method corresponds to *expected* frequencies, the *empirical* method corresponds to *observed* frequencies.

$$p(A) = \frac{\text{number of times } A \text{ has occurred}}{\text{total number of events}}$$

EXAMPLE: Consider a fair six-sided die (die = singular of dice). What is the probability of rolling a 6? According to the *a priori* method, p(6) = 1/6. But to compute p(6) according to the *empirical* method, we would have to roll the die some number of times (preferably a large number), count the number of sixes, and divide by the number of rolls. As alluded to earlier, statistics like *chi-squared* involve comparison of *a priori* and empirical probabilities.

1.4 METHODS OF COUNTING: COMBINATIONS & PERMUTATIONS

The preceding definitions indicate that calculating a probability value entails *counting*. The number of things (or events) to be counted is often enormous. Therefore, conventional methods of counting are often inadequate. We now turn to some useful methods of counting large numbers of events.

Permutations

An *ordered* arrangement of *r* distinct objects is called a permutation. For example, there are 6 permutations of the numbers 1, 2, and 3:

1,2,3	1,3,2
2,1,3	2,3,1
3,1,2	3,2,1

In general, there are “r-factorial” permutations of r objects. In this case, *r* = 3, and 3-factorial = 3x2x1 = 6. “Factorial” is symbolized with an exclamation mark, the formula would look like this: 3! = 3x2x1 = 6.

Note that one is not always interested in taking all r cases at a time. If we had 4 numbers rather than 3, for example, the number of permutations of those 4 numbers would be 4! = 4x3x2x1 =

24. But what if we only wished to take the numbers 2 at a time? In this case, there would be only 12 permutations:

1,2	1,3	1,4
2,1	2,3	2,4
3,1	3,2	3,4
4,1	4,2	4,3

The number of ways of ordering *n* distinct objects taken *r* at a time is given by the following:

$$P_r^n = n(n-1)(n-2)...(n-r+1) = \frac{n!}{(n-r)!}$$

EXAMPLE: Suppose that there are 10 horses in a race, and that event A has been defined as correct selection the top 3 finishers in the correct order (i.e., correct selection of the *Win*, *Place* and *Show* horses). What is the probability of A?

$$p(A) = \frac{\text{number of ways to pick top 3 horse in correct order}}{\text{number of permutations of 10 horses chosen 3 at a time}}$$

$$= \frac{1}{\frac{10!}{(10-3)!}} = \frac{1}{720} = .00139$$

> Note that there is often more than one way to solve a probability problem, and that some ways are easier than others. Later on we will solve this same problem in another way that I find more intuitively appealing.

Combinations

A combination is similar to a permutation, but differs in that **order** is not important. ABC and BCA are different permutations, but ABC is the same combination as BCA. The number of combinations of *n* things taken *r* at a time is given by:

$$C_r^n = \frac{P_r^n}{r!} = \frac{n!}{r!(n-r)!}$$

EXAMPLE: A camp counselor is supervising 6 campers, 3 girls and 3 boys. The counselor chooses 3 campers to chop wood and 3 to wash the dishes. If the assignment of children to chores is random, what is the probability that the 3 girls will be asked to wash dishes, and the 3 boys to chop wood?

Solution: Let A = assignment of 3 girls to dishwashing and 3 boys to wood-chopping. There is only 1 way for A to happen. The total number of possible outcomes = the number of combinations of 6 campers chosen 3 at a time. So,

$$p(A) = \frac{1}{C_3^6} = \frac{1}{\frac{6!}{3!3!}} = \frac{1}{\frac{720}{36}} = \frac{1}{20} = 0.05$$

Thus, if the assignment of campers to chores was random, the probability of 3 girls being assigned to dish-washing and 3 boys to wood-chopping is 0.05 (or 5%).

NOTE: There are always r! permutations for every combination.

A More General Combinatorial Formula

The number of ways of partitioning $n_\bullet$ distinct objects into k distinct groups containing $n_1, n_2, \ldots n_k$ objects respectively is given by:

$$C^{n_\bullet}_{n_1, n_2, \ldots n_k} = \frac{n_\bullet}{n_1!n_2!\ldots n_k!} \qquad \text{where } n_\bullet = \sum_{i=1}^{k} n_i$$

EXAMPLE: How many ways can an ice hockey coach assign 12 forwards to 4 forward lines (where each line consists of 3 players)?

$$C^{12}_{3,3,3,3} = \frac{12!}{3!3!3!3!} = 369,600$$

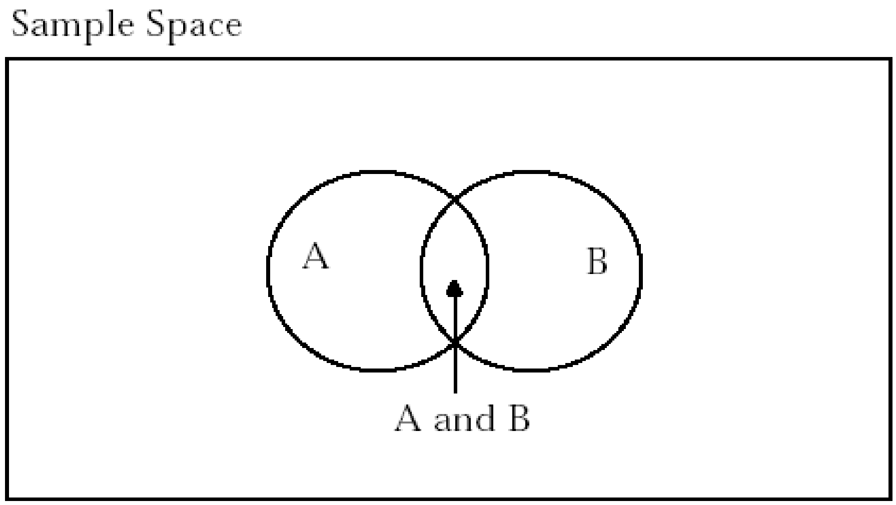
NOTE: The formula for combinations is a special case of this more general formula with *k* = 2.

Venn Diagrams

Venn diagrams are commonly used to illustrate probabilities and relationships between events. A few words of explanation may be helpful.

The first part of a Venn diagram is a rectangle that represents the *sample space*. The sample space encompasses all possible outcomes in the situation being considered. For example, if you are considering all possible outcomes when drawing one card from a standard deck of playing cards, the sample space consists of the 52 cards.

Inside of the sample space, circles are drawn to represent the various *events* that have been defined. If the sample space is 52 playing cards, event A might be getting a red card, and event B might be getting a King. Event A in this case would consist of 26 outcomes (the 26 red cards); and event B would consist of 4 outcomes (the 4 Kings). Note that there would be two outcomes common to A and B, the King of Hearts and the King of Diamonds. This would be represented in the Venn diagram by making the two circles *overlap*.



One other important about Venn diagrams is this: **If they are drawn to scale** (which is **not** the case here), the probability of event A is equal to the *area of circle A* divided by the *area of the entire sample space*. In the example above, event A was getting a red card when drawing one card from a standard deck. In this case, p(A) = .5; and so a Venn diagram drawn to scale would show circle A taking up exactly half of the sample space.

ADDITION RULE: How to calculate p(A or B)

Definition of "or"

In everyday use, the word "or" is typically used in the **exclusive** sense. That is, when we say "A or B", we usually mean one or the other, but **not both**. In probability (and logic) though, "or" is **inclusive**. In other words, "A or B" includes the possibility of "A **and** B":

“or” in probability (and logic) = “and/or” in everyday language

Finally, note that because or is inclusive, p(A or B) is often expressed as $p(A \cup B)$, which stands for “A union B”.

EXAMPLE: Imagine drawing 1 card at random from a well-shuffled deck of playing cards. And define events A, B, and C as follows:

A = the card is a face card (Jack, Queen, King)
B = the card is a red card (Hearts, Diamonds)
C = A **or** B

In this example, C is true if the card drawn is either a face card (A), or a red card (B), or both a face card and a red card (both A and B).

Addition Rule: General Case

To calculate p(A or B), begin by adding p(A) and p(B). Note however, that if events A and B intersect (see the Venn diagram on p. 6), events that fall in the intersection (A and B) have been counted *twice*. Therefore, they must be subtracted *once*. Thus, the formula for computing the probability of (A or B) is as follows:

$$p(A \text{ or } B) = p(A) + p(B) - p(A \text{ and } B)$$

Special Case for Mutually Exclusive Events

Two events are **mutually exclusive** if they cannot occur together. In other words, A and B are mutually exclusive if the occurrence of A precludes the occurrence of B, and vice versa. In other words, if events A and B are mutually exclusive, then $p(A \text{ and } B) = 0$; and if $p(A \text{ and } B) = 0$, then events A and B are mutually exclusive.

Therefore, when the addition rule is applied to mutually exclusive events, it simplifies to:

$$p(A \text{ or } B) = p(A) + p(B)$$

NOTE: In a Venn diagram, mutually exclusive events are *non-overlapping*.

Addition Rule for more than 2 mutually exclusive events

The addition rule may also be applied in situations where there are more than two events, provided that the occurrence of one of the events precludes the occurrence of all others. The rule is as follows:

$$p(A \text{ or } B \text{ or } C ... \text{or } Z) = p(A) + p(B) + p(C) ... + p(Z)$$

*Mutually exclusive and **exhaustive** sets of events*

A set of events is **exhaustive** if it includes all possible outcomes. For example, when rolling a 6-sided die, the set of outcomes of 1,2,3,4,5, or 6 is exhaustive, because it contains all possible outcomes. The outcomes in this set are also mutually exclusive, because the occurrence of one number precludes all others. *When a set of events is exhaustive, and the outcomes in that set are mutually exclusive, then the sum of the probabilities for the events must equal 1.* For a fair 6-sided die, for example, if we let X = the number showing, X would have the following probability distribution:

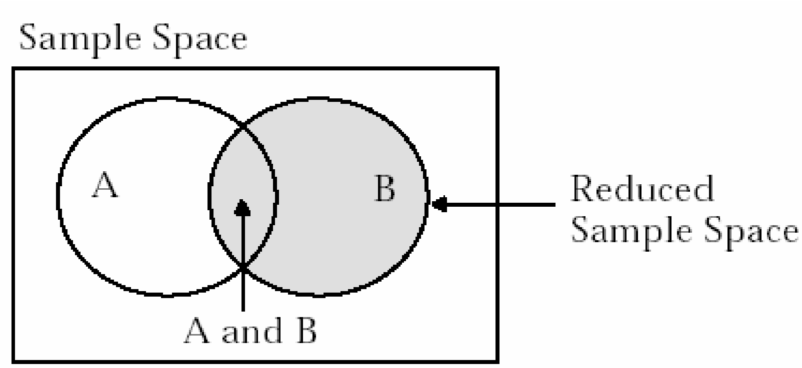
X	$p(X)$	$\sum p(X)$
1	1/6	1/6
2	1/6	2/6
3	1/6	3/6
4	1/6	4/6
5	1/6	5/6
6	1/6	6/6 = 1

CONDITIONAL PROBABILITY

The concept of *conditional* probability is very important in hypothesis testing. It is standard to use the word "given" when talking about conditional probability. For example, one might want to know the probability of event A *given* that event B has occurred. This conditional probability is written symbolically as follows:

$$p(A \mid B)$$

When you are *given* (or told) that event B (or any other event) has occurred, this allows you to *delimit, or reduce the sample space*. In the following Venn Diagram that illustrates conditional probability, for example, when computing *p(A|B),* we can immediately delimit the sample space to the area occupied by circle B. In effect, circle B has become the sample space, and anything outside of it does not exist. And so, when a Venn diagram is drawn to scale, *p(A|B)* is equal to the ratio of the *area in the (A and B) intersection* divided by the *area of circle B*.



Venn diagram illustrating conditional probability

Putting these ideas into symbols, we get the following:

$$p(A \mid B) = \frac{p(A \text{ and } B)}{p(B)} \quad \text{and} \quad p(B \mid A) = \frac{p(A \text{ and } B)}{p(A)}$$

NOTE: In calculating conditional probabilities, always divide by the probability of *that which was given*.

MULTIPLICATION RULE

General Rule

Related to conditional probability is the question of how to compute $p(A \text{ and } B)$. Note the formulae for computing conditional probabilities can be rearranged to isolate the expression $p(A \text{ and } B)$:

$$(1)\quad p(A \text{ and } B) = p(A)p(B \mid A)$$
$$(2)\quad p(A \text{ and } B) = p(B)p(A \mid B)$$

These two equations represent the *general multiplication rule*.

Special Case for Independent Events

There is also a special multiplication rule for *independent* events. Before looking at it, let us define independence. **Two events, A and B, are independent if the *occurrence of one has no effect on the probability of occurrence of the other***. Thus, A and B are independent if, and only if:

$$p(A \mid B) = p(A \mid \overline{B}) = p(A) \quad \text{and}$$
$$p(B \mid A) = p(B \mid \overline{A}) = p(B)$$

And conversely, if A and B are independent, then the two equations shown above must be true.

So then, if A and B are independent, $p(A \mid B) = p(A)$ and $p(B \mid A) = p(B)$. Therefore, the equations that represent the general multiplication rule (shown above) simplify to the following *special multiplication rule* for two independent events:

$$p(A \text{ and } B) = p(A)p(B)$$

If you have more than 2 events, and they are all independent of each other, the special multiplication rule can be extended as follows:

$$p(A \text{ and } B \text{ and } C\text{... and } Z) = p(A)p(B)\text{...}p(Z)$$

Three or more related events

When you have two related events, the general multiplication rule tells you that

$p(A \text{ and } B) = p(A)p(B \mid A)$. When you have 3 events that are all related to one another, this can be extended as follows:

$$p(A \text{ and } B \text{ and } C) = p(A)p(B \mid A)p(C \mid AB)$$

And for 4 related events:

$$p(A \text{ and } B \text{ and } C \text{ and } D) = p(A)p(B \mid A)p(C \mid AB)p(D \mid ABC)$$

These extensions to the general multiplication rule are useful in situations that entail *sampling without replacement*. For example, if you draw 5 cards from a standard playing deck without replacement, what is the probability that all 5 will be Diamonds? The probability that the first card drawn is a Diamond is 13/52, or .25. If the first card is in fact a Diamond, note that the number of remaining Diamonds is now 12. And of course, the number of cards remaining is 51. Therefore, the probability that the 2nd card is a Diamond, *given* that the first card was a Diamond, is 12/51. Using the same logic, the probability that the 3rd card is a Diamond, given that the first two were both Diamonds, is 11/50, and so on. So the solution to this problem is given by:

$$p(\text{all 5 are Diamonds}) = \left(\frac{13}{52}\right)\left(\frac{12}{51}\right)\left(\frac{11}{50}\right)\left(\frac{10}{49}\right)\left(\frac{9}{48}\right) = .0005$$

For another example of how to use these extensions of the general multiplication, let us return to the horse racing problem described in the section on permutations. The problem stated that there are 10 horses in a race. Given that you have no other information, what is the probability that you will correctly pick the top 3 finishers (in the right order)? We first solved this problem by using the formula for permutations. But given what we now know, there is another way to solve the problem that may be more intuitively appealing. The probability of picking the winner is 1/10. The probability of picking the 2nd place finisher *given* that we have successfully picked the winner is 1/9. And finally, the probability of picking the 3rd place finisher *given* that we have picked the top two horses correctly, is 1/8. So the probability of picking the top 3 finishers in the right order is:

$$p(\text{top 3 horses in right order}) = \left(\frac{1}{10}\right)\left(\frac{1}{9}\right)\left(\frac{1}{8}\right) = \frac{1}{720} = .00139$$

In my opinion, this is a much easier and more understandable solution than the first one we looked at.

GENERAL POINT: A general point worth emphasizing again is that there is usually more than one way to solve a probability problem, and that some ways are easier than others—sometimes a *lot* easier.

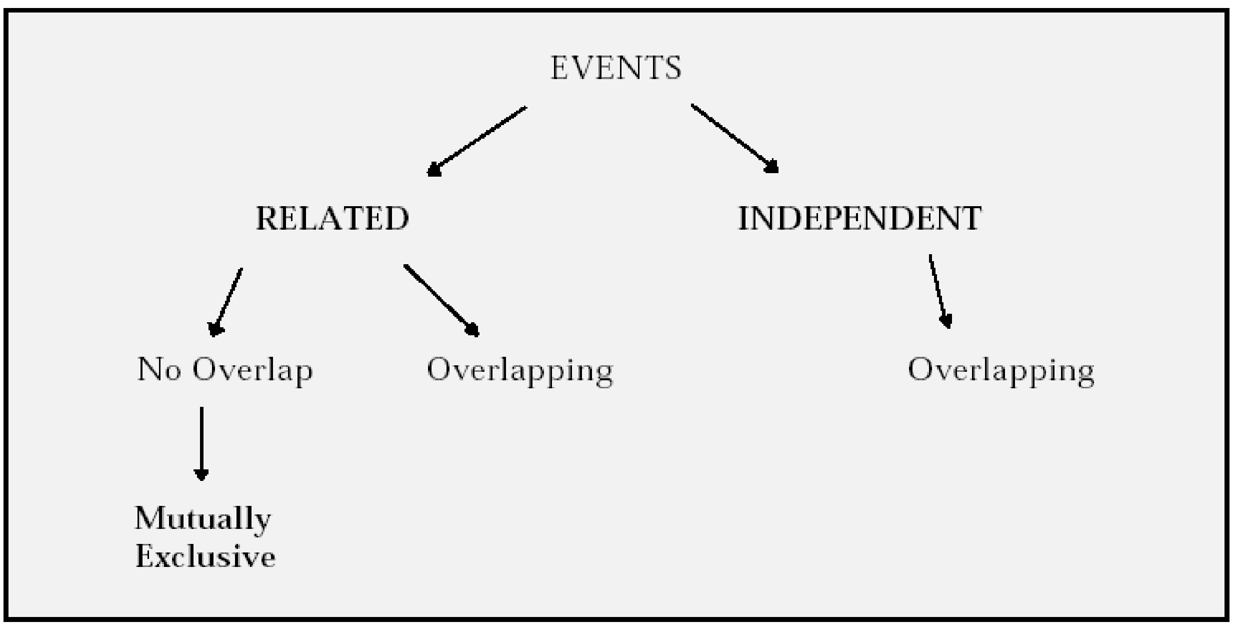
Mutually Exclusive Events

Two events, A and B, are mutually exclusive if both cannot occur at the same time. Thus, if A and B are mutually exclusive, $p(A \text{ and } B) = 0$. Note also that if you are told that $p(A \text{ and } B) = 0$, you can conclude that A and B are mutually exclusive.

NOTE: Because $p(A \text{ and } B) = 0$ for mutually exclusive events, $p(A|B)$ *and* $p(B|A)$ must both be equal to 0 as well (see section on conditional probability).

A TAXONOMY OF EVENTS

In the preceding sections, I have been throwing around terms such as independent, mutually exclusive, and related. There often seems to be a great deal of confusion about what these terms mean. Students are prone to thinking that independence and mutual exclusivity are the same thing. They are not, of course. One way to avoid this confusion is to lay out the different kinds of events in a chart such as this:



From this chart, it should be clear that independence is not the same thing as mutual exclusivity. If events A and B are not independent, then they are related. (Think of related and independent are opposites.) Because mutually exclusive events are related events, they cannot be independent. This chart also illustrates that overlap is a *necessary* but not a *sufficient* condition for two events to be independent. That is, if two events do not overlap, they definitely are not independent. But if they do overlap, they may or may not be independent. The following points, some of which are repeated, are also noteworthy in this context:

1. Sampling *with replacement* produces a situation in which successive trials are independent.

2. Sampling *without replacement* produces a situation in which successive trials are related.
3. The *complementary* events *A* and *not A* are *mutually exclusive and exhaustive*. Therefore, the following are true:

$$p(A) + p(\overline{A}) = 1$$
$$p(A) = 1 - p(\overline{A})$$

It was suggested earlier that there is often more than one way to figure out the probability of some event, but that some ways are easier than others. If you are finding that direct calculation of p(A) is very difficult or confusing, always remember that an alternative approach is to figure out *p*(*A*) and subtract it from 1. In some cases, this approach will be much easier. Consider the following problem, for example:

> *If you roll 3 fair 6-sided dice, what is the probability that at least one 2 or one 4 will be showing?*

One way to approach this is by directly calculating the probability of the event in question. For example, I could define the following events:

Let A = 2 or 4 showing on die 1 p(A)=1/3
Let B = 2 or 4 showing on die 2 p(A)=1/3
Let C = 2 or 4 showing on die 3 p(A)=1/3

And now the event I am interested in is (A or B or C). I could apply the rules of probability to come up with following:

$$p(A \text{ or } B \text{ or } C) = p(A) + p(B) + p(C) - p(AB) - p(AC) - p(BC) + p(ABC)$$

I have left out several steps, which you don't need to worry about.[1] The important point is that this method is relatively cumbersome. A much easier method for solving this problem falls out if you recognize that the complement of *at least one 2 or one 4* is *no 2's and no 4's*. In other words, if there are NOT no 2's and no 4's, then there must be at least one 2 or one 4. So the answer is 1 - p(no 2's and no 4's). With this approach, I would define events as follows:

Let A = no 2's and no 4's on 1st die p(A) = 2/3
Let B = no 2's and no 4's on 2nd die p(B) = 2/3
Let C = no 2's and no 4's on 3rd die p(C) = 2/3

In order for there to be no 2's or 4's, all 3 of these events must occur. If we assume that the 3 dice are all independent of each other, which seems reasonable, then *p(ABC) = p(A) p(B) p(C)* = 8/27. And the answer to the original question is 1- 8/27, or .7037.

---

[1] If you're curious about where this formula comes from, draw yourself a Venn diagram with 3 overlapping circles for events A-C, and see if you can figure it out.

PROBABILITY AND CONTINUOUS VARIABLES

Most of the cases we have looked at so far in our discussion of probability have concerned discrete variables. Thus, we have thought about probability of some event A as the number of outcomes that we can classify as A divided by the total number of possible outcomes.

Many of the dependent variables measured in experiments are continuous rather than discrete, however. When a variable is continuous, we must conceptualize the probability of A in a slightly different way. Specifically,

$$p(A) = \frac{\text{area under the curve that corresponds to } A}{\text{total area under the curve}}$$

Note that if you are dealing with a **normally distributed continuous** variable, you can carry out the appropriate z-transformations and make use of a table of the standard normal distribution to find probability values.

THE BINOMIAL DISTRIBUTION

Definition: The **binomial distribution** is a discrete probability distribution that results from a series of N **Bernoulli** trials. For each trial in a Bernoulli series, there are only two possible outcomes. These outcomes are often referred to as success (S) and failure (F). These two outcomes are *mutually exclusive* and *exhaustive*, and there is *independence between trials*. Because of the independence between trials, the probability of a success (p) remains constant from trial to trial.[2]

Example 1: Tossing a Fair Coin N Times

When the number of Bernoulli trials is very small, it is quite easy to generate a binomial distribution without recourse to any kind of formula: You can just list all possible outcomes, and apply the special multiplication rule for independent events. Consider the following examples where we toss a fair coin *N* times, and let X = the number of Heads obtained.

a) Fair coin with *N*=1 toss

X	p(X)	Sequences that yield X
0	0.5	{ T }
1	0.5	{ H }
	1.0	

[2] Coin flipping is often used to illustrate the binomial distribution. Independence between trials refers to the fact that the coin has no memory. That is, if the coin is fair, the probability of a Head is .5 every time it is flipped, regardless of what has happened on previous trials.

When $N = 1$, we have a 50:50 chance of getting a Head, and a 50:50 chance of getting a Tail.

b) Fair coin with $N = 2$ tosses

X	p(X)	Sequences that yield X
0	0.25	{ TT }
1	0.50	{ HT, TH }
2	0.25	{ HH }
	1.00	

When $N = 2$, four different sequences are possible: TT, HT, TH, or HH. Given that the two tosses are independent, each of these sequences is equally likely. There is one way to get X=0; two ways to get X=1; and one way to get X=2.

c) Fair coin with $N = 3$ tosses

X	p(X)	Sequences that yield X
0	0.125	{ TTT }
1	0.375	{ HTT, THT, TTH }
2	0.375	{ HHT, HTH, THH }
3	0.125	{ HHH }
	1.00	

In this case, there are 8 equally likely sequences of Heads and Tails, as shown above. Because there are 3 such sequences that yield $X = 2$, p(x=2) = 3/8 = .375.

Example 2: Tossing a ***Biased*** *Coin N Times*

Now let us assume that we have biased coin. Heads and Tails are not equally likely when we flip this coin. Let $p = .8$ and $q = .2$, where p stands for the probability of a Head, and q the probability of a Tail. Toss this biased coin N times, and let x = the number of Heads obtained for:

a) Biased coin with N=1 toss

X	p(X)	Sequences that yield X
0	0.2	{ T }
1	0.8	{ H }
	1.0	

b) Biased coin with $N = 2$ tosses

X	p(X)	Sequences that yield X
0	0.04	{ TT }
1	0.32	{ HT, TH }
2	0.64	{ HH }
	1.00	

When $N = 2$, four different sequences are possible: TT, HT, TH, or HH. What is the probability of a particular sequence occurring? If you remember that the trials are independent (i.e., the coin has no memory for what happened on the previous toss), then you can use the *special multiplication rule for independent events* to figure out the probability of any sequence. Take for example getting a Tail on both trials:

$$p(\text{Tails on both trials}) = p(\text{Tail on trial 1})p(\text{Tail on trial 2}) = (.2)(.2) = .04$$

Applying the special multiplication rule to all of the possible sequences we get the following:

$$p(\text{TT}) = (.2)(.2) = .04$$
$$p(\text{HT}) = (.8)(.2) = .16$$
$$p(\text{TH}) = (.2)(.8) = .16$$
$$p(\text{HH}) = (.8)(.8) = .64$$

Note that the only way for X=0 to occur is if the sequence TT occurs. The probability of this sequence is equal to .04, as shown above. Likewise, the only way for X=2 to occur is if the sequence HH occurs. The probability of this sequence is equal to .64, as shown above.

That leaves X=1. There are 2 ways we could solve $p(X=1)$. The easier way is to recognize that the sum of the probabilities in any binomial distribution has to be equal to 1, and that we the final probability can be obtained by subtraction. Putting that into symbols, we have:

$$p(X=1) = 1 - [p(X=0) + p(X=2)] = 1 - [.04 + .64] = .32$$

The other way to solve $p(x=1)$ is to recognize that the two sequences that yield X=1 are **mutually exclusive** sequences. In other words, you know that if the sequence HT occurred when you tossed a coin twice, the sequence TH could not possibly have occurred. The occurrence of one precludes occurrence of the other, which is exactly what we mean by mutual exclusivity. We also know that X=1 if HT occurs, **or** if TH occurs. So $p(X=1)$ is really equal to p(HT **or** TH). Applying the special addition rule for mutually exclusive events, we get

$$p(\text{HT or TH}) = p(\text{HT}) + p(\text{TH}) = .16 + .16 = .32$$

c) Biased coin with $N = 3$ tosses

X	p(X)	Sequences that yield X
0	0.008	{ TTT }
1	0.096	{ HTT, THT, TTH }
2	0.384	{ HHT, HTH, THH }
3	0.512	{ HHH }
	1.00	

In this case there are 8 possible sequences. Applying the special multiplication rule, we can arrive at these probabilities for the 8 sequences:

$$
\begin{aligned}
p(\text{TTT}) &= (.2)(.2)(.2) = (.8)^0(.2)^3 = .008 \\
p(\text{HTT}) &= (.8)(.2)(.2) = (.8)^1(.2)^2 = .032 \\
p(\text{THT}) &= (.2)(.8)(.2) = (.8)^1(.2)^2 = .032 \\
p(\text{TTH}) &= (.2)(.2)(.8) = (.8)^1(.2)^2 = .032 \\
p(\text{HHT}) &= (.8)(.8)(.2) = (.8)^2(.2)^1 = .128 \\
p(\text{HTH}) &= (.8)(.2)(.8) = (.8)^2(.2)^1 = .128 \\
p(\text{THH}) &= (.2)(.8)(.8) = (.8)^2(.2)^1 = .128 \\
p(\text{HHH}) &= (.8)(.8)(.8) = (.8)^3(.2)^0 = .512
\end{aligned}
$$

And application of the special addition rule for mutually exclusive events allows us to calculate the probability values shown in the preceding table. For example,

$$
\begin{aligned}
p(X = 1) &= p(\text{HTT or THT or TTH}) \\
&= p(\text{HTT}) + p(\text{THT}) + p(\text{TTH}) \\
&= .032 + .032 + .032 = .096
\end{aligned}
$$

A formula for calculating binomial probabilities

In the examples above, note that *each sequence that yields a particular value of X has the same probability of occurrence*. This is the case for all binomial distributions. And because of it, you don't have to compute the probability of each sequence and then add them up. Rather, you can *compute the probability of one sequence, and multiply by the number of sequences*. In the case we have just considered, for example,

$$p(X = 1) = (3\text{ sequences})(.032\text{ per sequence}) = .096$$

When the number of trials in a Bernoulli series becomes even just a bit larger (e.g., 5 or more), the number of possible sequences of successes and failures becomes very large. And so, in these cases, it is not at all convenient to generate binomial distributions in the manner shown above. Fortunately, there is a fairly straightforward formula that can be used instead. It is shown below:

$$p(X) = \frac{N!}{X!(N-X)!)} p^X q^{N-X}$$

where N = the number of trials in the Bernoulli series
X = the number of "successes" in N trials
p = the probability of success on each trial
$q = 1 - p$ = probability of failure on each trial

To confirm that it works, you should use it to work out the following binomial distributions:

1) Let $N = 3$, $p = .5$
2) Let $N = 3$, $p = .8$
3) Let $N = 4$, $p = .5$
4) Let $N = 4$, $p = .8$

Note that the first two are distributions that we worked out earlier, so you can compare your answers.

The formula shown above can be broken into two parts. You should recognize the first part, *N!/[X!(N-X)!],* as giving the number of *combinations of N things taken x at a time*. In this context, it may be more helpful to think of it as giving the number of sequences of successes and failures that will result in there being X successes and (N-X) failures. To make this a little more concrete, think about flipping a fair coin 10 times:

Let $N = 10$
Let X = the number of Heads
Let p = .5, q = .5

To calculate the probability of exactly 6 heads, for example, the first thing you have to know is *how many sequences of Heads and Tails result in exactly 6 Heads?* The answer is: 10!/(6!4!) = 210. There are 210 sequences of Heads and Tails that result in exactly 6 Heads.

Note that these sequences are all mutually exclusive. That is to say, the occurrence of one sequence excludes all others from occurring. Therefore, if we knew the probability of each sequence, we could figure out the probability of getting sequence 1 or sequence 2 or...sequence 210 by adding up the probabilities for each one (special addition rule for mutually exclusive events).

Note as well that each one of these sequences has the same probability of occurrence. Furthermore, that probability is given by the second part of our formula, $p^X q^{N-X}$.

And so, conceptually, the formula for computing binomial probability
could be stated as: *(the number of sequences of Successes and Failures that results*

in X Successes and (N-X) Failures) multiplied by (the probability of each of those sequences).

$$p(X) = \left[\frac{N!}{X!(N-X)!} \right] \times \left[p^X q^{N-X} \right]$$

The number of sequences of Successes and Failures that results in *X* Successes and *N-X* Failures

The probability of each one of those sequences.

It is also possible to generate binomial probabilities using the *binomial expansion.* For example, for N = 4,

$$\begin{aligned}(p+q)^4 &= p^4q^0 + 4p^3q^1 + 6p^2q^2 + 4p^1q^3 + p^0q^4 \\ &= p^4 + 4p^3q^1 + 6p^2q^2 + 4p^1q^3 + q^4\end{aligned}$$

But many students find the formula I gave first easier to understand.

Note as well that some (older) statistics textbooks have Tables of the binomial distribution, but usually only for particular values of *p* and *q* (e.g., multiples of .05).

HYPOTHESIS TESTING

Suppose I have two coins in my pocket. One is a **fair** coin—i.e., p(Head) = p(Tail) = 0.5. The other coin is **biased toward Tails**: p(Head) = .15, p(Tail) = .85.

I then place the two coins on a table, and allow you to choose one of them. You take the selected coin, and flip it 11 times, noting each time whether it showed Heads or Tails.

Let *X* = the number of Heads observed in 11 flips.

Let hypothesis A be that you selected and flipped the fair coin.

Let hypothesis B be that you selected and flipped the biased coin.

Under what circumstances would you decide that hypothesis A is true?
Under what circumstances would you decide that hypothesis B is true?

	<u>FAIR COIN</u> p = p(Head) = 0.5 q = p(Tail) = 0.5
	<u>BIASED COIN</u> p = p(Head) = 0.15 q = p(Tail) = 0.85

A good way to start is to think about what kinds of outcomes you would expect for each hypothesis. For example, if hypothesis A is true (i.e., the coin is fair), you expect the number of Heads to be somewhere in the middle of the 0-11 range. But if hypothesis B is true (i.e., the coin is biased towards tails), you probably expect the number of Heads to be quite small.

Note as well that a very large number of Heads is improbable in either case, but is *less* probable if the coin is biased towards tails.

In general terms, therefore, we will decide that the coin is biased towards tails (hypothesis B) if the number of Heads is quite low; otherwise, we will decide that the coin is fair (hypothesis A).

IF the number of heads is LOW

THEN decide that the coin is biased

ELSE decide that the coin is fair.

But an obvious problem now confronts us: That is, **how low is low**? Where do we draw the line between low and middling?

The answer is really quite simple. The key is to recognize that the variable X (the number of Heads) has a binomial distribution. Furthermore, if hypothesis A is true, X will have a binomial distribution with $N = 11$, $p = .5$, and $q = .5$. But if hypothesis B is true, then X will have a binomial distribution with $N = 11$, $p = .15$, and $q = .85$.

We can generate these two distributions using the formula you learned earlier.

Two Binomial Distributions with $N = 11$ and $X = \#$ of Heads

X	$p(X \mid H_A)$	$p(X \mid H_B)$
0	.0005	.1673
1	.0054	.3248
2	.0269	.2866
3	.0806	.1517
4	.1611	.0536
5	.2256	.0132
6	.2256	.0023
7	.1611	.0003
8	.0806	.0000
9	.0269	.0000
10	.0054	.0000
11	.0005	.0000
	1.0000	1.0000

Note that each of these binomial probability distributions
is really a *conditional* probability distribution.

We are now in a position to compare conditional probabilities for particular experimental outcomes. For example, if we actually did carry out the coin tossing experiment and obtained 3 Heads (X=3), we would know that the probability of getting exactly 3 Heads is lower if hypothesis A is true (.0806) than it is if hypothesis B is true (.1517). Therefore, we might decide that hypothesis B is true if the outcome was $X = 3$ Heads.

But what if we had obtained 4 Heads (X=4) rather than 3? In this case the probability of exactly 4 Heads is *higher* if hypothesis A is true (.1611) than it is if hypothesis B is true (.0536). So in this case, we would probably decide that hypothesis A is true (i.e., the coin is fair).

A decision rule to minimize the overall probability of error

In more general terms, we have been comparing the (conditional) probability of a particular outcome if hypothesis A is true to the (conditional) probability of that same outcome if hypothesis B is true. And we have gone with whichever hypothesis yields the higher conditional probability for the outcome. We could represent this decision rule symbolically as follows:

if $p(X \mid \text{A}) > p(X \mid \text{B})$ then choose A
if $p(X \mid \text{A}) < p(X \mid \text{B})$ then choose B

Note that even if the coin is biased towards tails, it is possible for the number of Heads to be very large; and if the coin is fair, it is possible to observe very few Heads. No matter which hypothesis we choose, therefore, there is always the possibility of making an error.[3] However, the use of the decision rule described here will *minimize the overall probability of error*. In the present example, this rule would lead us to decide that the coin is biased if the number of Heads was 3 or less; but for any other outcome, we would conclude that the coin is fair (see below).

Decision rule to minimize the overall probability of error

X	$p(X \mid H_A)$	$p(X \mid H_B)$	
0	.0005	.1673	
1	.0054	.3248	
2	.0269	.2866	Choose B
3	.0806	.1517	
4	.1611	.0536	
5	.2256	.0132	
6	.2256	.0023	
7	.1611	.0003	Choose A
8	.0806	.0000	
9	.0269	.0000	
10	.0054	.0000	
11	.0005	.0000	
	1.0000	1.0000	

Null and Alternative Hypotheses

In any experiment, there are two hypotheses that attempt to explain the results. They are the **alternative hypothesis** and the **null hypothesis**.

Alternative Hypothesis (H_1 or H_A). In experiments that entail manipulation of an independent variable, the alternative hypothesis states that the results of the experiment are due to the effect of the independent variable. In the coin tossing example above, H_1 would state that the biased coin had been selected, and that *p*(Head) = 0.15.

Null Hypothesis (H_0). The null hypothesis is the complement of the alternative hypothesis. In other words, if H_1 is not true, then H_0 must be true, and vice versa. In the foregoing coin tossing situation, H_0 asserts that the fair coin was selected, and that *p*(Head) = 0.50.

Thus, the decision rule to minimize the overall p(error) can be restated as follows:

[3] For the moment, the distinction between Type I and Type II errors is not important. We will get to that shortly.

if p(X | H_0) > p(X | H_1) then do not reject H_0
if p(X | H_0) < p(X | H_1) then reject H_0

Rejecting H_0 is essentially the same thing as choosing H_1, but note that many statisticians are very careful about the terminology surrounding this topic. According to statistical purists, it is only proper to *reject the null hypothesis* or *fail to reject the null hypothesis*. *Acceptance* of either hypothesis is strictly forbidden.

Rejection Region

As implied by the foregoing, the **rejection region** is a range containing outcomes that lead to rejection of H_0. In the coin tossing example above, the rejection region consists of 0, 1, 2, and 3. For $X > 3$, we would fail to reject H_0.

Type I and Type II Errors

When one is making a decision about H_0 (i.e., either to reject or fail to reject it), it is possible to make two different types of errors.

Type I Error. A Type I Error occurs when H_0 is TRUE, but the experimenter decides to reject H_0. In other words, the experimenter attributes the results of the experiment to the effect of the independent variable, when in fact the independent variable had no effect. The probability of Type I error is symbolized by the Greek letter alpha:

$$p(\text{Type I Error}) = \alpha$$

Type II Error. A Type II Error occurs when H_0 is FALSE (or H_1 is TRUE), and the experimenter fails to reject H_0. In this case, the experimenter concludes that the independent variable has no effect when in fact it does. The probability of Type II error is symbolized by the Greek letter beta:

$$p(\text{Type II Error}) = \beta$$

The two types of errors are often illustrated using a 2x2 table as shown below.

	THE TRUE STATE OF NATURE	
YOUR DECISION	**H_0 is TRUE**	**H_0 is FALSE**
Reject H_0	**Type I error (α)**	**Correct rejection of H_0**
Fail to reject H_0	**Correct failure to reject H_0**	**Type II error (β)**

EXAMPLE: To illustrate how to calculate α and β, let us return to the coin tossing experiment discussed earlier.

Decision rule to minimize the overall probability of error[4]

X	$p(X \mid H_0)$	$p(X \mid H_1)$	
0	**.0005**	.1673	
1	**.0054**	.3248	
2	**.0269**	.2866	Reject H_0
3	**.0806**	.1517	
4	.1611	**.0536**	
5	.2256	**.0132**	
6	.2256	**.0023**	
7	.1611	**.0003**	Fail to reject H_0
8	.0806	**.0000**	
9	.0269	**.0000**	
10	.0054	**.0000**	
11	.0005	**.0000**	
	1.0000	1.0000	

$\alpha = .0005 + .0054 + .0269 + .0806 = .1134$
$\beta = .0536 + .0132 + .0023 + .0003 + ... = .0694$

Let us begin with α, the probability of Type I error. According to the definitions given above, *Type I error can only occur if the null hypothesis is TRUE*. And if the null hypothesis is true, then we know that we should be taking our probability values from the distribution that is conditional on H_0 being true. In this case, that means we want the distribution on the left (binomial with $N = 11$, $p = .5$).

We now know which distribution to use, but still have to decide which region to take values from, above the line or below the line. We know that H_0 is true, and we know that an error has been made. In other words, we have rejected H_0. We could reject H_0 for any value of X equal to or less than 3. And so, the probability of Type I error is the sum of the probabilities in the rejection region (from the distribution that is conditional on H_0 being true):

$$\alpha = .0005 + .0054 + .0269 + .0806 = .1134$$

[4] See Appendix A for a graphical representation of the information shown in this table.

REVIEW. Why can we add up probabilities like this to calculate α? Because we want to know the probability of (X=0) **or** (X=1) **or** (X=2) **or** (X=3), and because these outcomes are all mutually exclusive of the others. Therefore, we are entitled to use the *special addition rule for mutually exclusive events*. We will use it again in calculating β.

From the definitions given earlier, we know that a *Type II error can only occur if H_0 is FALSE*. So in calculating β, the probability of Type II error, we must take probability values from the distribution that is conditional on H_1 being true—i.e., the right-hand distribution (binomial with $N = 11$, $p = .15$, $q = .85$). If H_1 is true, but an error has been made, the decision must have been *FAIL to reject* H_0. That means that we are looking at outcomes *outside* of the rejection region, and in the distribution that is conditional on H_1 being true. The probability of Type II error is the sum of the probabilities in that region:

$$\beta = .0536 + .0132 + .0023 + .0003 + ... = .0694$$

Decision Rule to Control α

The decision rule to minimize overall probability of error is rarely if ever used in psychological or medical research. Perhaps many of you have never heard of it before. There are at least two reasons for its scarcity. The first reason is that in order to minimize overall probability of an error, you must be able to specify **exactly** the sampling distribution of the statistic given that H_1 is true. This is rarely possible. (Note that in situations where you cannot specify the distribution that is conditional on H_1 being true, you cannot calculate β either. We will return to this later.)

The second reason is that the two types of errors are said to have different costs associated with them. In many kinds of research, Type I Errors are considered by many to be more costly in terms of wasted time, effort, and money. (We will return to this point later too.)

Therefore, in most hypothesis testing situations, the experimenter will want to ensure that the probability of a Type I error is less than some pre-determined (but arbitrary) value. In psychology and related disciplines, it is common to set $\alpha = 0.05$, or sometimes $\alpha = 0.01$. To ensure that $p(\text{Type I Error}) \leq \alpha$, one must ensure that the sum of the probabilities in the rejection region (assuming H_0 to be true) is equal to or less than α.

EXAMPLE: For the coin tossing example described above, find the decision rule and calculate β for:

a) $\alpha = 0.05$
b) $\alpha = 0.01$

a) In order to find the rejection region that maintains $\alpha \leq .05$, we need to start with the most extreme outcome that is favourable to H_1, and work back towards the middle of the distribution. In this case, that means starting with $X = 0$:

X	$p(X \mid H_0)$	$\sum p(X \mid H_0)$	
0	**.0005**	.0005	
1	**.0054**	.0059	Reject H_0
2	**.0269**	.0327	
3	.0806	.1133	

Here, the right hand column shows the *sum of the probabilities from 0 to X*. As soon as that sum is greater than .05, we know we have gone one step to far, and must back up. In this case, the sum of the probabilities exceeds .05 when we get to X=3. Therefore, the .05 rejection region consists of X values less than 3 (X=0, X=1, and X=2). Note that the actual α level is .0328, which is well below .05.

If X=3 is no longer in the rejection region, then it must now fall into the region used to calculate β. The probability that $X = 3$ given that H_1 is true = .1517 (see the previous table). So β will be equal to the β value we calculated before *plus* .1517:

$$\beta = \mathbf{.1517} + .0536 + .0132 + .0023 + .0003 + ... = .2211$$

b) To find the .01 rejection region, we go through the same steps, but stop and go back one step as soon as the sum of the probabilities exceeds .01.

X	$p(X \mid H_0)$	$\sum p(X \mid H_0)$	
0	**.0005**	.0005	
1	**.0054**	.0059	Reject H_0
2	.0269	.0327	
3	.0806	.1133	

In this case, we are left with a rejection region consisting of only two outcomes: X=0 and X=1. The actual α level is .0059, which is well below the .01 level we were aiming for. Again, because we have subtracted one outcome from the rejection region, that outcome must go into the FAIL to reject region. And from the earlier table, we see that $p(X=2 \mid H_1) = .2866$. Therefore,

$$\beta = \mathbf{.2866} + .1517 + .0536 ...+ .0000 = .5077$$

Finally, note that the more severely α is limited (e.g., .01 is more severe than .05), the more β will increase. And the more β increases, the less sensitive (or powerful) the experiment. (We will address the issue of power soon.)

Directional and Non-directional Alternative Hypotheses

We now turn to the issue of directional (or *one-tailed*) and non-directional (or *two-tailed*) alternative hypotheses. Perhaps the best way to understand what this means is by considering some examples. In each case, set $\alpha \leq .05$, toss a coin 13 times, count the number of Heads, and let the null hypothesis be that it is a fair coin. In each example, we will entertain a different alternative hypothesis.

Example: H_0: $p \leq 0.5$ (i.e., the coin is fair, or biased towards Tails)
H_1: $p > 0.5$ (i.e., the coin is biased towards Heads)

NOTE: As mentioned earlier, H_0 and H_1 are complementary. That is, they must encompass all possibilities. So in this case, according to H_0, $p \leq .5$ rather than $p = .5$. In other words, because the alternative hypothesis states that the coin is biased towards Heads, the null hypothesis must state that it is fair *or* biased towards Tails. In generating the H_0 probability distribution, however, we use $p = .5$. This provides what you could think of as a **worst case for the null hypothesis**. That is, if you can reject the null hypothesis with $p = .5$, you would certainly be able to reject it with any value of $p < .5$.

Let us once again think about what kinds of outcomes are expected under the two hypotheses. If the null hypothesis is true (i.e., the coin is *not* biased towards Heads), we do not expect a particularly large number of Heads. But if the alternative hypothesis is true (i.e., the coin is biased towards Heads), we do expect to see a large number of Heads. So only a relatively large number of Heads will lead to rejection of H_0.

X	$p(X \mid H_0)$	$\sum p(X \mid H_0)$
0	.0001	.0001
1	.0016	.0017
2	.0095	.0112
3	.0349	.0461
4	.0873	.1334
5	.1571	.2905
6	.2095	.5000
7	.2095	.7095
8	.1571	.8666
9	.0873	.9539
10	.0349	.9888
11	.0095	.9983
12	.0016	.9999
13	.0001	1.0000
	1.0000	

If H_0 is true, then X, the number of Heads, will have a binomial distribution with $N = 13$ and $p = .5$ (see above).

Because a large number of Heads will lead to rejection of H_0, let us begin constructing our rejection region with $X = 13$, and working back towards the middle of the distribution. As soon as the sum of the probabilities exceeds .05, we will stop and go back one step:

X	$p(X \mid H_0)$	$\sum p(X \mid H_0)$	
13	.0001	.0001	
12	.0016	.0017	
11	.0095	.0112	Reject H_0
10	.0349	**.0461**	
9	.0873	.1334	
8	.1571	.2905	Fail to reject H_0
etc			

So in this case we would reject H_0 if the number of Heads was 10 or greater, and the probability of Type I error would be .0461.

Example 2: H_0: $p \geq 0.5$ (i.e., the coin is fair, or biased towards Heads)
H_1: $p < 0.5$ (i.e., the coin is biased towards Tails)

Again, if H_0 is true, X, the number of Heads will have a binomial distribution with $N = 13$ and $p = .5$. But this time, we would expect a *small* number of Heads if the alternative hypothesis is true. Therefore, we start constructing the rejection region with $X = 0$, and work towards the middle of the distribution:

X	$p(X \mid H_0)$	$\sum p(X \mid H_0)$	
0	.0001	.0001	
1	.0016	.0017	
2	.0095	.0112	Reject H_0
3	.0349	**.0461**	
4	.0873	.1334	
5	.1571	.2905	Fail to reject H_0
etc			

Now we would reject H_0 if the number of Heads was 3 or less; and again, the probability of Type I error would be .0461.

Example 3: H_0: $p = 0.5$ (i.e., the coin is fair)
H_1: $p \neq 0.5$ (i.e., the coin is biased)

As in the previous two examples, *X*, the number of Heads, will have a binomial distribution with $N = 13$ and $p = .5$ if the null hypothesis is true. But this example differs in terms of what we expect if the alternative hypothesis is true. The alternative hypothesis is that the coin is biased, but no direction of bias is specified. Therefore, we expect either a small or a large number of heads under the alternative hypothesis. In other words, our rejection region must be two-tailed. In order to construct a two-tailed rejection region, we must work with *pairs* of outcomes rather than individual outcomes. We start with the most extreme pair (0,13), and work towards the middle of the distribution:

X	$p(X \mid H_0)$	$\sum p(X \mid H_0)$	
0	.0001	.0001	
1	.0016	.0017	Reject H_0
2	.0095	.0112	
3	.0349	.0461	
4	.0873	.1334	
5	.1571	.2905	
6	.2095	.5000	Fail to reject H_0
7	.2095	.7095	
8	.1571	.8666	
9	.0873	.9539	
10	.0349	.9888	
11	.0095	.9983	
12	.0016	.9999	Reject H_0
13	.0001	1.0000	
	1.0000		

It turns out that we can reject H_0 if the number of Heads is less than 3 or greater than 10; otherwise we cannot reject H_0. The probability of Type I error is the sum of the probabilities in the rejection region, or .0224, which is well below .05.

Note that $X = 10$ was included in the rejection region in Example 1; and $X = 3$ was included in the rejection region in Example 2. But if we include the pair (3,10) in the rejection region in this example, α rises to .0922. Therefore, this pair of values cannot be included if we wish to control α at the .05 level, 2-tailed.

NOTE: It is the **ALTERNATIVE HYPOTHESIS** that is either directional or non-directional. Note as well that in the three examples we have just considered, it is not possible to calculate β, because H_1 is not precise enough to specify a particular binomial distribution.

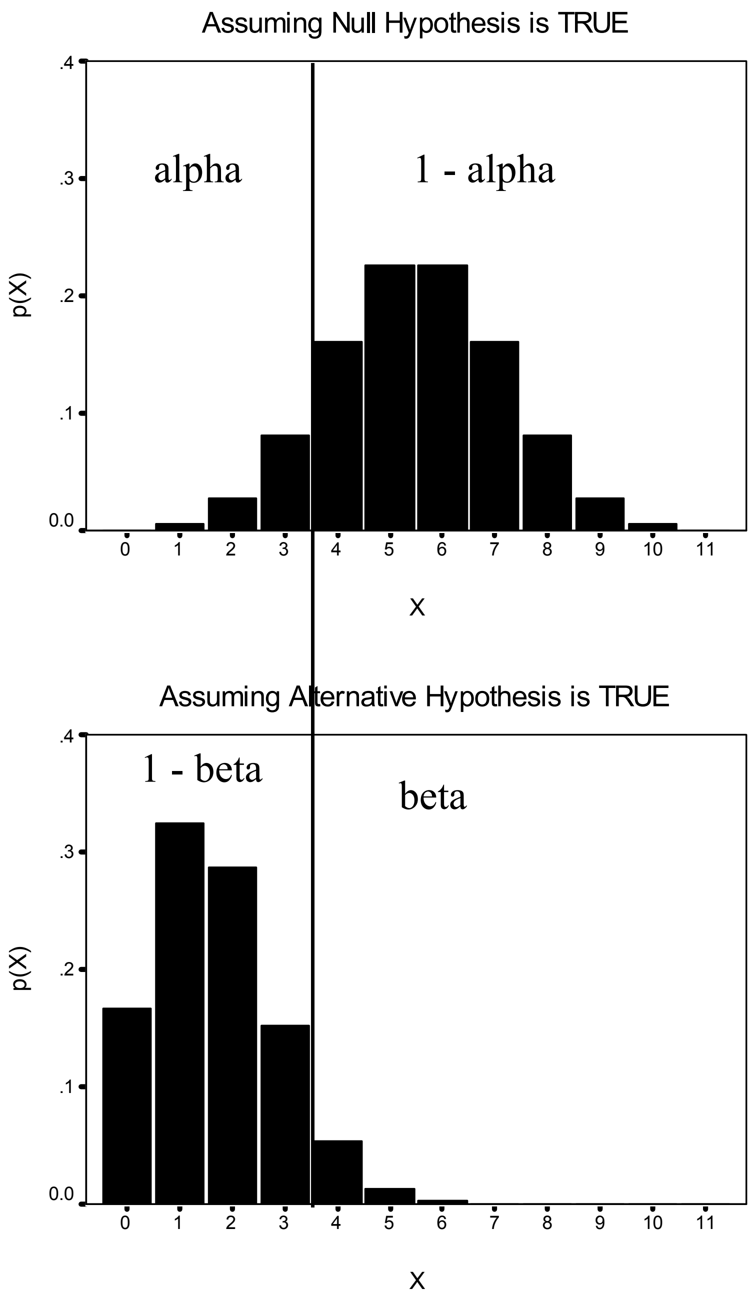
STATISTICAL POWER

We are now in a position to define "power". The power of a statistical test is the **conditional probability that a false null hypothesis will be correctly rejected**. Note that power is also equal to 1 - β.

As we saw earlier, **it is only possible to calculate β when the sampling distribution conditional on H_1 is known**. The sampling distribution conditional on H_1 is rarely known in psychological research. Therefore, it is worth bearing in mind that power analysis (which is becoming increasingly trendy these days) is necessarily based on *assumptions* about the nature of the sampling distribution under the alternative hypothesis, and that the power *estimates* that are produced are only as good as those initial assumptions.

Appendix A: Graphical Illustration of alpha, beta, and power

The figure shown below illustrates the same information that was presented in tabular form on page 23. Alpha = p(Type I error), beta = p(Type II error), and 1-beta = power.



Appendix B: The Likelihood Ratio

Earlier, we stated our decision rule for minimizing the overall probability of an error as follows:

if $p(X \mid H_0) > p(X \mid H_1)$ then do not reject H_0
if $p(X \mid H_0) < p(X \mid H_1)$ then reject H_0

This way of stating the rule implies that we simply use the **difference** between two conditional probabilities as the basis of our decision. In actual practice, we (or at least statisticians) use the **ratio** of these probabilities. The so-called **likelihood ratio** (*l*) is computed as follows:

$$l = \frac{p(X \mid H_0)}{p(X \mid H_1)}$$

Note as well that:

$$\text{if } p(X \mid H_0) < p(X \mid H_1)\text{, then } \frac{p(X \mid H_0)}{p(X \mid H_1)} < 1.0$$

and

$$\text{if } p(X \mid H_0) > p(X \mid H_1)\text{, then } \frac{p(X \mid H_0)}{p(X \mid H_1)} > 1.0$$

Therefore our decision rule to minimize the overall probability of an error can be restated as follows:

if $l < 1$ then reject H_0; if $l > 1$ then do not reject H_0