Binomial Probability Models

By the end of this set of lessons, you should be able to do the following:

1. Identify Bernoulli Trials and Binomial Process.
2. Calculate related probabilities for a binomial process.

Background

You should like this process! Even if you don’t, I know you need to appreciate its usefulness in business and science. Let’s haul out that time tested probability chestnut to start the thinking.

What is the probability that a family with three children will have two boys?

We have all the tools to handle this problem. We can list the sample space:

<table>
<thead>
<tr>
<th>bbb</th>
<th>bbg</th>
<th>bgb</th>
<th>bgg</th>
</tr>
</thead>
<tbody>
<tr>
<td>gbb</td>
<td>gbg</td>
<td>ggb</td>
<td>ggg</td>
</tr>
</tbody>
</table>

We can count the successful results (purely a matter of opinion). There are three outcomes that have two boys. So the probability is 3 in 8.

Now let's try to handle this:

What is the probability that a family with ten children will have two boys?

I would prefer not to list the sample space. It has $2^{10} = 1024$ possible outcomes! We need a different strategy.

Let's go back to the first, simpler problem and develop a “better” way. “Better” because it will handle larger sample spaces.

Observe that our problem has only two possible outcomes ($b$ or $g$) at each step. Also notice that the probabilities step by step do not change. They are independent. The $p(b) = 0.5$ at each birth. We can use that to our advantage.

What we want to do now is to construct the simplest outcome that meets our criteria, two boys. The outcome is $bbg$. The probability for that outcome is calculated using independence to the right.

Now we notice that this is very restrictive. We can place the girl first, second or last. So there are three ways to place the girl. Then there must be three ways to place the two boys also.

We recall from counting that we can calculate how many ways using combinations as $\binom{3}{2}$. We combine the basic probability calculation with the number of ways the factors can be combined to get our final result.

Let's try this process on the big problem, two boys in 10 children. First we see that the $p(b) = 0.5$ at each birth. Second we notice that we still only have two possibilities ($b$ or $g$) as before. Finally, we know that independence still applies.

| $p(bbg) = p(b)p(b)p(g) = \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)$ |
| $p(bbg) = p(b)^2 p(g) = \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)$ |

$$p(2b, 1g) = 3 \binom{2}{1} p(b)^2 p(g) = \frac{3!}{2!1!} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)$$
Our simplest successful outcome puts the two boys first: b b g g g g g g g g. Then we calculate its probability as before:

\[ p(bb\text{gggggggg}) = p(b)p(b)p(g)p(g)p(g)p(g)p(g)p(g)p(g) = \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^8 \]

I took a little observational shortcut to get to the bottom line. I also know that I have restricted my thinking to only one successful outcome, boys first. The boys could be in any of the ten positions. So just as before, we have to think of all combinations that include exactly two boys, There are \( \binom{10}{2} \) possibilities.

The total probability is\[ p(2b\rightarrow 8g) = \binom{10}{2} p(b)^2 p(g)^8 = \binom{10}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^8 = \frac{45}{1024}. \]

Let's flip this problem. What is the probability that this family with ten children would have exactly eight girls and two boy children? We can rearrange factors to see that it is

\[ p(8g \text{ and } 2b) = \binom{10}{8} p(b)^2 p(g)^8 = \binom{10}{8} \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^2 = \frac{45}{1024}. \]

Examining the three calculations we have done, a formula pops up very quickly:

\[ p(m \text{ successes out of } n \text{ trials}) = \binom{n}{m} p(\text{success})^m p(\text{failure})^{n-m} \]

This formula is presented in various textbooks as \( b(n,m,p) = \binom{n}{m} p^m (1 - p)^{n-m} \) where \( p \) is probability of success for each of \( m \) trials and we want exactly \( n \) successes. Then, since there are only two complementary possibilities, the probability of failure is \((1 - p)\).

The situation described above is usually called a Bernoulli Trial. The model is called a Binomial Probability model.

---

**Bernoulli Trial Defined**

Any time a single process creates the same outcome space of mutually exclusive outcomes, then a series of the processes (trials) is a Bernoulli experiment. The formula previously stated does an excellent job of simplifying our calculations.

The boy/girl problems are Bernoulli trials. Regardless of the number of children under consideration, we expect to see only boys or girls and, at least for our simplification, the probability of a boy is one-half. As you will see in a more realistic boy/girl Bernoulli trial process, that probability is actually a simplification to introduce the concept.
Binomial Probability Models

More Examples.

Example 1: A teacher gives a ten-question true/false test. What is the probability that a student will score exactly 80% by purely guessing the correct response?

Let's stop to think about what we consider success here. It is not true or false. It is a correct or incorrect problem. We don't know for a fact that all the answers are true. Since we are guessing, our probability of guessing correctly is 50%. We either guess right or guess wrong, two equally likely options.

The problem is a paraphrase of the conditions for the \( g, b \) problem! Now though, \( g \) is good (correct) and \( b \) is bad (incorrect) the calculation is

\[
p(8g \text{ and } 2b) = {10 \choose 8} p(b)^2 p(g)^8 = {10 \choose 8} \left( \frac{1}{2} \right)^8 \left( \frac{1}{2} \right)^2 = \frac{45}{1024}.
\]

Let's change the question.

Example 2: Assume 70% is passing. What is the probability that by purely guessing on a ten-problem T/F quiz, a student will pass the test?

The student passes with 7 or 8 or 9 or 10 correct responses.

The passing situations are mutually exclusive. That allows us to add the individual probabilities because there is no possibility of an overlap in the probabilities. The calculation is tedious, but it is simple.

\[
p(\text{pass}) = b(10, 7, .5) + b(10, 8, .5) + b(10, 9, .5) + b(10, 10, .5)
\]

Example 3: A teacher gives a 5-question, multiple-choice test. Each problem has exactly one correct choice from among four possibilities. Assume 70% is passing. What is the probability that a student, purely by guessing, can score a passing grade?

A grim reality, to pass at the 70% level you must get either 4 or 5 correct. The problem is in the binomial world. For each problem, the probability of guessing correctly is 1 in 4. The probability of guessing incorrectly is 3 in 4. So, looking at the formal \( n = 5 \), \( p = 0.25 \) and \( m \) is 4 or 5.

Then

\[
p(\text{pass}) = b(5, 4, .25) + b(5, 5, .25).
\]

Approximating Non-Bernoulli Processes

The next example is definitely not a Bernoulli process, but we can get away with using the formula because the probability changes are so small as to be unnoticeable.

Example 4: A company produces widgets. They have found over time that after producing billions of widgets, the defect rate is about 0.0003%. What is the probability that in selecting 25 widgets from their production line at random that they would find exactly three defective widgets?

As previously stated, this is not Bernoulli. The defect rate changes from selection to selection since it is without replacement. However the change is so minuscule that we can discount it. The answer is

\[
b(25, 3, 0.000003) = _{25}C_3 (0.000003)^3 (1 - 0.000003)^{25-3} \approx 6.2096 \times 10^{-14}
\]

This is so close to zero that if the company did get three bad widgets out of 25, they should shut down the line and re-calibrate the equipment! In fact, that is why they might do this kind of sampling process, to check the condition of their production process.
Binomial Probability Models

Example 5: A student has prepared well for a 25-question T/F test. In practicing for the test, he has successfully answered all T/F questions asked of him 95% of the time. What is the probability he will a score less than 90% on the test?

We have a success rate, \( p = 0.95 \). We need for our student to miss more than two problems to score less than 90%. So we calculate the complementary condition: He misses 2 or 1 or none. Then we subtract from 1.

\[
p(\text{score} < 90\%) = 1 - \left[ b(25,0,0.05) + b(25,1,0.05) + b(25,2,0.05) \right]
\]

Example 6: Let’s go back to the first problem about boys and girls. The probability of a boy is not 50%. Read the extract below:

Sex Ratios, Family Size, and Birth Order.


Biggar, Robert J. 1; Wohlfahrt, Jan 2; Westergaard, Tine 2; Melbye, Mads 2

Abstract:

In many countries, the male:female ratio at birth has varied significantly over the past century, but the reasons for these changes have been unclear. The authors observed a close parallel between decreasing family size and declining male:female sex ratio in Denmark from 1960 to 1994. To explain this finding, they examined the sex ratio and birth order of 1,403,021 children born to 700,030 couples. Overall, 51.2% of the first births were male. However, families with boys were significantly more likely than expected to have another boy (biologic heterogeneity). By the fourth birth to families with three prior boys, 52.4% were male. The increase varied directly with the number of prior boys (p for trend = 0.0007). Furthermore, couples with boys were more likely to continue to have children. In summary, the authors found that the declining male:female ratio in Denmark and probably other European populations is mainly attributable to three effects: declining family size, biologic heterogeneity, and child sex preference. Why families with boys are more likely to have additional boys is unknown.

Copyright 1999 by The Johns Hopkins University School of Hygiene and Public Health, Baltimore, Maryland, USA. All rights reserved.

Now let’s use the data to improve our calculation. We will not take it to the extreme though. Technically we should apply the conditional process to get the “most correct” answer. Let’s just assume that the probability of male birth is 51.5% . Calculate the probability of having three girls in three births.

Just “thinking it through” the answer is \((1 - 0.515)^3 \approx 0.112\).

Applying a formulaic approach, \( b(3,0,0.515) = \binom{3}{0} 0.515^0 (1 - 0.515)^3 \)

I chose to use the \( p = 0.515 \) and zero boys approach. The result vary from the original uniform of \( \frac{1}{6} \) result by a little more than 10%.

Your next lesson is on cumulative distributions. While this is introduced in the Bernoulli Trials section, it is not dependant on the process. From there we will move into the continuous random variable processes.