# CHAPTER 5. BINOMIAL EXPERIMENTS.

In this chapter and the next, we are concerned with the following: we take a given experiment and a given event for that experiment; we do repeated independent trials of the experiment; and we consider the occurrence or non-occurrence of the event on each trial. In particular, if $A$ is the given event, if $p$ is the value of $P(A)$ in a chosen probability space for the experiment, and if we make $n$ independent trials of the experiment, we ask: what is the probability that the event $A$ occurs exactly $x$ times in those trials and fails to occur exactly $n-x$ times?

This is a fundamental question in the theory of probability, because if $A$ is observed to occur exactly $x$ times in those $n$ trials, then $\frac{x}{n}$ will be the observed relative frequency of $A$ in those trials. Hence we would want the theoretical answer to our question to agree with the observed stability of relative frequencies. That is to say, we would want our theory to show that values of $\frac{x}{n}$ close to $p$ are more likely than values of $\frac{x}{n}$ far from $p$ We begin with a specific example.

If we roll a die five times, what is the probability that the number 6 occurs exactly twice? Let us say that a roll is a <u>success</u> (<u>S</u>) if we get a 6 , and a <u>failure</u> (<u>F</u>) if we do not. Then we can get exactly two 6's in five rolls by having two successes and three failures in any of the following orders:

| | | | | |
|---|---|---|---|---|
| <u>SSFFF</u> | <u>SFSFF</u> | <u>SFFSF</u> | <u>SFFFS</u> | <u>FSSFF</u> |
| <u>FSFSF</u> | <u>FSFFS</u> | <u>FFSSF</u> | <u>FFSFS</u> | <u>FFFSS</u> |

As we saw at the end of Chapter 4, by the multiplication law for independent events, each of these ten orders has probability $(\frac{1}{6})^2(\frac{5}{6})^3$. Hence the desired probability is:

$$P(\underline{\text{exactly two 6's}}) = 10(\tfrac{1}{6})^2(\tfrac{5}{6})^3 = 0.16.$$

We would expect to get exactly two 6's about 16% of the time.

More generally, if we roll a die $n$ times, what is the probability that the number 6 will occur exactly $x$ of those times? Clearly we need to multiply $(\frac{1}{6})^x(\frac{5}{6})^{n-x}$ by the number of different orders in which we can get exactly $x$ successes and $n-x$ failures. From Chapter 3 we know that this number of orders is given by the formula

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \quad \text{(from standard formula 4.)}$$

Thus we see that the probability of exactly $x$ 6's in $n$ rolls of a die can be written

$$\binom{n}{x}(\tfrac{1}{6})^x(\tfrac{5}{6})^{n-x} .$$

More generally, if we have a single experiment and an event $A$ in that experiment, if we speak of occurrence of $A$ as <u>success</u> and occurrence of $\bar{A}$ as <u>failure</u>, if the probability of success is $P(A) = p$, and if we make $n$ <u>independent</u> trials of the experiment, then the probability of exactly $x$ successes must be

$$\binom{n}{x}p^x(1-p)^{n-x} .$$

This expression is called the binomial formula. (We know from Chapter 3 that it is closely related to the binomial theorem of of algebra). For convenience, we abbreviate the binomial formula as:

$$b(x;n,p) = \binom{n}{x}p^x(1-p)^{n-x}.$$

It is easy to show, from the binomial theorem, that

$$b(0;n,p) + b(1;n,p) + \ldots + b(n;n,p) = 1.$$

Our new over-all experiment (in which we make $n$ independent trials of the original single experiment and look at the number of occurrences of a chosen event $A$ of the original experiment) is called a binomial experiment. Any single experiment can be made part of a larger binomial experiment of length $n$, if we choose an event $A$ to call success, if we can make $n$ independent repetitions of the given experiment, and if we take, as outcome, the number of successes observed.

More generally, for a single experiment and for a chosen event $A$ of that experiment (where we call the occurrence of $A$ a success), a sequence of repeated independent trials is called a sequence of Bernoulli trials. We use the more specific phrase binomial experiment when we are interested only in the number of successes observed. If we are interested not just in the number of successes but also in the order in which successes and failures occur, we no longer use the phrase binomial experiment, but instead use the more general phrase Bernoulli trials. The question which asks for the probability of at least 7 successes in 10 trials is a question about a binomial experiment of length 10. The question which asks for the probability of having an uninterrupted <u>run</u> of at least 7 successes in 10 trials is a question about Bernoulli trials. To put it another way, when we speak of a <u>binomial experiment</u> of length $n$, we <u>may</u>, if we wish, take the set $\{0,1,\ldots,n\}$ of possible numbers of successes as our sample space; and when we speak of a <u>Bernoulli trials</u> experiment of length $n$, we <u>must</u> take the set $\{SS\ldots S, SS\ldots SF, SS\ldots SFS, \ldots\}$ of possible patterns of success and failure in $n$ trials as our sample space. The first sample space has $n+1$ points; the second has $2^n$ points.

The binomial formula is one of the most useful formulas in all of probability theory. Here is an example. Suppose that each battery for a radio has probability 0.2 of operating for more than 300 hours. If we test 10 different batteries, what is the probability that <u>exactly</u> $x$ of them operate for more than 300 hours? The answer is, of course, given by the binomial formula

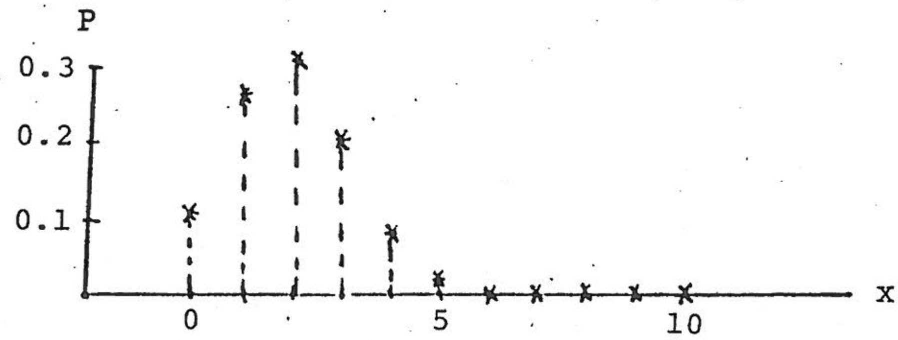
$$b(x;10,0.2) = \binom{10}{x}(0.2)^x(0.8)^{10-x} .$$

Carrying out the calculation, we obtain the following table.

| x: | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| P(<u>exactly x exceed 300 hours</u>): (= b(x;10,0.2)) | 0.107 | 0.268 | 0.302 | 0.201 | 0.088 | 0.026 | 0.006 | 0.001 | 0.000 | 0.000 | 0.000 |

where the final three probabilities are zero out to three decimal places.

Such a table of values for a binomial formula (with fixed n and p) is called a <u>binomial distribution</u>. If we graph the above distribution, we get




<u>Figure 5.1</u>

This graph, with a high point and then decreasing values on either side, is typical of binomial distributions. The only exceptions are that in certain cases there may be two adjacent points of highest (and equal) probability value, rather than a single highest point, and that in certain cases the highest point (or two highest points) may occur at the left side of the graph (if p is small enough) or at the right side (if p is close enough to 1.). The <u>most probable value</u> of x (the value of x for the highest point) gives a value of $\frac{x}{n}$ close to p. (In the case of Figure 5.1, we get $\frac{x}{n} = \frac{2}{10}$ and $p = 0.2$). This is always true of binomial distributions, as we shall see. (Recall that $\frac{x}{n}$ is the observed relative frequency of success in a binomial experiment.)

When we consider a binomial experiment of length n, there are, as we have already indicated, two possible and obvious choices of sample space. We can use the underlying sample space of all possible sequences of success and failure ($2^n$ sample points), or we can use the set of possible values for the total number of successes ($n+1$ sample points). Each point $x$ in the second space represents an event in the first space (an event consisting of $\binom{n}{x}$ points from the first space.) To get a probability function for either of these sample spaces, we must choose a value for $p$, the probability of success in a single trial. Then the probability function for the first space assigns the value $p^x(1-p)^{n-x}$ to each point having exactly $x$ successes, and the probability function for the second space assigns the value $b(x;n,p)$ to the point $x$. Thus we see that a binomial distribution is a probability function for this second choice of sample space. Indeed, we shall usually refer to the resulting probability space itself as a "binomial distribution", since the table for a binomial distribution fully describes that space.

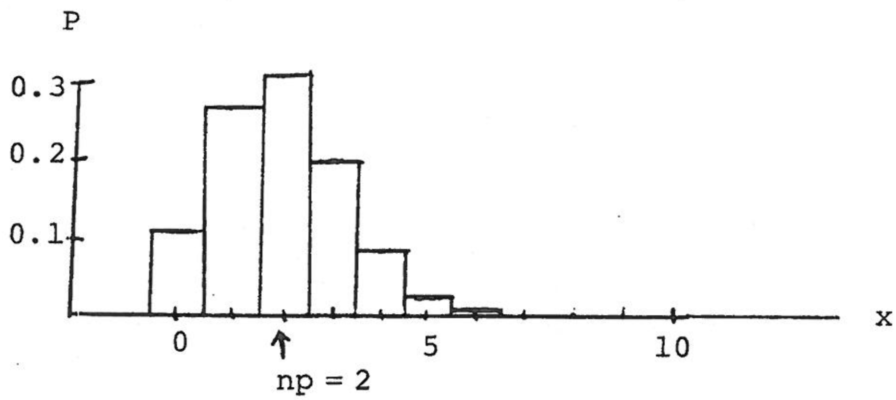
When we have a probability space whose sample space is a set of numbers, we shall often refer to that probability space as a "distribution". (We shall later refer to such a probability space as a "random variable".) We shall also use the word "distribution" to refer to the probability function (or probability density function in the case of a continuous probability space) of that probability space.

Since a binomial distribution can be looked upon as a probability space, we can use it to calculate the probability of any given event in that probability space. Thus, in the example above for testing 10 batteries, the probability that at least 2 but fewer than 6 of the 10 batteries operate for more than 300 hours would be given by:

$$P(2 \leqslant x \leqslant 5) = P(x = 2) + P(x = 3) + P(x = 4) + P(x = 5)$$

$$= 0.302 + 0.201 + 0.088 + 0.026 = 0.617$$

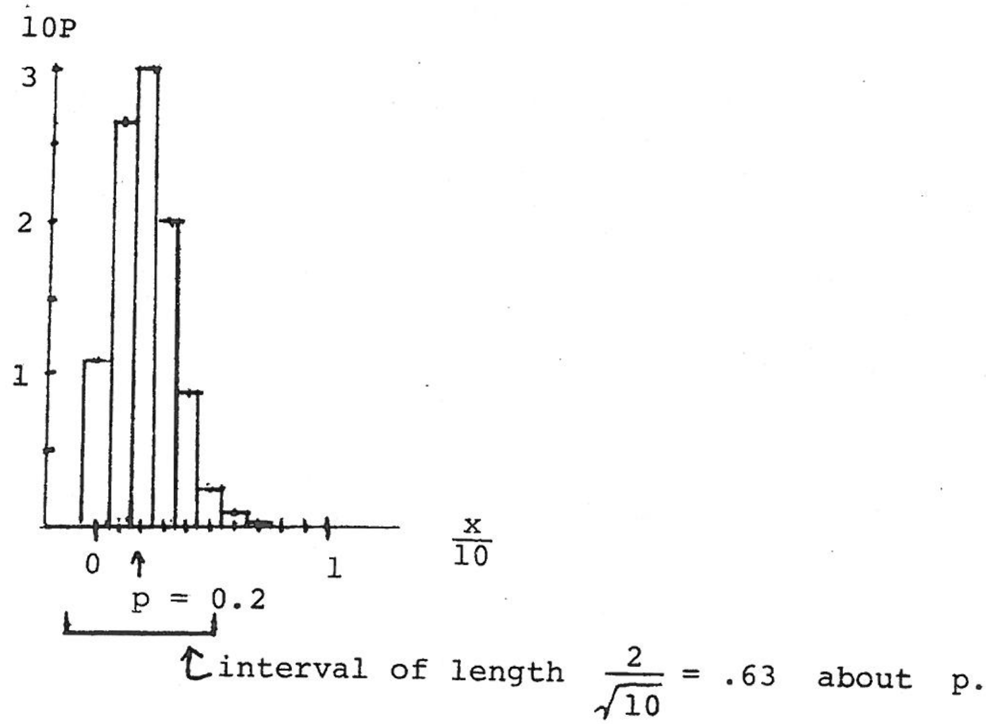
The graphical information in Figure 5.1 can be given in several other different ways. In Figure 5.2, we replace each vertical line by a bar. Since the width of each bar is 1 unit, the <u>area</u> of each bar gives the probability for that $x$ value, and the total area of all the bars must be 1. (In the figure, <u>area</u> of a bar means width in horizontal units multiplied by height in vertical units.)




<u>Figure 5.2</u>

Note that in this figure, for example, $P(2 \leqslant x \leqslant 5)$ is the area of the graph between $x = 2 - \frac{1}{2} = 1.5$ and $x = 5 + \frac{1}{2} = 5.5$. We shall speak of a graph of the Figure 5.2 kind as a <u>type A</u> graph for a binomial distribution.

In Figure 5.3, we contract the graph of Figure 5.2 horizontally by using values of $\frac{x}{n}$, rather than values of $x$, along the horizontal axis, and by centering each bar at the corresponding value of $\frac{x}{n}$. The width of each bar is now $\frac{1}{n}$. At the same time, we lengthen each bar so that its area remains equal to its probability. (Since each bar's width is changed from 1 to $\frac{1}{n}$, its length must be multiplied by $n$ when we go from Figure 5.2 to Figure 5.3.) Thus the total area of the graph remains 1. (In Figure 5.3 below, units of length along the two axes are changed to make the figure fit more easily onto the page.)




<u>Figure 5.3</u>

Note that in Figure 5.3, $P(2 \leqslant x \leqslant 5) = P(\frac{2}{10} \leqslant \frac{x}{n} \leqslant \frac{5}{10})$ is the area of the graph between $\frac{x}{n} = \frac{2 - 1/2}{10} = 0.15$ and $\frac{x}{n} = \frac{5 + 1/2}{10} = 0.55$. We shall speak of a graph of the Figure 5.3 kind as a <u>type B</u> graph for a binomial distribution.

What happens to graphs of the kind given in Figures 5.2 and 5.3 if larger values of $n$ are used while the value of p remains fixed? If one studies this matter in detail, one finds the following. The total area of each graph must remain 1. Graphs of the Figure 5.2 kind (type A) will cover a longer and longer portion of the horizontal axis. At the same time, the height of the highest bar will decrease and be approximately proportional to $\frac{1}{\sqrt{n}}$ (see below). Thus the graph of Figure 5.2 spreads and flattens if we increase n (while keeping p fixed). Also, the position of the highest bar (or bars) will move to the right and be given by a value of x close to np. (In fact, the highest bar will either be the bar standing over the horizontal value np or it will be next to the bar over np.)

On the other hand, graphs of the Figure 5.3 kind (type B) must remain between the horizontal values 0 and 1 (or, more exactly, between $-\frac{1}{2n}$ and $1+\frac{1}{2n}$). The height of the highest bar will be seen to increase (and to be approximately proportional to $\sqrt{n}$), and the position of the highest bar will be close to the value p on the horizontal axis. Also, we find, as n increases, that almost all of the area of the graph lies above a narrower and narrower interval around the horizontal value p. This is in agreement with the observed weak stability of relative frequencies as described in Chapter 1. Indeed, Chapter 1 tells us that nearly all of the area of a type B graph should fall in a horizontal interval of total length $\frac{2}{\sqrt{n}}$ about the value p. (For the case of the particular graph given in Figure 5.3 this interval is marked at the bottom.) In Chapter 6, we shall see that this does in fact occur, and hence that the theory of binomial experiments thus agrees with the empirically observed square root of $n$ law.

Thus, in summary, the graph of Figure 5.3 becomes higher, narrower, and more spiked as we increase $n$ (while keeping $p$ fixed), while the location of the highest bar remains near $p$.

Remark. If we have repeated trials, but they are not independent, then the binomial formula cannot be used. For example, if we are in a country where every day is either sunny or rainy, we can make repeated trials, on successive days, of the experiment: observe the weather for one day. As we saw in Example 2 of Chapter 4, these trials may not be independent. (In that example, sunny days had a greater tendency to follow sunny days, and rainy days to follow rainy days.) As we noted in Chapter 4, a probability space for repeated trials which may not be independent is sometimes called a stochastic process. Bernoulli trials give a special case of stochastic process where the trials are independent.

Notation. For the binomial distribution, the letters $n$ and $p$ are traditionally used as above for the total number of trials and for the probability of success on a single trial. The letter $q$ is also often used for the probability of failure on a single trial. Thus $q = 1-p$, and $b(x;n,p)$ can be written

$$\binom{n}{x} p^x q^{n-x} .$$

Note on proof. The fact that the most probable value of $x$ is close to $np$ can be seen from the ratio

$$\frac{b(x;n,p)}{b(x+1;n,p)} = \frac{n!(x+1)!(n-x-1)!p^x q^{n-x}}{x!(n-x)!n!\ p^{x+1}q^{n-x-1}} = \frac{(x+1)q}{(n-x)p}$$

This ratio is less than 1 when $x < np - q$ and greater than 1 when $x > np - q$. Hence the distribution increases until $x$ passes $np - q$, and then decreases. (If $x = np - q$, then the probability values at $x$ and $x+1$ are equal and maximum. If $np - q < x < np - q + 1$ $(= np + p)$, then the probability value at $x$ is a single maximum.)

The fact that in a type A graph the height of the highest bar (near $x = np$) is approximately proportional to $1/\sqrt{n}$ can be proved by using Stirling's formula to evaluate the factorials in the binomial formula. This gives

$$\frac{n!}{x!(n-x)!}\ p^x q^{n-x} \approx \frac{n^{n+\frac{1}{2}}e^{-n}\ \sqrt{2\pi}\ p^x q^{n-x}}{x^{x+1/2}e^{-x}\sqrt{2\pi}\,(n-x)^{(n-x+1/2)}e^{-(n-x)}\sqrt{2\pi}}\ .$$

When we substitute $x = np$ (and $n - x = n - np = nq$) and cancel factors, this yields

$$\frac{1}{\sqrt{2\pi npq}} = \frac{1}{\sqrt{2\pi pq}} \cdot \frac{1}{\sqrt{n}}\ .$$

<u>de Méré's paradox</u>. If a die is rolled 4 times, what is the probability that a <u>six</u> occurs <u>at least once</u>? Let $x$ = the number of occurrences. Using the binomial formula, we immediately get $P(x \geq 1) = 1 - P(x = 0) = 1 - (5/6)^4 = 0.518$ (for a binomial experiment with $n = 4$ and $p = 1/6$). If a pair of dice is rolled 24 times, what is the probability that a <u>twelve</u> occurs <u>at least once</u>? By a similar calculation, for a binomial experiment with $n = 24$ and $p = 1/36$, we get $P(x \geq 1) = 1 - (35/36)^{24} = 0.491$.

In the early days of probability theory, the fact that these two results differ was a surprise to gamblers, who argued that because the ratio of 4 to 6 is the same as the ratio of 24 to 36, it ought to follow that the two probabilities be the same. We see, from our calculations, that such an argument is false. Indeed, one gambler (de Méré) had already noted from observed relative frequencies that the first probability appeared to lie above 0.5 while the second appeared to fall below 0.5. Students in probability should be wary of apparently simple arguments, like the false argument above, which are not based on theory.

The gamblers' wrong conclusion is correct in one respect. If we take $p = \frac{1}{m}$ and $n = \frac{2}{3}m$ as above, we get

$$P(x \geqslant 1) = 1 - \left(\frac{m-1}{m}\right)^{(2/3)m} = 1 - \left(1 - \frac{1}{m}\right)^{m(2/3)} .$$

If we let m increase, then, by a familiar result from calculus, $P(x \geqslant 1)$ approaches the limiting value $1 - e^{-2/3} = 0.487$.

Conceptual remark. The reader should again note the special role that independence plays in getting the probability space for a binomial experiment. If we did not assume independence of the trials, then many different choices of probability function would be possible. Our knowledge of the physical situation tells us that independence should hold. Assuming independence as a mathematical fact about the probability space, along with using the same p for each trial, then enables us to find a unique and correct probability function.

## EXERCISES FOR CHAPTER 5.

5-1. (a) What is the probability of getting exactly eight heads in 20 tosses of a fair coin?

(b) What is the probability of getting exactly two sixes in twelve rolls of a fair die?

5-2. What is the probability of getting fewer than 3 elevens in ten rolls of two dice?

5-3. Two thirds of the students in a large class have drivers' licenses. Four students are selected at random to go on a field trip. Give an approximate value for the probability that at least two are licensed drivers.

5-4. A four-engine airplane will complete a certain trip provided that at least two of its engines continue to operate for the full trip. Otherwise it will not complete the trip. Assume that for each engine, the probability that the engine will fail during the trip is 0.1, and that failure of an engine is independent of failure of the other engines. What is the probability that the airplane will not complete the trip?

5-5. (a) A baseball player's batting average is .300. We understand this to mean that on the average he gets three base-hits in every ten official times at bat. What is the probability that he gets exactly 2 hits in

4 times at bat? What is the probability that he gets at least 2 hits in 4 times at bat?

(b) Are there any reasons for believing that the number of hits in four times at bat does not follow a binomial distribution?

(c) Assume that the same player will have four times at bat in each of 10 successive games. What is the probability that he gets at least one hit in every game?

(d) What would the probability in (c) be for a player whose batting average is .400? For a player whose batting average is .500?

5-6. An examination has 10 multiple choice questions. Each question has 5 alternative answers. A student answers each question by making a random guess. What is the probability that the student gets 7 or more right?

5-7. A certain operation for cancer patients has a five-year survival rate of 80 percent. That is to say, about four patients in five survive for five years or more from the time of the operation.

(a) What is the probability that exactly four of the next five patients operated on survive for five years or more?

(b) What is the probability that exactly eight of the next ten patients operated on survive for five years or more?

5-8. (a) A binomial experiment consists of 3 trials and the probability of exactly 2 successes is 12 times as great as the probability of exactly 3 successes. Find p, the probability of success on a single trial.

(b) In a certain sequence of Bernoulli trials, the probability of 3 successes in 5 trials is 2/3 of the probability of 3 successes in 6 trials. Find p, the probability of success in a single trial.

5-9. Assume that one penny in every 100,000,000 minted has two heads. The rest are normal. A coin is chosen at random and tossed 20 times. The coin comes up <u>heads</u> on each of the 20 tosses. What is the probability that it has two heads?

5-10. A box contains a fair coin and a biased coin. The biased coin comes up heads 75 percent of the time when it is tossed. A coin is chosen at random and tossed four times. It comes up heads on three of those tosses. What is the probability that it is the biased coin?