

January 25, 2008

## Chapter 3: Probability

Suppose that for a particular type of cancer, 50% of the people with the cancer go into remission. In other words, there is a 50-50 chance that a person with this cancer will go into remission with no treatment. A new drug is developed in the hope of improving the remission rate. 100 patients with the cancer take the experimental drug. If the drug does not work, then we would expect about only 50 of the patients to get better. Suppose that we observe 58 of the 100 patients get better after taking the drug – can we then claim that the drug works?

An analogous experiment is to take a fair coin and flip it 100 times. We would expect to get about 50-50 heads and tails. This does not mean that we expect to see exactly 50 tails and 50 heads. It is pretty clear that if we observe 95 tails out of 100 flips that there is something wrong with the coin. What if we observe 58 tails out of 100 flips. Is this unusual? In the cancer example, if we see 58 patients out of 100 getting better, would this be unusual if the drug did not work, i.e. just due to chance? Or is it unusual to observe 58 successes out of 100 trials if the drug did not work?

In order to answer these questions we need a way to quantify the likelihood of observing particular outcomes in an experiment. The branch of mathematics called *probability* allows us to answer the questions posed above. Probability is needed to make statistical inferences, such as whether or not there is strong evidence the new drug is effective.

The results of many scientific studies are presented using *p*-values (see chapter 6) which are probabilities associated with the likelihood of particular outcomes. In order to understand scientific literature, one must understand *p*-values and hence, some knowledge of probability is needed.

The notes for this chapter provide a very brief introduction to probability.

**Definition.** Suppose an experiment is conducted and several different outcomes are possible. Let  $S$  denote the set of all possible outcomes. Then  $S$  is called the *sample space*.

**Definition.** A subset of the sample space is called an *event*.

Notice that the word “set” is used in both of these definitions. Probability theory is closely related to set theory.

Let us consider some very simple examples to illustrate the terminology.

**Example.** Suppose a couple has a child and the sex of the child is noted ( $M$ =boy,  $F$  = girl). Then the sample space is  $S = \{M, F\}$ .

**Example.** Suppose a couple has three children and the sex of each of the children is noted. Then one way to express the sample space is

$$S = \{FFF, FFM, FMF, MFF, FMM, MFM, MMF, MMM\},$$

where the outcome  $FFM$  means the two older children are girls and the youngest is a boy. Define an event  $A$  as the event that the couple has two boys. Then  $A$  is a subset of  $S$ :

$$A = \{FMM, MFM, MMF\}.$$

For a given sample space, we can define many different events. For the couple that has three children, we can define  $A$  as the event that the couple has two boys and the event  $B$  as the event the couple has at least one girl. Then  $B = \{FMM, MFM, MMF, FFM, FMF, MFF, FFF\}$ .

Given any two sets (or events), we can form new sets (events) using *unions* and *intersections*.

**Definition.** The *union* of two events  $A$  and  $B$  is the event that either  $A$  occurs or  $B$  occurs, or they both occur. The union is denoted

$$A \cup B.$$

The key word for union is “or.”

**Definition.** The *intersection* of two events  $A$  and  $B$  is the event that both  $A$  and  $B$  occurs. The intersection is denoted by

$$A \cap B.$$

The keyword indicating an intersection is “and”.

Using the previous example,

$$A \cup B = \{FMM, MFM, MMF, FFM, FMF, MFF, FFF\},$$

and

$$A \cap B = \{FMM, MFM, MMF\}.$$

Note that in this example, the event  $A$  is a subset of the event  $B$ . In other words, if the event  $A$  occurs (i.e. the couple has two boys), then the event  $B$  must have occurred (they had at least one girl).

**Definition.** Two events are *mutually exclusive* if they cannot both simultaneously occur. In terms of set theory, mutually exclusive means the two events are disjoint, i.e. their intersection is the empty set.

In the example of the couple having three children, let  $C$  be the event that the couple has at least two girls:  $C = \{FFF, FFM, FMF, MFF\}$ . Note that events  $A$  and  $C$  are mutually exclusive. The couple cannot have two girls *and* two boys if they only have three children.

**Definition.** The *complement* of an event  $A$ , denoted  $\bar{A}$  is the event that  $A$  does not occur.

In the previous example, if we consider the event that the couple has at least one girl, then the complement of this event is that the couple did not have any girls (i.e. they had all boys).

The goal of probability is to assign a number to events indicating how likely it is that the events will occur. The probability of an event  $A$  is denoted

$$P(A).$$

Here are some properties of probabilities.

1. For any event  $A$ ,

$$0 \leq P(A) \leq 1.$$

That is, probabilities are always numbers between zero and one. The closer the probability of an event is to zero, the more unlikely it is that the event will occur. If an event has probability zero, then the event cannot happen. On the other hand, if an event has probability one, then the event must happen. If an event has probability 0.95 say, then it is very likely to occur.

2.  $P(S) = 1$ . That is, the probability of the sample space is always one. Since the sample space contains all possible outcomes, it is certain (probability equal to one) that one of these outcomes will occur.
3. If two events  $A$  and  $B$  cannot both occur simultaneously (i.e. mutually exclusive), then

$$P(A \cup B) = P(A) + P(B).$$

This last property extends to any countable number of disjoint events: the probability of their union is the sum of their probabilities.

From these properties of probability, one can deduce the following results:

1. (**Additive rule**) for any two events  $A$  and  $B$ :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

2. (**Law of Complements**) For any event  $A$ ,

$$P(A) = 1 - P(\bar{A}).$$

The law of complements can be very useful for computing probabilities of complicated events. For example, there is a famous problem known as the birthday problem. What is the probability that at least two people in a classroom (of unrelated students) share the same birthday? Intuition seems to tell us that the probability would be low if the class was not too large since there are 365 days in a year. The event that at least two people share a birthday is very complicated because there are lots of ways it can happen. However, the complement is easier to deal with. The complement is the event that no two people share a common birthday. The reason this problem is popular is that the answer is not intuitive. The probability that at least two people share a common birthday exceeds 50% when there are at least 23 people in the room.

There are two distinct types of sample spaces encountered in practice: discrete and continuous. A discrete sample space corresponds to experiments where there are only a finite number of possible outcomes or the number of outcomes is *countably infinite* which means that the outcomes can be listed:  $x_1, x_2, x_3, \dots$ . The example of the couple having three children is a discrete example because there are only eight possible outcomes. Another discrete type example with a countably infinite number of possible outcomes is counting the number of days until the next major earthquake in the Ohio Valley. The sample space in this example can be expressed as  $S = \{1, 2, 3, 4, 5, \dots\}$ . In order to compute probabilities for discrete sample spaces, all one has to do is add up the probabilities of all the sample points in the event of interest.

Continuous sample spaces are sets consisting of a continuum of sample points. For example, suppose the experiment of interest is to record the weight of a newborn child. Weight is a continuous variable that can take values on the positive real line. It is impossible to enumerate every possible weight in a list. Other examples with continuous sample spaces involve examples of measuring lengths, volumes, amount of time, etc. In order to compute probabilities for continuous sample spaces, we need to add up probabilities, but over a continuum. This is done using integral calculus. Calculus is not a prerequisite for these notes and details of integration for finding probabilities will not be carried out.

So far we have not addressed the problem of actually assigning probabilities to outcomes of an experiment. For discrete probability spaces, we need to assign a probability to each outcome in the sample space and these probabilities must add up to be one since the total probability must be one. There are various ways probabilities are assigned.

1. Relative Frequency of an event: Suppose we want to assign a probability to an event  $A$ . One way to do this is to perform the experiment many times and let

$$P(A) = \lim_{n \rightarrow \infty} \left[ \frac{\text{number of times } A \text{ occurs}}{\text{number of times experiment is repeated}} \right].$$

2. Knowledge of the experiment: Randomly select a ball from an urn that contains three red balls and two blue balls. Then it makes sense to assign a probability of  $2/5$  to the event of picking a blue ball from the urn. If we flip a fair coin, it makes sense to assign a probability of 0.5 to tails and 0.5 to heads.

**Equally Likely Outcomes.** A common type of experiment in discrete examples is one where the outcomes are equally likely. For example, flip a fair coin 10 times. There are  $2^{10} = 1024$  possible outcomes of head-tails sequences and they are all equally likely. Thus, the probability of any one sequence, say all heads, would be  $1/1024$ . If you play poker with a typical deck of 52 cards, then there is a well-known hierarchy of hands. For instance, a full-house beats a flush. Why is this? If a deck is randomly shuffled, then all possible five card hands are equally likely. The probability of an event of getting a full-house or a flush is computed by counting the number of ways a full-house or flush can occur and dividing these by total number of poker hands possible (which is 2,598,960). For instance, the probability of getting a poker hand of four aces is  $48/(2,598,960)$  since there are only 47 possible five card poker hands with four aces.

### Conditional Probability.

In many experimental situations we often want to know the probability of particular events given that some other event has occurred. For instance, what is the probability an adult over the age 50 will develop high blood pressure? We can estimate this probability by counting the number of people over the age of 50 with high blood pressure and dividing this by the total number of people over the age of 50. Let  $A$  denote the event that a person over the age of 50 has high blood pressure. For the sake of argument, suppose  $P(A) = 0.6$ . Suppose now that we want to know that given the person is male, what is the probability he will develop high blood pressure? This is an example of a *conditional probability*.

**Definition.** The conditional probability of an event  $A$  given  $B$  is defined as

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

That is, if the event  $B$  has occurred and we want to know the probability of the event  $A$ , then we can restrict our sample space to  $B$  only.

Returning to the blood pressure example, let  $B$  be the event the person is male and suppose  $P(B) = 0.5$ . Suppose that the probability that a randomly selected person over the age 50 is male *and* has high blood pressure is 0.4. Note that we have used the word “and” which implies intersection. That is,  $P(A \cap B) = 0.4$ . Then the probability that a randomly selected person over the age of 50 has high blood pressure given that the person is a man is

$$P(A|B) = P(A \cap B)/P(B) = 0.4/0.5 = 4/5 = 0.8.$$

Thus, given the person is male, there is an 80% chance of having high blood pressure. What is the probability of high blood pressure given the person is female?

$$P(A|\bar{B}) = P(A \cap \bar{B})/P(\bar{B}) = 0.2/(1 - P(B)) = 0.2/0.5 = 0.4.$$

**Question.** Why does  $P(A \cap \bar{B})$  equal 0.2?

Now the probability of developing high blood pressure over the age of 50 depends on numerous factors. For instance, the probability of developing high blood pressure may depend on a person's body mass index (BMI) defined as a person's weight (in kg) divided by their squared height (in m):  $\text{BMI} = \text{kg}/\text{m}^2$ . BMI is a continuous measurement. *Logistic regression* is a statistical technique used to compute the probabilities of certain events given the value of a *covariate* (like BMI).

**Independence** Suppose that the probability of having high blood pressure does not depend on whether or not you are male or female. Then we would say that the event of having high blood pressure is *independent* of sex. This was not the case in our hypothetical example above. In that example, the probability of having high blood pressure was higher if you were male. In this example, high blood pressure is *dependent* on sex.

**Definition 1.** Events  $A$  and  $B$  are independent if  $P(A|B) = P(A)$ . That is, if the probability of the event  $A$  occurring given that  $B$  has occurred is the same as if we did not know that  $B$  occurred. Otherwise, the events are dependent. It can easily be shown that if events  $A$  and  $B$  are independent, then  $P(B|A) = P(B)$ .

Let  $A$  and  $B$  be two independent events. Then using the definition of conditional probability and independence, we can write

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

Multiplying both sides of this equation by  $P(B)$  shows that  $P(A \cap B) = P(A)P(B)$  when events  $A$  and  $B$  are independent. This provides us a more convenient definition of independence:

**Definition 2.** Events  $A$  and  $B$  are independent if and only if

$$P(A \cap B) = P(A)P(B).$$

Otherwise, the two events are dependent.

The definition of independence can be extended to any number of events:

**Definition 3.** The events  $A_1, A_2, \dots, A_k$  are *mutually independent* if and only if the probability of the intersection of any sub-collection of the events is equal to the product of the probabilities of the events in the sub-collection. In particular,  $P(A_i \cap A_j) = P(A_i)P(A_j)$  for any pairs  $A_i$  and  $A_j$  (pairwise independence) and

$$P(A_1 \cap A_2 \cap \dots \cap A_k) = P(A_1)P(A_2) \cdots P(A_k).$$

We shall now illustrate this example by introducing the binomial probability distribution. The binomial probability distribution is probably the most well-known discrete probability distributions.

### Example: Binomial Distribution.

A tank of water contains  $n = 10$  fish and the water has been contaminated with a toxin. The probability that a fish will die due to the amount of toxin in the water is  $p = 0.8$ . What is the probability that all the fish die? What is the probability that exactly 2 of the 10 fish die? Let  $A_i$  denote the event that the  $i$ th fish dies for  $i = 1, 2, \dots, 10$ . Assuming the events  $A_1, A_2, \dots, A_{10}$  are mutually independent, then

$$P(\text{All the fish die}) = P(A_1 \cap A_2 \cap \dots \cap A_{10})$$

$$\begin{aligned}
&= P(A_1)P(A_2) \cdots P(A_{10}) \quad (\text{by independence}) \\
&= 0.8 \times 0.8 \times \cdots \times 0.8 \\
&= 0.8^{10} \\
&= 0.10737418
\end{aligned}$$

To find the probability that exactly 2 of the 10 fish die is a bit harder. For example, the event

$$A_1 \cap A_2 \cap \bar{A}_3 \cap \bar{A}_4 \cap \bar{A}_5 \cap \bar{A}_6 \cap \bar{A}_7 \cap \bar{A}_8 \cap \bar{A}_9 \cap \bar{A}_{10}$$

is the event that the first two fish die and the rest survive. By independence, the probability of this event is

$$(0.8)(0.8)(1 - 0.8)^8 = p^2(1 - p)^8.$$

However, this event is just one of many ways that exactly 2 of the 10 fish die (another example is that the first and third fish die and the rest survive). The probability for each of the distinct events where 2 of the 10 fish die is  $p^2(1 - p)^8$ . Thus,

$$P(\text{Exactly 2 of the 10 fish die}) = (\text{Number of ways exactly 2 of 10 fish can die})p^2(1 - p)^8.$$

Computing the probability then becomes a counting problem, i.e. count the number of distinct ways that exactly 2 of the 10 fish die.

This fish example is an example of a **binomial experiment**:

1.  $n$  trials where the outcome of each trial is either a success (S) or a failure (F).
2. The outcomes of the trials are independent of each other.
3. The probability of success on each trial is the same and is denoted by  $p$ . The probability of failure then must be  $1 - p$ , which we will denote by  $q$ :  $q = 1 - p$ .

In the fish example, we have  $n = 10$  trials with success probability  $p = 0.8$ . In order to efficiently count the number of ways of obtaining 2 successes out of 10 trials, do the following: 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:

$$n! = n(n - 1)(n - 2) \cdots 2 \cdot 1 \quad (\text{"n factorial"}).$$

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

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

where  $k = 2$ . 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  $k = 2$  slots out of the  $n = 10$  possible slots to place successes in is

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

The same logic can be applied when  $k = 3$  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  $n!/((n-k)!k!)$  which is the same formula we derived when  $k = 2$ . This formula is the general formula for all values of  $k = 0, 1, \dots, n$ . This expression for counting the number of *combinations of  $n$  objects taken  $k$  at a time* is given by the *binomial coefficient* which is denoted by  $\binom{n}{k}$ :

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (1)$$

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

Returning to our original probability computation, we have

$$\begin{aligned} P(\text{Exactly 2 of the 10 fish die}) &= (\text{Number of ways exactly 2 of 10 fish can die})p^2(1-p)^8 \\ &= \binom{n}{k} p^k (1-p)^{(n-k)} \\ &= \binom{10}{2} 0.8^2 (1-0.8)^{(10-2)} \\ &= \frac{10!}{(10-2)!2!} (0.8^2)(0.2^8) \\ &= 0.000073728. \end{aligned}$$

**Question.** Given a typical deck of 52 playing cards. How many distinct poker hands (5 cards) are possible?

To finish up our discussion of the binomial distribution, we give the following definition:

**Definition.** A *random variable* (typically denoted by  $X$  or  $Y$ ) is a function that associates a number to each outcome in a sample space.

In the binomial example, we can define  $X$  to be the number of successes out of  $n$  trials. In this case,  $X$  is known as a *binomial random variable*.

**Definition.** A *probability mass function* gives the probability associated with each value that a random variable can assume.

If we let  $X$  denote the number of fish that die (out of 10), then, as we have seen,

$$P(X = 10) = p^n = 0.10737418,$$

and

$$P(X = 2) = \binom{10}{2} 0.8^2 (1 - 0.8)^{(10-2)} = 0.000073728.$$

The preceding derivation generalizes for any binomial random variable and the formula for the probability mass function is

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{(n-k)}, \quad (2)$$

for  $k = 0, 1, \dots, n$ .

### Expected Value and Variance of a Discrete Random Variable

Suppose  $X$  is a discrete random variable. That is,  $X$  assigns to each element in a discrete (finite or countably infinite) sample space a unique number. Let  $x_1, \dots, x_k$ , denote the values that  $X$  can assume. What is the average or *expected value* of  $X$ ? The answer is given by the following definition:

**Definition: Expected Value** of a discrete random variable  $X$  is

$$E[X] = \mu = \sum_{i=1}^k x_i P(X = x_i). \quad (3)$$

That is, the expected value is a weighted average of  $X$ , weighted by the probabilities associated with each of its possible values. We denote the expected value by either  $\mu$  or  $E[X]$ . The expected value is the center of gravity for the distribution.

Just as we saw for descriptive statistics, it is useful to have a measure of spread for random variables.

**Definition: The Variance** of a discrete random variable  $X$  is defined to be

$$\sigma^2 = E[(X - \mu)^2] = \sum_{i=1}^k (x_i - \mu)^2 P(X = x_i). \quad (4)$$

If  $X$  is a random variable, then we can define a new random variable  $(X - \mu)^2$ . The variance is simply the expected value of this new random variable. Another way to put it is that the variance is the average squared deviation of  $X$  about its mean  $\mu$ .

For a binomial random variable  $X$ , the expected value is

$$E[X] = \sum_{k=0}^n k \binom{n}{k} p^k (1 - p)^{n-k}.$$

This is a bit complicated to compute, however some algebraic simplification shows that

$$E[X] = np \quad \text{Binomial Random Variables only.}$$

In addition,

$$\sigma^2 = np(1 - p) \quad \text{Binomial Random Variables only.}$$

In the fish example where  $n = 10$  and  $p = 0.8$ , if  $X$  equals the number of fish that die (out of 10), then

$$E[X] = np = 10(0.8) = 8.$$

We would expect 8 fish to die. However if this experiment is carried out, the exact number of fish that actually die may vary from 8. If the experiment were carried out numerous times, the average number of fish that die over all experiments would be 8. The variance is  $\sigma^2 = np(1 - p) = 10(0.8)(1 - 0.8) = 1.6$  and the standard deviation of  $X$  is  $\sigma = \sqrt{np(1 - p)} = \sqrt{1.6} \approx 1.25$ .

**Example.** 50% of the bass fish in a large lake suffer from tumors due to a toxin in the water. There is concern that because the amount of toxin in the lake has increased due to a spill that the proportion of bass fish with tumors has increased. A random sample of  $n = 100$  fish are captured and examined to see if they have the tumors or not. Suppose 66 of the 100 captured fish have the tumors, what can we conclude?

For the sake of argument, *suppose the proportion of fish with tumors has remained the same after the spill.* Let  $X$  denote the number of fish out of the 100 captured that have tumors. Then  $X$  has a binomial distribution with  $n = 100$  and  $p = 0.50$ . The mean (or expected value) of  $X$  is  $\mu = np = 100(0.5) = 50$ . That is, if the percentage of fish with tumors remains at 50%, then we would expect to see 50 fish out of 100 with tumors on average. However, the exact number of fish with tumors would likely vary from 50. This is analogous to flipping a fair coin 100 times. The expected number of heads would be 50 but the actual number of heads is likely to vary from 50. How far from 50 can we expect the number of heads to vary before we conclude the coin is not fair? We can compute the probability of observing 66 or more fish with tumors using (2):

$$\begin{aligned} P(X \geq 66) &= P(X = 66) + P(X = 67) + \dots + P(X = 100) \\ &= \binom{100}{66} 0.5^{66} 0.5^{100-66} + \binom{100}{67} 0.5^{67} 0.5^{100-67} + \dots + \binom{100}{100} 0.5^{100} 0.5^{100-100}. \end{aligned}$$

This is an awfully tedious computation to carry out.

The SAS function “probbnml” computes *cumulative* binomial probabilities which can be used to solve this problem. In SAS, the function  $\text{probbnml}(p, n, k) = P(X \leq k)$  where  $X$  is a binomial random variable. The probability  $P(X \leq k)$  is known as a **cumulative** probability because it represents the cumulative probability of all  $X$  values less than or equal to  $k$ . To compute  $P(X \geq 66)$  using SAS’s `probbnml`, note that from the law of complements that

$$P(X \geq 66) = 1 - P(X < 66) = 1 - P(X \leq 65).$$

The following SAS code will do this computation for us automatically:

```
data;
x=1-probbnml(0.5,100,65);
proc print;
run;
```

The results from SAS give the probability of 66 or more successes out of 100 trials when  $p = 0.5$  as 0.000895. In other words, if the toxin spill did not increase the rate of tumors among the fish, there is only a probability of about 0.000895 of observing 66 or more fish with tumors. We did observe 66 fish with

tumors but if the tumor rate stayed at 50%, it would be very unlikely to observe such an outcome. Thus the assumption that the toxin spill did not increase the tumor rate appears to be false. The probability value of 0.000895 is an example of a  $p$ -value from hypothesis testing.

Another way to get an idea of whether or not 66 successes is plausible if the tumor rate stayed the same is to use the empirical rule. The binomial distribution with  $p = 0.5$  and  $n = 100$  will have a bell-shape (due to the central limit theorem which we will introduce later). If  $p = 0.5$  after the toxin spill, then the standard deviation of  $X$  is  $\sigma = \sqrt{np(1-p)} = \sqrt{100(0.5)(0.5)} = 5$ . If the mean is  $\mu = 50$ , then we see that a value of 66 is more than three standard deviations above the mean. According to the empirical rule, it is highly unusual to see 66 or more fish with tumors if the tumor rate remained at 0.5. In fact, the empirical rule says that the probability of observing values more than 3 standard deviations above the mean is roughly  $(1 - .997)/2 = 0.0015$  which again is a very small probability.

**Hypergeometric Distribution.** Consider the following problem: Suppose we want to estimate the proportion of children in a large school of  $N = 500$  students that have elevated lead levels in their blood. Unbeknownst to school health officials, the number of students with elevated lead levels is  $M = 100$  and therefore  $p = 100/500$  or 20% of the students have elevated blood lead levels. Suppose  $n = 25$  students are random chosen to have their blood tested. For each student, we can record a 0 or 1 response for not-elevated or elevated lead levels respectively. If we let  $X$  equal the number of students (out of 25 sampled) with elevated lead levels in their blood, then  $X$  does **NOT** have a binomial distribution. This is because the sampling is done *without replacement*. That is, once a student has been tested, we will not pick this student again to be tested. The 25 students are all distinct students. Therefore, the 25 trials are not independent and the success probability does not stay the same. If the first student tested tests negative, then the proportion of the remaining students in the population with elevated lead levels is  $100/499$ .

If  $N$  and  $M$  are large numbers, then the binomial distribution will closely approximate the distribution for  $X$ . The exact distribution for  $X$  is known as the *hypergeometric* distribution. One can show that the probability of observing  $k$  successes out of a sample of  $n$  is

$$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}},$$

where  $k = 1, 2, \dots, \min(n, M)$ .

**Multinomial Distribution.** Recall that the binomial distribution is used to compute the probability of the number of “successes” out of  $n$  independent and identical trials (e.g. a coin flip). For the binomial distribution, there are only two possible outcomes for each trial: success and failure. However, there are many examples of experiments consisting of independent and identical trials where there are more than two possible outcomes. For instance, in a random survey of people from a large population, recording the sex of the individuals will result in a binomial distribution (approximately). However, recording the marital status (single, married, divorced say) will result in a *multinomial distribution*. The binomial distribution is a special case of the multinomial distribution. Suppose there are  $k$  potential outcomes for a given multinomial trial where  $p_1, p_2, \dots, p_k$  are the probabilities associated with the  $k$  outcomes (note:  $p_1 + \dots + p_k = 1$ ). Out of  $n$  trials, the probability of observing  $y_1$  occurrences of the first outcome (say single),  $y_2$  occurrences of the second outcome (say married), on up to  $y_k$  occurrences of the  $k$ th outcome

is given by:

$$\frac{n!}{y_1! \cdots y_k!} p_1^{y_1} \cdots p_k^{y_k}.$$