1 Normal Distribution.

1.1 Introduction

A Bernoulli trial is simple random experiment that ends in success or failure. A Bernoulli trial can be used to make a new random experiment by repeating the Bernoulli trial and recording the number of successes. Now repeating a Bernoulli trial a large number of times has an irritating side effect. Suppose we take a tossed die and look for a 3 to come up, but we do this 6000 times. This is a Bernoulli trial with a probability of success of \( \frac{1}{6} \) repeated 6000 times. What is the probability that we will see exactly 1000 success?

This is definitely the most likely possible outcome, 1000 successes out of 6000 tries. But it is still very unlikely that an particular experiment like this will turn out so exactly. In fact, if 6000 tosses did produce exactly 1000 successes, that would be rather suspicious. The probability of exactly 1000 successes in 6000 tries almost does not need to be calculated. Whatever the probability of this, it will be very close to zero. It is probably too small a probability to be of any practical use. It turns out that it is not all that bad, 0.014. Still this is small enough that it means that have even a chance of seeing it actually happen, we would need to repeat the full experiment as many as 100 times. All told, 600,000 tosses of a die.

When we repeat a Bernoulli trial a large number of times, it is unlikely that we will be interested in a specific number of successes, and much more likely that we will be interested in the event that the number of successes lies within a range of possibilities. Instead of looking for the probability that exactly 1000 successes occur, we are more likely interested in the number of success being around 1000, say off by no more than 15 in either direction. What is the probability that the number of successes is between 9985 and 1015? It turns out that this is a more reasonable 0.41. Unfortunately to do this calculation precisely, you need a pretty powerful computer or calculator.

It would seem that the calculations needed to deal with a Bernoulli trial repeated a large number of times would make such probability models too difficult to deal with. However, we noticed something about the graphs of these probability models. When we take any probability \( p \) and \( n \) large enough, all the graphs of the probability models look very similar. As it turns out, they are more than similar, they are all basically the same. There is one universal probability distribution that will approximate all repeated Bernoulli experiments with enough repetitions. What this means is:

**Remark 1** In a repeated Bernoulli experiment with enough repetitions, there is an easy method of calculating the probabilities of events of the form "The number of successes is less than or equal to \( k_0 \)," "The number of successes is at most \( k_1 \) and at least \( k_0 \)," and "The number of successes is greater than or equal
to \( k_0 \). That is to say, we can quickly compute

\[
\begin{align*}
p(k & \leq k_0) \\
p(k_0 & \leq k \leq k_1) \\
p(k & \geq k_1)
\end{align*}
\]

for any success-probability \( p \) when \( n \) is any number large enough.

The trick to this shortcut is to use one common scale on all our probability models. To find this common scale we look back at our measures of the center of a data set and the variation in that data set. We compute a mean value for our probability model, an variance for the model, and a standard deviation for the model. It turns out that there is one table of values that will approximate all sorts of tables obtained from Bernoulli trials with success probability \( p \) and large enough \( n \). That table, however, gives probabilities in terms of standard deviations from the mean. To use it in any particular example, we must convert the example probability model to a scale of standard deviations from the mean.

### 1.2 The Mean, Variance and Standard Deviation of a Repeated Bernoulli Trial

Our first step is to consider the mean, variance and standard deviation of a repeated Bernoulli trial. Suppose we have a Bernoulli trial with probability of success \( p_0 \) and repeat it \( n \) times. We first commute the mean, variance and standard deviation of its probability model as a set of weighted data. We will then convert the probability model tabulated in terms of the number of successes to a model tabulated by standard deviations from the mean. The result will look weird at first, but it is the key to making later calculations much easier.

As usual we start with a small \( n \) and use a very nice probability of success

1.2.1 Example

For example, start with the simplest example: a Bernoulli trial with probability of success \( \frac{1}{2} \) and \( n = 2 \). Thus the probability model we obtain is

<table>
<thead>
<tr>
<th># Success</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(#) )</td>
<td>( \frac{1}{4} )</td>
<td>( \frac{1}{2} )</td>
<td>( \frac{1}{4} )</td>
</tr>
</tbody>
</table>

The mean of this is computed by multiplying each number of success by its probability and taking the sum. We keep with tradition and use the Greek letter \( \mu \), \( \mu \), to stand for this mean.

\[
\mu = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1.
\]

We compute the variance by multiplying the square of the distance from the mean by the probability of each number of success. Again tradition dictates
we denote this by $\sigma^2$. Then

$$\sigma^2 = (0 - 1)^2 \cdot \frac{1}{4} + (1 - 1)^2 \cdot \frac{1}{2} + (2 - 1)^2 \cdot \frac{1}{4} = \frac{1}{2}.$$ 

Once we have a variance, we bring it back to normal units by taking the square root. The result is the standard deviation, and we denote it with the Greek letter sigma, $\sigma$.

$$\sigma = \frac{1}{\sqrt{2}} \approx 0.70711.$$ 

We convert each number of success to a scale based on standard deviations from the mean. Thus 0 successes is $-1$ success from the mean of 1. Each standard deviation measures as 0.70711 successes. So $-1$ success is $\frac{-1}{0.70711} = -1.4142$ standard deviations. We start our chart as

<table>
<thead>
<tr>
<th># Success</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Deviations</td>
<td>$-1.4142$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p(#)$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{4}$</td>
</tr>
</tbody>
</table>

Next 1 success is 0 successes from the mean of 1. Each standard deviation measures as 0.70711 successes. So 0 success is 0 standard deviations. We continue our chart as

<table>
<thead>
<tr>
<th># Success</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Deviations</td>
<td>$-1.4142$</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$p(#)$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{4}$</td>
</tr>
</tbody>
</table>

Then 2 successes is +1 successes from the mean of 1. Each standard deviation measures as 0.70711 successes. So +1 success is $\frac{1}{0.70711} = 1.4142$ standard deviations. We finish our chart with

<table>
<thead>
<tr>
<th># Success</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Deviations</td>
<td>$-1.4142$</td>
<td>0</td>
<td>$1.4142$</td>
</tr>
<tr>
<td>$p_{\sigma}(#)$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{4}$</td>
</tr>
</tbody>
</table>

Notice that the formula that converts the number of success $k_0$ into $x_0$ the number of standard deviations from the mean is

$$x_0 = \frac{k_0 - \mu}{\sigma}.$$ 

All we do next, is drop the old scale and keep the new scale

| $x_0$ | $-1.4142$ | 0 | $1.4142$ |
| $p_{\sigma}(x = x_0)$ | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ |
1.2.2 Example

Just to be thorough, we will do one more small example explicitly; \( n = 3 \). Thus the probability model we obtain is

<table>
<thead>
<tr>
<th># Success</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(#) )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{3}{8} )</td>
<td>( \frac{3}{8} )</td>
<td>( \frac{1}{8} )</td>
</tr>
</tbody>
</table>

We compute the mean as before:

\[
\mu_3 = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = \frac{3}{2}.
\]

We compute the variance by multiplying the square of the distance from the mean by the probability of each number of success:

\[
\sigma_3^2 = (0 - \frac{3}{2})^2 \cdot \frac{1}{8} + (1 - \frac{3}{2})^2 \cdot \frac{3}{8} + (2 - \frac{3}{2})^2 \cdot \frac{3}{8} + (0 - \frac{3}{2})^2 \cdot \frac{1}{8} = \frac{9}{32} + \frac{3}{32} + \frac{3}{32} + \frac{3}{32} = \frac{24}{32} = \frac{3}{4}.
\]

Once we have a variance, we also have the standard deviation:

\[
\sigma_3 = \frac{\sqrt{3}}{2} \approx 0.86603.
\]

We convert each number of success to a scale based on standard deviations from the mean. Thus 0 successes is \(-\frac{3}{2}\) success from the mean of \(\frac{3}{2}\). Each standard deviation measures as 0.86603 successes. So \(-\frac{3}{2} = -1.5\) success is \(-1.732\) standard deviations. Next 1 success is \(-\frac{1}{2}\) success from the mean of \(\frac{3}{2}\). So \(-\frac{1}{2} = -0.5\) success is \(-0.57735\) standard deviations. Then 2 successes is \(\frac{1}{2}\) success from the mean of \(\frac{3}{2}\). So \(\frac{1}{2} = 0.5\) success is \(0.57735\) standard deviations. And finally 3 successes is \(\frac{3}{2}\) success from the mean of \(\frac{3}{2}\). So \(\frac{3}{2} = 1.5\) success is \(-1.732\) standard deviations.

<table>
<thead>
<tr>
<th># Success</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Deviations</td>
<td>-1.732</td>
<td>-0.57735</td>
<td>0.57735</td>
<td>-1.732</td>
</tr>
<tr>
<td>( p_3(#) )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{3}{8} )</td>
<td>( \frac{1}{8} )</td>
</tr>
</tbody>
</table>

Cleaning up, we get

\[
\begin{array}{c|cccc}
\hline
x_0 & p_3(x = x_0) & 1.732 & -0.57735 & 0.57735 & -1.732 \\
\hline
\frac{1}{2} & \frac{1}{8} & \frac{1}{8} & \frac{3}{8} & \frac{1}{8} \\
\end{array}
\]

where \( x \) is the number of standard deviations from the mean.

These are just examples of a small number of repetitions while we are interested cases where there are a large number of repetitions. It would seem that calculating the mean, variance and standard deviations for a probability with a large number of values would make the steps above too lengthy to carry out. Fortunately, we have a shortcut.
Theorem 2 Suppose we have a Bernoulli trial with a probability of success given by \( p \). Suppose the trial is repeated \( n \) times and the number of successes counted. Then the mean, variance and standard deviation of the probability model for the various numbers of success is given by

\[
\begin{align*}
\mu_n &= n \cdot p \\
\sigma_n^2 &= n \cdot p \cdot (1 - p) \\
\sigma_n &= \sqrt{n \cdot p \cdot (1 - p)}
\end{align*}
\]

1.3 Cumulative Probability Tables

The next step is finding a shortcut for finding probabilities in repeated Bernoulli experiments is to change the way we tabulate the probabilities that we know. Up to now, we have been giving our tables by listing the probabilities for each possible number of successes one at a time. Thus our last table was

<table>
<thead>
<tr>
<th>( x_0 )</th>
<th>-1.732</th>
<th>-0.57735</th>
<th>0.57735</th>
<th>1.732</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_3(x = x_0) )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{3}{8} )</td>
<td>( \frac{3}{8} )</td>
<td>( \frac{1}{8} )</td>
</tr>
</tbody>
</table>

This tells us that

\[
\begin{align*}
p(x &= -1.732) = \frac{1}{8}; \\
p(x &= -0.57735) = \frac{3}{8}; \\
p(x &= 0.57735) = \frac{3}{8}; \\
p(x &= 1.732) = \frac{1}{8};
\end{align*}
\]

From this we can create an "accumulation table." In an accumulation table, we give the values of the probabilities \( p(x \leq x_0) \) in stead of the values of \( p(x = x_0) \). Thus

<table>
<thead>
<tr>
<th>( x_0 )</th>
<th>-1.732</th>
<th>-0.57735</th>
<th>0.57735</th>
<th>1.732</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_3(x \leq x_0) )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{3}{8} )</td>
<td>( \frac{3}{8} )</td>
<td>( \frac{1}{8} )</td>
</tr>
</tbody>
</table>

Thus the individual probabilities accumulate as the \( x \) grows larger. This table is

<table>
<thead>
<tr>
<th>( x_0 )</th>
<th>-1.732</th>
<th>-0.57735</th>
<th>0.57735</th>
<th>1.732</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_3(x \leq x_0) )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{3}{8} )</td>
<td>( \frac{3}{8} )</td>
<td>1</td>
</tr>
</tbody>
</table>

The last probability in the table is 1 because all of the events listed are included in the event \( x \leq 1.732 \). The probability that something happens is 1.

At this point we can convert the probabilities to decimals and tilt the table on its side.

<table>
<thead>
<tr>
<th>( x_0 )</th>
<th>( p_3(x \leq x_0) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.732</td>
<td>0.025</td>
</tr>
<tr>
<td>-0.57735</td>
<td>0.500</td>
</tr>
<tr>
<td>0.57735</td>
<td>0.975</td>
</tr>
<tr>
<td>1.732</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Consider a larger problem: the case where a fair coin is tossed 50 times, and each time a heads occurs, that is recorded as a success. We can draw a graph of the probability model for this:

Most of the success numbers in this probability model have such a low probability of success that they do not register on this scale. We can trim the chart down to

We need a complete table of the probability model for this experiment. The table could just give us the probabilities to say, three digits accuracy, and that will actually help keep the table manageable. We have already noticed that many the probabilities, those further out from the center, will be so small that they will round to 0.000. In fact, all the probabilities for the number of successes below 14 would round off to 0.000. As would all the probabilities above 36.
We can compute the nonzero part of the table

<table>
<thead>
<tr>
<th>Successes</th>
<th>Probability</th>
<th>Successes</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>0.001</td>
<td>25</td>
<td>0.112</td>
</tr>
<tr>
<td>15</td>
<td>0.002</td>
<td>26</td>
<td>0.108</td>
</tr>
<tr>
<td>16</td>
<td>0.004</td>
<td>27</td>
<td>0.096</td>
</tr>
<tr>
<td>17</td>
<td>0.009</td>
<td>28</td>
<td>0.079</td>
</tr>
<tr>
<td>18</td>
<td>0.016</td>
<td>29</td>
<td>0.060</td>
</tr>
<tr>
<td>19</td>
<td>0.027</td>
<td>30</td>
<td>0.042</td>
</tr>
<tr>
<td>20</td>
<td>0.042</td>
<td>31</td>
<td>0.027</td>
</tr>
<tr>
<td>21</td>
<td>0.060</td>
<td>32</td>
<td>0.016</td>
</tr>
<tr>
<td>22</td>
<td>0.079</td>
<td>33</td>
<td>0.009</td>
</tr>
<tr>
<td>23</td>
<td>0.096</td>
<td>34</td>
<td>0.004</td>
</tr>
<tr>
<td>24</td>
<td>0.108</td>
<td>35</td>
<td>0.002</td>
</tr>
<tr>
<td>25</td>
<td>0.112</td>
<td>36</td>
<td>0.001</td>
</tr>
</tbody>
</table>

In this experiment we can use our theorem to compute the mean, variance, and standard deviation. The formula are

\[
\begin{align*}
\mu_n & = n \cdot p \\
\sigma^2_n & = n \cdot p \cdot (1 - p) \\
\sigma_n & = \sqrt{n \cdot p \cdot (1 - p)}
\end{align*}
\]

Thus these formula, we compute the mean:

\[
\mu = np = \frac{50}{2} = 25;
\]

we compute the variance

\[
\sigma^2 = np(1 - p) = \frac{50}{4} = 12.5,
\]

and we compute the standard deviation

\[
\sigma = \sqrt{\sigma^2} = \sqrt{12.5} = 3.5355.
\]

From these we can compute \(x_0\) (the distance from the mean in standard deviations) from \(k_0\) (the actual number of successes) using the formula

\[
x_0 = \frac{k_0 - \mu}{\sigma}.
\]
Eventually the table looks like

<table>
<thead>
<tr>
<th>$x_0$</th>
<th>$p(x = x_0)$</th>
<th>$x$</th>
<th>$p(x = x_0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.11</td>
<td>0.001</td>
<td>0.00</td>
<td>0.112</td>
</tr>
<tr>
<td>-2.83</td>
<td>0.002</td>
<td>0.28</td>
<td>0.108</td>
</tr>
<tr>
<td>-2.55</td>
<td>0.004</td>
<td>0.57</td>
<td>0.096</td>
</tr>
<tr>
<td>-2.26</td>
<td>0.009</td>
<td>0.85</td>
<td>0.079</td>
</tr>
<tr>
<td>-1.98</td>
<td>0.016</td>
<td>1.13</td>
<td>0.060</td>
</tr>
<tr>
<td>-1.70</td>
<td>0.027</td>
<td>1.41</td>
<td>0.042</td>
</tr>
<tr>
<td>-1.41</td>
<td>0.042</td>
<td>1.70</td>
<td>0.027</td>
</tr>
<tr>
<td>-1.13</td>
<td>0.060</td>
<td>1.98</td>
<td>0.016</td>
</tr>
<tr>
<td>-0.85</td>
<td>0.079</td>
<td>2.26</td>
<td>0.009</td>
</tr>
<tr>
<td>-0.57</td>
<td>0.096</td>
<td>2.55</td>
<td>0.004</td>
</tr>
<tr>
<td>-0.28</td>
<td>0.108</td>
<td>2.83</td>
<td>0.002</td>
</tr>
<tr>
<td>0.00</td>
<td>0.112</td>
<td>3.11</td>
<td>0.001</td>
</tr>
</tbody>
</table>

The table above gives the individual probabilities, and we want an accumulation table. That table follows from this using some addition:

<table>
<thead>
<tr>
<th>$x_0$</th>
<th>$p(x \leq x_0)$</th>
<th>$x$</th>
<th>$p(x \leq x_0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.11</td>
<td>0.001</td>
<td>0.00</td>
<td>0.556</td>
</tr>
<tr>
<td>-2.83</td>
<td>0.003</td>
<td>0.28</td>
<td>0.664</td>
</tr>
<tr>
<td>-2.55</td>
<td>0.008</td>
<td>0.57</td>
<td>0.760</td>
</tr>
<tr>
<td>-2.26</td>
<td>0.016</td>
<td>0.85</td>
<td>0.839</td>
</tr>
<tr>
<td>-1.98</td>
<td>0.032</td>
<td>1.13</td>
<td>0.899</td>
</tr>
<tr>
<td>-1.70</td>
<td>0.059</td>
<td>1.41</td>
<td>0.944</td>
</tr>
<tr>
<td>-1.41</td>
<td>0.101</td>
<td>1.70</td>
<td>0.960</td>
</tr>
<tr>
<td>-1.13</td>
<td>0.161</td>
<td>1.98</td>
<td>0.984</td>
</tr>
<tr>
<td>-0.85</td>
<td>0.240</td>
<td>2.26</td>
<td>0.992</td>
</tr>
<tr>
<td>-0.57</td>
<td>0.336</td>
<td>2.55</td>
<td>0.997</td>
</tr>
<tr>
<td>-0.28</td>
<td>0.444</td>
<td>2.83</td>
<td>0.999</td>
</tr>
<tr>
<td>0.00</td>
<td>0.556</td>
<td>3.11</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Each entry in this chart is computed from a (more accurate) version of the original.

To see how we might use an accumulation table to compute probabilities, consider the probability that the number of successes is less than or equal to
19. This is the probability in the red part of the chart:

Now 19 successes converts to \( \frac{k - \mu}{\sigma} = \frac{19 - 25}{5/\sqrt{50}} = -1.6971 \approx -1.70 \) standard deviations from the mean. So we want \( p(x \leq -1.6971) \). We can read this probability directly from the table:

\[
p(k \leq 19) = p(x \leq -1.6971) = 0.027.
\]

Next compute the probability that the number of success in 50 tries is less than or equal to 25. The event that the number of successes is 25 or below looks like the red part in

Now 25 successes is the mean, so it is 0 standard deviations from itself. From the accumulation table, we read that \( p(k \leq 25) = p(x \leq 0) = 0.556 \). It should not be surprising that this is just a bit over 0.5. Now 25 success is the mean, and therefore the exact center of the probability model. So it should be close to 0.5, but \( k = 25 \) is included in the event we are looking. Because the center is included, the probability that the number of success is less than or equal to that center should be just a bit more than 0.5.
Another clear probability is $p(k \leq 36)$. That appears as the red part of

There "is" some blue here, it is just too small to be visible. Now 36 is more than 3 standard deviations above the mean. Thus from the chart and the picture $p(k \leq 36) = p(x \leq 311) = 1.00$. This is an approximation accurate to three decimal places.

So far, we have just been reading probabilities directly from the table, but there are other possibilities. Consider the probability that there are 20 or more successes. This time the event we are considering is the blue part of

The table gives the probabilities in the red part, but that still gives us the answer we want. The total of all the colors on the graph is 1, so

$$
p(k \geq 20) = 1 - p(k < 20) = 1 - p(k \leq 19).
$$

$$
p(k \leq 19) = p(x \leq -1.6971) = 0.027.
$$

So the probability we are looking for is $p(k \geq 20) = 0.973$.

This shows us how to use the table to find the probability of an interval. Consider the probability that the number of successes is between 22 and 28
including both. We are looking for the probabilities in the red areas of

But we can find the probability that the number of success is at most 28 from the chart.

$$p(k \leq 28) = 0.839.$$  

We can also find the probability that the number of success is at most 22 there:

$$p(k \leq 21) = 0.161.$$  

The probability we want is

$$p(22 \leq k \leq 28) = p(k \leq 28) - P(k < 22)$$

$$= p(k \leq 28) - P(k \leq 21)$$

$$= 0.839 - 0.161$$

$$= 0.678.$$  

In the graph

That is we found the amount in the red and blue sections, $p(k \leq 28)$ and subtracted the total in the blue sections $p(k \leq 21)$ to get the total in the red alone.
1.4 The Similarities of Probability Models

The example we have been working with is a Bernoulli trial with probability of success \( p = \frac{1}{2} \) repeated \( n = 50 \) times. The methods we developed worked because we calculated a cumulative probability table for this exact case. But what if we change the example a bit? Not that much, as it turns out. The amazing thing that eventually follows from this observation is that there is one universal table of cumulative probabilities that closely approximates the cumulative probability for any probability model that approaches this ideal shape.

Look at the following graphs of experiments where a Bernoulli trial is repeated 50 times, but where the probability of success varies.

\( p = 0.5 \)

\[ \text{Graph for } p = 0.5 \]

\( p = 0.4 \)

\[ \text{Graph for } p = 0.4 \]
$p = 0.3$

$p = 0.2$

$p = 0.1$
The first three of these look very much alike, the later two, not so much. The trouble with the later two is that the mean is getting a little too close to the limit of 0 successes. These graphs are all scaled by the number of successes \( k \). If we switch to scale then by the number of standard deviations from the mean \( x \), the result is quite striking. Rather than keep the long tails of zeros, we will always trim the graphs down to the values of \( x \) where the probabilities are non-zero to 3 decimal places.

\[ p = 0.05 \]

\[ p = 0.5 \]

\[ p = 0.4 \]
$p = 0.3$

$\begin{array}{cccccccccccccccccccc}
-3.39 & -3.09 & -2.78 & -2.47 & -2.16 & -1.85 & -1.54 & -1.23 & -0.93 & -0.62 & -0.31 & 0 & 0.31 & 0.62 & 0.93 & 1.23 & 1.54 & 1.85 & 2.16 & 2.47 & 2.78 & 3.09 & 3.39 \\
0.000 & 0.020 & 0.040 & 0.060 & 0.080 & 0.100 & 0.120 & 0.140
\end{array}$

$p = 0.2$

$\begin{array}{cccccccccccccccccccc}
-2.83 & -2.47 & -2.12 & -1.77 & -1.41 & -1.06 & -0.71 & -0.35 & 0 & 0.35 & 0.71 & 1.06 & 1.41 & 1.77 & 2.12 & 2.47 & 2.83 & 3.18 & 3.54
\end{array}$

$p = 0.1$

$\begin{array}{cccccccccccccccccccc}
-2.36 & -1.89 & -1.41 & -0.94 & -0.47 & 0 & 0.47 & 0.94 & 1.41 & 1.89 & 2.36 & 2.83 & 3.3 & 3.77
\end{array}$
Just like before, the first three of these look very much alike. In the later two, however, the mean is getting too close to the limit of 0 successes, and the graph is getting skewed. We could choose probabilities larger that 0.5, but the results would be the same. The graph would take on one basic shape, but that shape would begin to deteriorate as the mean got closer and closer to the upper limit of 50 successes.

This shape might be familiar. It is the shape of the famous "Bell Curve." As long as the mean stays away from the extreme possibilities, the probability model will approximate this famous curve. If we use the standard deviation scale, we will even see a strong similarity range of the interesting part of the graph.

Just to be thorough we will state this fact formally, and include a scary formula for the bell shaped curve. The formal name for this curve and its formula is "The Normal Distribution."

**Theorem 3** Suppose we have a Bernoulli trial with a probability of success given by \( p \). Suppose the trial is repeated \( n \) times and the number of successes counted. If \( n \) is large enough, then the probability that the number of successes is less than \( s \) standard deviations from the mean is approximated by the area under the curve

\[
y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}
\]
to the left of the value \( x = s \). The graph of this curve looks like

This formula looks pretty formidable, but there is no need to worry. There are plenty of tables, plenty of calculators and plenty of computer programs that give the areas under this curve. A table of the area under this curve up to the value \( x = x_0 \) is an approximate accumulation table for any probability model that resembles the bell shaped curve. If we can find a table of values for the areas of shaded areas like

we have a sort of universal probability accumulation table for lots of \( p = p_0 \) and \( n = n_0 \). All we need is for \( n_0 \) to be large enough for the probability model to take on the right shape.

### 1.5 Applications of the Normal Distribution

So the question is: how can we use this? Well, while the table that lists all the probabilities \( p_{50}(i \leq k) \) is mythical, it is very easy to find very complete tables of \( p_{\infty}(x \leq s) \), where this stands for the area under the curve

\[
y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}
\]

to the left of the value \( x = s \).
The curve is called "the Normal Distribution" or "the Gaussian curve." It is also known colloquially as "the Bell Curve." Tables of the area under this curve may refer to it any of these ways. All such tables come with a picture of the areas they give. Typically the table will take values of $s$ with two decimal digits and give the area under the curve to the left of $x = s$. These tables generally give the data in rows that give the first digits, and columns that give the last. Thus a row that looks like

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.3</td>
<td>.9032</td>
<td>.9049</td>
<td>.9066</td>
<td>.9082</td>
<td>.9099</td>
<td>.9115</td>
<td>.9131</td>
<td>.9147</td>
<td>.9162</td>
<td>.9177</td>
</tr>
</tbody>
</table>

tells you that

$$p_\infty(x \leq 1.32) \simeq 0.9066$$
$$p_\infty(x \leq 1.36) \simeq 0.9131.$$  

**Example 1** Now suppose you wish to use such a table to approximate in 50 repetitions of a fair coin toss, the probability of the event: "The number of successes is between 22 and 28." We do this by completing the following steps.

### 1.5.1 Step 1: Compute the mean and standard deviation

In a fair coin toss the probability of success is $p = \frac{1}{2}$. Since our experiment calls for 50 repetitions, $n = 50$. Thus using our theorem, we compute the mean:

$$\mu = np = \frac{50}{2} = 25;$$

we compute the variance

$$\sigma^2 = np(1 - p) = \frac{50}{4} = 12.5,$$

and the standard deviation

$$\sigma = \sqrt{\sigma^2} = \sqrt{12.5} = 3.5355.$$  

### 1.5.2 Step 2: Convert the event interval into standard deviations

We are interested in the event that the number of successes is between 22 and 28.

$$p_{50}(22 \leq k \leq 28).$$

There is, however, a trick to getting an accurate approximation. We think of the two extremes of the interval as integer approximations of real numbers before we convert to the standard deviations. Thus for a real number to round off to a number between 22 and 28, it could be as small as 21.5 and as large as 28.5. Thus we "unround" the interval of integers $k$ where $22 \leq k \leq 28$ to the
interval of real numbers $x$ where $21.5 \leq x < 28.5$. The real number interval we need to convert to standard deviations is 

$$21.5 \leq x < 28.5$$

Now the left had side of this interval 21.5 is 3.5 successes below the mean, and the right hand side 28 is 3.5 successes above the mean. Now we set up a conversion

$$1 \text{ standard deviation} : 3.5355 \text{ successes.}$$

Thus

$$\frac{1}{3.5355} \text{ standard deviation} : 1 \text{ success.}$$

And so

$$\frac{3.5}{3.5355} \text{ standard deviation} : 3 \text{ successes.}$$

So we can give the interval

- between $-3.5$ and $+3.5$ success from the mean

To

- between $-0.989959$ and $+0.989959$ standard deviations from the mean.

The table we are using has the area under the normal curve given to two decimal degrees of accuracy. So we need to round off to use the table. So I give the interval as

- between $-0.99$ and $+0.99$ standard deviations from the mean.

### 1.5.3 Step 3: Using the table

Our table gives me the probability under the curve with $x \leq 0.99$ as

$$p_\infty(x \leq 0.99) \approx 0.8389$$

We also need the area under the curve with $x \leq -0.99$. (It doesn’t matter if include $-0.99$ in this inequality or not.) In my table we find

$$p_\infty(x \leq -0.99) \approx 0.1611$$

Thus we can approximate the probability we need:

$$p_\infty(-0.85 \leq x \leq 0.85) = p_\infty(x \leq 0.85) - p_\infty(x \leq -0.85) \approx 0.8389 - 0.1611 \approx 0.6778$$

Notice that this approximation is so accurate that it gives the same answer we get by rounding off the result of our direct calculation 0.677764.
This method of estimating probabilities for repeated Bernoulli trials works basically the same way for all probabilities of success \( p \) and all numbers of repeats \( n \) large enough. How large is "large enough?" For \( p \) near \( \frac{1}{2} \) not large at all; for \( p \) closer to 0 or 1, \( n \) might need to be a lot greater. A good rule of thumb is that if both 0 successes and \( n \) success are more than 3 standard deviations \( \sigma \) away from the mean \( \mu \) then any approximation of the probability of an interval will be pretty good. If both 0 successes and \( n \) success are more than 2 standard deviations \( \sigma \) away from the mean \( \mu \) then an approximation will still be good, but it will be better for intervals on the side with more room. If both 0 successes and \( n \) success are less than 2 standard deviations \( \sigma \) away from the mean \( \mu \), then you begin to take a risk using the normal curve to approximate a probability, especially on the "short" side.

**Example 2** Consider tossing one fair die, and counting a 6 as a success where we repeat this 30 times. What is the probability that the number of successes will be between, and including, 4 and 7. We could draw the probability model of this to see how it fits the bell shape. An easier thing to do is see how far the mean is from the extreme edges of 0 and 30 successes.

We compute the mean, variance and standard deviations with \( n = 30 \) and \( p = \frac{1}{6} \):

\[
\mu = np = \frac{30}{6} = 5
\]
\[
\sigma^2 = np(1 - p) = 30 \times \frac{5}{6} = \frac{25}{6}
\]
\[
\sigma = \sqrt{np(1 - p)} \approx 2.0412
\]

Notice that 3 standard deviations is over 6 successes. This means that the lower limit 0 successes is within 3 standard deviations of the mean 5. According to our rule of thumb, it may not be safe to use the normal distribution to approximate probabilities. It is close though, and we will take a chance. Still we must remember that this approximation might be off a bit more than the example above.

Our interval \( 4 \leq k \leq 7 \) is centered in a probability model that is on the margin of our rule of thumb. That probably and only a small part of it is in the squashed side near 0. We might thrown caution to the wind, and try to approximate this using the normal distribution. We do this the same way we did the last time.

**1.5.4 Step 1: Compute the mean and standard deviation**

In our fair die toss the probability of success is \( p = \frac{1}{6} \). Since our experiment calls for 30 repetitions, \( n = 30 \). Thus using our theorem, we compute the mean,
variance and standard deviation:

\[
\begin{align*}
\mu &= np = \frac{30}{6} = 5 \\
\sigma^2 &= np(1 - p) = \frac{30}{6} \cdot \frac{15}{6} = \frac{25}{6} \\
\sigma &= \sqrt{np(1 - p)} \approx 2.0412
\end{align*}
\]

1.5.5 Step 2: Convert the event interval into standard deviations

We are interested in the event that the number of successes is between 4 and 7. We think of the two extremes of the interval as integer approximations of real numbers before we convert to the standard deviations. Thus for a real number to round off to a number between 4 and 7, it could be as small as 3.5 and as large as 7.5. Thus we need to convert the interval of real numbers

\[3.5 \leq x < 7.5\]

to standard deviations. Now the left-hand side of this interval 3.5 is 1.5 successes below the mean of 5, and the right-hand side 7.5 is 2.5 successes above the mean 5. Now we set up a conversion for the low end

\[
1 \text{ standard deviation : } 2.0412 \text{ successes.}
\]

Thus

\[
\frac{1}{2.0412} \text{ standard deviation : } 1 \text{ success.}
\]

And so

\[
\frac{1.5}{2.0412} \text{ standard deviation : } 1.5 \text{ successes}
\]

\[
\text{and } \frac{2.5}{2.0412} \text{ standard deviation : } 2.5 \text{ successes.}
\]

So -1.5 successes from the mean corresponds to -0.73486 standard deviations from the mean. Also 1.5 successes from the mean corresponds to 1.2248 standard deviations from the mean. So we can give the interval

between \(-3.5\) and \(+3.5\) successes from the mean

to

between \(-0.73486\) and \(+1.2248\) success from the mean.

The table I am using has the area under the normal curve given to two decimal degrees of accuracy. So I need to round off to use the table. So I give the interval as

between \(-0.73\) and \(+1.22\) success from the mean.
1.5.6 Step 3: Using the table

My table gives me the probability under the curve with \( x \leq 1.22 \) as

\[ p_\infty(x \leq 1.22) \simeq 0.8888 \]

I also need the area under the curve with \( x \leq -0.73 \). (It doesn't matter if include \(-0.73\) in this inequality or not.) In my table I find

\[ p_\infty(x \leq -0.73) \simeq 0.2327 \]

Thus we can approximate the probability we need:

\[ p_\infty(-0.73 \leq x \leq 1.22) = p_\infty(x \leq 1.22) - p_\infty(x \leq -0.73) \]

\[ \simeq 0.8888 - 0.2327 \]

\[ \simeq 0.6561 \]

If we cared to, we could break out a calculator and find this probability from the formula we used earlier, without using the normal distribution to approximate. This answer turns out to be 0.64669. So even when our rule of thumb cautions us against using a normal distribution approximation, it may still do a reasonably good job.

We will do one last example. A high end casino will have a "single zero" roulette wheel. This is a roulette wheel with 37 slots for the numbers 0, 1, 2, 3, \ldots, 35, 36. The wheel spins and a ball is released so that it will eventually settle into one slot on the wheel. Players bet on the final slot the ball will end in before the ball is released.

**Example 3** One bet a player can make is on the ball ending in a slot with an odd number. If the player bets $1 that the number comes up odd and it does, the player gets his dollar back plus $1 in winnings. (In roulette, the number 0 is neither odd nor even.) If a player makes this bet 200 times, he must win at least 100 times to break even. What is the probability that this will happen?

Now one bet of this type is a Bernoulli trial. The probability of success is \( p = \frac{18}{37} \). This is because there are 37 equally likely numbers in the sample space \( \{0,1,2,3\ldots35,36\} \) since you must include the 0. We are repeating this Bernoulli trial \( n = 200 \) times, and we are interested in the probability

\[ p(k \geq 100) \]

It would take a great deal of effort to compute this probability using the formula for all the values \( k = 100, 101, 102 \ldots 199, 200 \). It is better to use the normal distribution to estimate this. We do it the same way as before.

1.5.7 Step 1: Compute the mean and standard deviation

In our fair die toss the probability of success is \( p = \frac{18}{37} \). Since our experiment calls for 200 repetitions, \( n = 200 \). Thus using our theorem, we compute the
mean, variance and standard deviation:

\[
\begin{align*}
\mu &= np = 200 \cdot \frac{18}{37} = 97.297 \\
\sigma^2 &= np(1 - p) = 200 \cdot \frac{18}{37} \cdot \frac{19}{37} = 49.963 \\
\sigma &= \sqrt{np(1 - p)} \approx 7.0685
\end{align*}
\]

1.5.8 Step 2: Convert the event interval into standard deviations

We are interested in the event that the number of successes is at least 100. We think of the extremes of this interval as integer approximations of a real numbers before we convert to standard deviations. Thus for a real number to round off to a number 100 or larger, it could be as small as 99.5. Thus we need to convert the interval of real numbers

\[99.5 \leq x\]

to standard deviations. Now the left hand side of this interval 99.5 is 2.203 successes above the mean of 97.297. Now we set up a conversion for the low end

\[1 \text{ standard deviation : } 7.0685 \text{ successes.}\]

Thus

\[\frac{1}{7.0685} \text{ standard deviation : } 1 \text{ success.}\]

And so

\[\frac{2.203}{2.0412} \text{ standard deviation : } 2.203 \text{ successes.}\]

So +2.203 successes from the mean corresponds to +1.0793 standard deviations from the mean. So we can give the interval

more that +2.203 successes from the mean
to

more that +1.0793 standard deviations from the mean.

The table I am using has the area under the normal curve given to two decimal degrees of accuracy. So I need to round off to use the table. So I give the interval as

more that +1.08 standard deviations from the mean.

1.5.9 Step 3: Using the table

My table gives me the probability under the curve with \(x \leq 1.08\) as

\[p_\infty(x \leq 1.08) \approx 0.8599\]

23
I also need the area under the complete curve which is 1. Thus we can approximate the probability we need:

\[ p_\infty(1.08 \leq x) = 1 - p_\infty(x \leq 1.08) \]
\[ \simeq 1.0000 - 0.8599 \]
\[ \simeq 0.1401. \]

Thus the player’s probability of walking away from the table without losing any money is approximately 0.14; not very good. That is why another good rule of thumb is that it is better own a casino than to gamble in one.

Prepared by: Daniel Madden and Alyssa Keri: May 2009