REPEATED TRIALS

Suppose you toss a “fair coin” one time. Let $E$ be the event that the coin lands heads. We know from basic counting that $p(E) = \frac{1}{2}$ since $n(E) = 1$ and $n(S) = 2$.

Now suppose we play a game 5 times, where the probability of winning is $p$ and the probability of losing is $q = 1 - p$. We could think of the last example where the game is tossing a coin, and if the coin lands heads then you win and if it lands tails then you lose. In this case $p = q = \frac{1}{2}$. We now ask the following question:
What is the probability of winning the first and third times, and losing the second, fourth, and fifth times you play the game?

Answer: Since the probability of winning the first game is $p$, and the probability of losing the second game is $q$, hence the probability of winning the first game and then losing the second game is the product $pq$. Continuing this reasoning, the probability of winning the first, losing the second, winning the third, losing the fourth and losing the fifth game is $pqpq = p^2q^3$.

More generally, suppose you are going to play the game $n$ times and suppose you have specified $k$ times when you win, hence you know the $n - k$ times when you lose. In our last example $n = 5$ and we specified $k = 2$ times, the first and third times, when you win, hence you lose $5 - 2 = 3$ times: the second, fourth, and fifth games.

The probability of winning those $k$ chosen times and losing the other times is then

$$p^k q^{n-k}.$$
We now ask a more complicated question. Suppose you play a game \( n \) times where the probability of winning is \( p \) and the probability of losing is \( q = 1 - p \). What is the probability of winning exactly \( k \) times?

Answer: We first count the number of ways of choosing exactly which \( k \) times you win the game. This is \( nC_k \). We just showed that \( p^kq^{n-k} \) is the probability of winning \( k \) times in a specified order, hence the probability of winning \( k \) times in \( n \) trials is the product \( nC_kp^kq^{n-k} \), which is the probability of winning \( k \) times in any order. If we denote \( p_n(k) \) as the probability of winning exactly \( k \) times out of \( n \) games, where the probability of winning a game is \( p \) and the probability of losing is \( q = 1 - p \), then

\[
p_n(k) = nC_kp^kq^{n-k}.
\]

Notation: Such repeated trials are also referred to as Bernoulli trials.

EXAMPLES

Problem: Suppose you toss a fair coin 6 times. What is the probability that the coin will land heads exactly 3 times?

Answer: 
\( n = 6, k = 3, p = q = \frac{1}{2} \), so 
\[
p_n(k) = nC_kp^kq^{n-k} = 6C_3(\frac{1}{2})^3(\frac{1}{2})^{6-3} = \frac{5}{16}.
\]

Problem: Suppose you toss a fair coin 9 times. List the probabilities that the coin lands exactly 0 times, exactly 1 times, exactly 2 times, \ldots, and exactly 9 times. Then construct a chart showing these probabilities.

\[
\begin{array}{c|c}
 n & k & p_n(k) = nC_kp^kq^{n-k} \\
\hline
 9 & 0 & 6C_0(\frac{1}{2})^0(\frac{1}{2})^9 = 1/512 \approx .00195 \\
 9 & 1 & 6C_1(\frac{1}{2})^1(\frac{1}{2})^8 = 9/512 \approx .0176 \\
 9 & 2 & 6C_2(\frac{1}{2})^2(\frac{1}{2})^7 = 9/128 \approx .0703 \\
 9 & 3 & 6C_3(\frac{1}{2})^3(\frac{1}{2})^6 = 21/128 \approx .164 \\
 9 & 4 & 6C_4(\frac{1}{2})^4(\frac{1}{2})^5 = 63/256 \approx .246 \\
 9 & 5 & 6C_5(\frac{1}{2})^5(\frac{1}{2})^4 = 63/256 \approx .246 \\
 9 & 6 & 6C_6(\frac{1}{2})^6(\frac{1}{2})^3 = 21/128 \approx .164 \\
 9 & 7 & 6C_7(\frac{1}{2})^7(\frac{1}{2})^2 = 9/128 \approx .0703 \\
 9 & 8 & 6C_8(\frac{1}{2})^8(\frac{1}{2})^1 = 9/512 \approx .0176 \\
 9 & 9 & 6C_9(\frac{1}{2})^9(\frac{1}{2})^0 = 1/512 \approx .00195 \\
\end{array}
\]
Problem:

100 points are placed randomly in the interval $(0, 10)$. What is the probability that exactly 20 of these points are in the interval $(4, 6)$?

Answer:

This can be considered as a problem in repeated trials. The first experiment is to place the first point, the second experiment is to place the second point, etc. Let $p$ be the probability that a single point lands within the subinterval $(4, 6)$. Then $p = 2/10 = 1/5$, the ratio of the lengths of the two intervals, and $q$ is the probability of a failure $= 4/5$. In this problem $n = 100$ and $k = 20$.

$$p_n(k) = p_{100}(20) = nC_k p^k q^{n-k} = 100C_{20}(1/5)^{20}(4/5)^{80} \approx 0.0993$$

More generally, the probability that $k$ points placed randomly in an interval $(a, b)$ lie within a subinterval $(c, d)$ is

$$p_n(k) = nC_k p^k q^{n-k}$$

where $p = \frac{b-a}{c-d}$ and $q = 1 - p$.

In our example, $n = 100$, $k = 20$, $a = 0$, $b = 10$, $c = 4$, and $d = 6$.

Problem:

An apartment building has residents living on the second and third floors. The residents use an elevator to get to their apartments. There are 20 people living on the second floor. The third floor has more luxurious, larger apartments and there are 6 people living on the third floor. Six random residents use the elevator to get to their apartments. What is the probability that exactly 4 of the people exit on the second floor and that 2 people exit on the third floor?

Answer:

We consider each time one of the resident uses the elevator as a separate event. In fact we consider it as playing a game. We consider a resident exiting on the second floor as winning and exiting on the third floor as losing. The probability of winning is $p = \frac{20}{26}$ and the probability of losing is $q = \frac{6}{26} = 1 - p$.

Our original question is equivalent to the following question: What is the probability of winning the game exactly $k = 4$ times when playing the game $n = 6$ times?

We use the formula $p_n(k) = nC_k p^k q^{n-k} = 6C_4(\frac{20}{26})^4(\frac{6}{26})^2 \approx 0.194$. 
Problem:

An apartment building has 11 residents. There are 7 residents on the second floor and 4 residents on the third floor. Six random residents use the elevator to get to their apartments.

a) Make a table listing the probabilities that exactly $k$ of the 6 residents exit on the second floor, for $k = 0, 1, 2, 3, 4, 5, 6$.

b) Make a chart showing the probabilities.

c) Make a graph with a smooth curve through the probabilities.

Answers:

a) Let $p$ be the probability that a resident exits on the second floor. Then $p = \frac{7}{11}$ since the number of ways of choosing a resident who lives on the second floor is 7 and the total number of ways of choosing a resident is 11. Then $q = 1 - p = \frac{4}{11}$. We consider the first resident using the elevator as the first trial, the second resident using the elevator as the second trial, etc. In this problem the number of trials $n = 6$. We then compute $p_n(k) = \binom{n}{k} p^k q^{n-k}$ with $n = 6$ and $k = 0, 1, 2, 3, 4, 5, 6$. and list the results in a table:

<table>
<thead>
<tr>
<th>n</th>
<th>k</th>
<th>$p_n(k) = \binom{n}{k} p^k q^{n-k}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0</td>
<td>$\binom{6}{0}(\frac{7}{11})^0(\frac{4}{11})^6 \approx 0.00231$</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>$\binom{6}{1}(\frac{7}{11})^1(\frac{4}{11})^5 \approx 0.0243$</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>$\binom{6}{2}(\frac{7}{11})^2(\frac{4}{11})^4 \approx 0.106$</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>$\binom{6}{3}(\frac{7}{11})^3(\frac{4}{11})^3 \approx 0.248$</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>$\binom{6}{4}(\frac{7}{11})^4(\frac{4}{11})^2 \approx 0.325$</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>$\binom{6}{5}(\frac{7}{11})^5(\frac{4}{11})^1 \approx 0.228$</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>$\binom{6}{6}(\frac{7}{11})^6(\frac{4}{11})^0 \approx 0.0664$</td>
</tr>
</tbody>
</table>

b) Enter the probabilities .00231, .0243, .106, .248, .325, .228, .0664 in to a column in an Excel spreadsheet and use the chart wizard to create a column chart. Change the labels and get a chart representing the probabilities. Click here to see such a chart.

c) Enter the probabilities .00231, .0243, .106, .248, .325, .228, .0664 in to a column in an Excel spreadsheet and use the chart wizard to create a smooth curve (look under Custom Graphs and select Smooth Lines.) Change the labels and get a smooth curve through the probabilities. Click here to see such a chart.
Problem:

An apartment building has 11 residents. There are 7 residents on the second floor and 4 residents on the third floor. Six people use the elevator at random to go to their apartments. What is the probability that at most 3 of the residents exit on the second floor?

Answer:
We use the notation \( p(0 \leq k \leq 3) \) to stand for the probability the number \( k \) of the 6 residents who exit on the second floor is between 0 and 3. The events that there are exactly \( k \) people who exit on the second floor are mutually exclusive, as \( k = 0, 1, 2, 3 \), so

\[
p(0 \leq k \leq 3) = p_0(0)+p_0(1)+p_0(2)+p_0(3) = 6C_0p^0q^6+6C_1p^1q^5+6C_0p^2q^4+6C_0p^3q^2
\]

In the above, \( p \) is the probability of a success (a resident exits on the second floor) = 7/11 and \( q \) is the probability of a failure (a resident exits on the third floor ) = 4/11. You can use your calculator or computer to verify that

\[
p(0 \leq k \leq 3) \approx .381
\]

In general, if the probability of a success is \( p \) and the probability of a failure is \( q \) (where \( p + q = 1 \)), and if the experiment is performed \( n \) times, the probability that the number \( k \) of successes is between \( k_1 \) and \( k_2 \),

\[
p_n(k_1 \leq k \leq k_2) = nC_{k_1}p^{k_1}q^{n-k_1} + \cdots + nC_{k_2}p^{k_2}q^{n-k_2}.
\]

If you are familiar with the summation notation, the formula may be written more compactly as:

\[
p_n(k_1 \leq k \leq k_2) = \sum_{k=k_1}^{k=k_2} nC_kp^kq^{n-k}.
\]

The notation means let \( k = k_1, k = k_1 + 1, \ldots, k = k_2 \) in the expression \( nC_kp^kq^{n-k} \) and add those terms.

In our previous example, \( n = 6, p = 7/11, q = 4/11, k_1 = 0, k_2 = 3 \).
A more useful problem:

A company manufactures memory chips. The probability that a randomly selected memory chip, after it is shipped to a customer, is defective equals 0.015. An order of 1000 memory chips is received. What is the probability that the total number of defective memory chips received does not exceed 12?

Answer:

In this example the probability that a memory chip is defective \( p = 0.015 \) and \( q = 1 - p = 0.985 \). Also, \( n = 1000 \), \( k_1 = 0 \) and \( k_2 = 12 \). Using our formula for the probability that the number of occurrences \( k \) is between \( k_1 \) and \( k_2 \),

\[
p_{1000}(0 \leq k \leq 12) = \sum_{k=0}^{k=12} \binom{1000}{k}(0.015)^k(0.985)^{1000-k}.
\]

If you use your calculator or computer, you can verify that the answer is approximately 0.27.

THE BINOMIAL THEOREM

We return briefly to the the formula \( p_n(k) = \binom{n}{k}p^kq^{n-k} \). Recall that \( p \) was the probability that the event \( E \) is a success and \( q \) is the probability that the event \( E \) is a failure. For example, in our last example about elevators, \( p = \frac{7}{11} \), \( q = \frac{4}{11} \), and \( \binom{n}{k} \) is the number of ways to choose which \( k \) of the \( n \) times there is a success. If you ask in how many ways can it happen that exactly \( k \) of the residents exit on the second floor, there are \( \binom{n}{k} \) ways to choose which \( k \) times a resident exits on the second floor. After that, for each such time there are 7 ways to choose a resident who exits on the second floor (we are allowing repetitions). So there are \( 7^k \) ways to choose those residents, and since the times that the residents exit on the second floor have already been chose, the remaining \( n - k \) times are determined for those residents who exit on the third floor, and there are \( 4^{n-k} \) ways to choose those residents. So there are \( \binom{n}{k}7^k4^{n-k} \) ways it can happen that exactly \( k \) residents exit on the second floor for \( n \) elevator rides. More generally if there are \( x \) residents on the second floor and \( y \) residents on the third floor, there are \( \binom{n}{k}x^ky^{n-k} \) ways it can happen that exactly \( k \) of the residents exit on the second floor for \( n \) elevator rides.
We now ask the following question: Suppose there are \( x \) people living on the second floor and \( y \) people living on the third floor. On a particular day, \( n \) people ride the elevator and exit on either the second or third floor. In how many ways can this be done?

Answer:

First way of counting:

There are \( x + y \) residents and so there are \( x + y \) ways to choose the first rider, and since repetitions are allowed, there are \( x + y \) ways to choose the second rider, etc. So in all there are 
\[
(x + y)^n
\]
ways to choose the \( n \) elevator users.

Second way of counting: We count the number of ways that all \( n \) of the riders can exit on the second floor, then we count the number of ways that \( n - 1 \) riders can exit on the second floor and hence that 1 rider exits on the third floor, etc., and finally the number of ways that 0 riders exit on the second floor and all riders exit on the third floor. These are mutually exclusive events, so we can find the total number of ways by taking the sum:
\[
\binom{n}{0} x^0 y^n + \binom{n}{1} x^1 y^{n-1} + \cdots + \binom{n}{n} x^n y^0.
\]
Using our summation notation, the number of ways the \( n \) residents can choose to ride the elevator is
\[
\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}
\]
You might notice that this is similar to the formula given earlier for \( p_n(k_1 \leq k \leq k_n) \) with \( k_1 = 0 \) and \( k_2 = n \) (allowing all possible values of \( k \)) except that the \( p_n(k_1 \leq k \leq k_n) \) represents probabilities, hence the terms \( p \) and \( q \) include denominators.

The same quantity computed two different ways give us the famous Binomial Theorem
\[
(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.
\]
FINAL OBSERVATIONS:

In the examples, you will notice that when you graph the probabilities \( p_n(k) = \binom{n}{k} p^k q^{n-k} \) either using a bar (column) chart or a smooth curve through the points \( \{(k, p_n(k))\} \), that when \( p = q = 1/2 \) then you get a bell-shaped curve which is symmetric about a vertical line through the highest point. Such a curve is called a normal distribution. On the other hand, when \( p \) and \( q \) are not equal, then the curve becomes distorted, or skewed, and it is still called a distribution, but it is not a normal distribution. Finally, we proved the Binomial Theorem only for the case that \( x \) and \( y \) are integers. It is easy to extend the formula to rational numbers and then using a little analysis (beyond the scope of this course) to all numbers.

END OF LECTURE
P6(k) for k=0,1,2,3,4,5,6
Smooth curve through $P_6(k)$