Topic 11: Conditional Probability and Independence

The question of the hot hand is really a question about conditional probability, namely "does the probability of success change for the player depending on what has happened in the previous trial of the experiment?". In this section we look at conditional probability and independence. We will use these concepts to extend our models from the previous section to basketball players with a probability of success greater than or less than 0.5. You can find more details on conditional probability and independence in [**?**]

1. Events

Since we are often interested in probabilities of outcomes or events which are not just simple outcomes of our sample space, we need to consider how to calculate probabilities of such events.

Definition 1.1. An Event *E is a subset or sub collection of the simple outcomes of the sample space S.*

All subsets of the sample space describe an event including the empty set(the event that nothing happens). Events with a single simple outcome correspond to the simple outcomes of the sample space. The entire sample space also qualifies as an event; it is the event that something happens. Given a verbal description of an event, the corresponding subset is given by all outcomes which fit the verbal description. When speaking about general rules of calculating probabilities for events, we usually name the events using capital letters A, B, C, D etc....

Example 1.1. *The experiment of rolling a six sided die and observing the number on the uppermost face has a sample space given by*

$$S = \{1, 2, 3, 4, 5, 6\}.$$

Find the subsets corresponding to the following verbal description of the given events:

(a) E = the event that the number is even.

(b) F = the event that the number is larger than two.

We say that an event **occurs** on a given trial of the experiment if the trial yields an outcome in the event set.

1.1. Probability of an Event.

To calculate the probability of an event A, denoted by $P(A)$, we add the probabilities of all of the simple outcomes in the event A.

Example 1.2. *If I roll a fair six sided die and observe the number on the uppermost face, then all faces are equally likely to appear. The probability distribution for the experiment is given in the table below*

Outcome	Probability
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Let O denote the event that the number on the uppermost face is odd. Then O corresponds to the subset of the sample space given by $O = \{1, 3, 5\}$. Then $P(O) = P(1) + P(3) + P(5) = 3/6$.

Example 1.3. *If I roll a fair six sided die twice and observe the number on the uppermost face for both rolls. The list of possible outcomes is shown below as pairs of numbers listed in the order in which they were observed.*

$$\text{Sample Space:} = \begin{array}{cccccc} \{(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6)\} \end{array}$$

We reason that all of these outcomes are equally likely, since the die is symmetric and all sides are equally likely to appear on the uppermost face. Since there are 36 equally likely simple outcomes in the sample space, we assign a probability of $1/36$ to each.

(a) Let E be the event that the numbers observed add to 7.

$$\begin{array}{c} \text{E} \\ \text{(in red)} \end{array} = \begin{array}{cccccc} \{(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & \color{red}{(1,6)} \\ (2,1) & (2,2) & (2,3) & (2,4) & \color{red}{(2,5)} & (2,6) \\ (3,1) & (3,2) & (3,3) & \color{red}{(3,4)} & (3,5) & (3,6) \\ (4,1) & (4,2) & \color{red}{(4,3)} & (4,4) & (4,5) & (4,6) \\ (5,1) & \color{red}{(5,2)} & (5,3) & (5,4) & (5,5) & (5,6) \\ \color{red}{(6,1)} & (6,2) & (6,3) & (6,4) & (6,5) & (6,6)\} \end{array}$$

What is the probability of E, $P(E)$?

(b) Let F be the event that the numbers observed add to 11. List the elements of the set F and calculate $P(F)$.

> **Equally likely outcomes** For an event A, we let $n(A)$ denote the number of outcomes in A. If all outcomes in a sample space are equally likely, then $P(A) = n(A)/n(S)$ where $n(S)$ is the number of outcomes in the sample space.

1.2. **Intersection of events.** Given two events A and B, we can define a new event $A \cap B$ which is the event that both A and B occur. The simple outcomes in the set corresponding to $A \cap B$ is the set of all simple outcomes which are in both events.

Example 1.4. *If our experiment is to roll a fair six sided die twice and observe the number on the uppermost face for both rolls. Let*
A be the event that the number on the first roll of the die is a 6
and let
B is the event that the number on the second roll of the die is a 6.
The outcomes in A are shown in red on the left below and the outcomes in B are shown in blue on the right.

$$\begin{array}{cccccc} \{(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ \color{red}{(6,1)} & \color{red}{(6,2)} & \color{red}{(6,3)} & \color{red}{(6,4)} & \color{red}{(6,5)} & \color{red}{(6,6)}\} \end{array} \qquad \begin{array}{cccccc} \{(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & \color{blue}{(1,6)} \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & \color{blue}{(2,6)} \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & \color{blue}{(3,6)} \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & \color{blue}{(4,6)} \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & \color{blue}{(5,6)} \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & \color{blue}{(6,6)}\} \end{array}$$

The event $A \cap B$ corresponds to the set of all outcomes which belong to both sets. Thus we see that $A \cap B = \{(6,6)\}$, since $(6,6)$ is the only outcome which belongs to both sets. Therefore $P(A \cap B) = 1/36$.

2. Conditional Probability

Sometimes our computation of the probability of an event is changed by the knowledge that a related event has occurred (or is guaranteed to occur) or by some additional conditions imposed on the experiment. We see some examples below:

For Example: In a previous example, we estimated that the probability that LeBron James will make his next attempted field goal in a major league game is 0.496. We used the proportion of field goals made out of field goals attempted (FG%) throughout his career to estimate this probability. If we look at `the statistics from season to season` and some of the `split statistics for the 2014/2015 season` for LeBron James shown below, we see that the FG% changes from season to season and also when calculated under specified conditions such as home games vs. away games. For example the FG% for games played after 2 days rest in the 2014/2015 season(to date) is 0.52 which is higher than the overall FG% for the season which is 0.489.

Regular Season | Postseason

REGULAR SEASON AVERAGES

SEASON	TEAM	GP	GS	MIN	FGM-A	FG%
'03-'04	CLE	79	79	39.5	7.9-18.9	.417
'04-'05	CLE	80	80	42.4	9.9-21.1	.472
'05-'06	CLE	79	79	42.5	11.1-23.1	.480
'06-'07	CLE	78	78	40.9	9.9-20.8	.476
'07-'08	CLE	75	74	40.4	10.6-21.9	.484
'08-'09	CLE	81	81	37.7	9.7-19.9	.489
'09-'10	CLE	76	76	39.0	10.1-20.1	.503
'10-'11	MIA	79	79	38.8	9.6-18.8	.510
'11-'12	MIA	62	62	37.5	10.0-18.9	.531
'12-'13	MIA	76	76	37.9	10.1-17.8	.565
'13-'14	MIA	77	77	37.7	10.0-17.6	.567
'14-'15	CLE	45	45	36.6	9.2-18.7	.489
Career		**887**	**886**	**39.4**	**9.9-19.9**	**.496**

Season: 2014-2015

2014-2015 PER GAME SPLITS

SPLIT	GP	MIN	FGM-FGA	FG%	3
Total	**45**	**36.6**	**9.2-18.7**	**.489**	
Home	26	35.2	8.2-17.1	.479	
Road	19	38.6	10.5-20.9	.500	
vs. Division	8	36.5	9.9-22.0	.449	
vs. Conference	27	36.5	9.1-19.0	.479	
0 Days Rest	11	38.7	10.8-21.3	.509	
1 Days Rest	20	35.9	8.2-17.9	.458	
2 Days Rest	10	35.3	9.3-17.9	.520	
3+ Days Rest	4	39.0	9.0-18.0	.500	
By Month	**GP**	**MIN**	**FGM-FGA**	**FG%**	**3**
October	2	42.5	9.5-22.5	.422	
November	13	36.7	8.5-17.6	.485	
December	14	37.6	9.0-17.9	.502	
January	10	36.3	10.6-20.8	.510	
February	6	33.0	8.3-18.3	.455	
Pre All-Star	45	36.6	9.2-18.7	.489	
By Result	**GP**	**MIN**	**FGM-FGA**	**FG%**	**3**

It may be more appropriate to estimate the probability that LeBron James' next field goal attempt in the current season (2014/2015) will be successful using the field goal percentage for the current season.

$$0.489 = \frac{\text{Field goals made in the 2014/2015 season}}{\text{Field goals attempted in the 2014/2015 season}}$$

Here we are estimating the probability that LeBron will make the field goal **given** the extra information that the attempt will be made in the current season. This is referred to as a **conditional probability**, because we have some extra or **prior information** about conditions under which the experiment will be performed.

If we know that the next field goal attempt will be made after 2 days rest, it might be appropriate to use the field goal percentage for games played in the current season after 2 days rest which is 0.52. (Note the data used for this statistic is from 10 games with an average of 17.9 shots per game, giving us a data set of sufficeint size to make such an estimate). Here we are estimating a **conditional probability given that the underlying conditions of the experiment are that the next shot will be taken in the 2014/2015 season in a game after 2 days rest.**

Definition 2.1. *If A and B are events in a sample space S, with $P(B) \neq 0$, the* **conditional probability** *that an event A will occur, given that the event B has occurred (or is guaranteed to occur) is denoted by*

$$P(A|B).$$

We can interpret this symbol as the proportion of times that we would expect the event A to happen if we were to repeat many trials of the experiment with the underlying condition that B is guaranteed to happen on each trial.

2.1. Calculating conditional probability.

2.1.1. *Reasoned Probability.* If we have used logic to create **a sample space for an experiment which has equally likely outcomes**, then we can calculate the probability that A will occur given that B is guaranteed to occur by restricting our attention to elements of the sample space where B is guaranteed to occur. We can then calculate the probability of A given that B is guaranteed to occur as the proportion of the outcomes in the set corresponding to B for which A also occurs. In other words when we have **equally likely outcomes in our sample space**, we get

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

In general the formula

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

remains true for all sample spaces, whether the outcomes are equally likely or not. This formula is reasonably intuitive as we will see below.

Example 2.1. *Let us consider the experiment from above where we roll a fair six sided die twice and observe the number on the uppermost face for both rolls.*

Let
A be the event that the number on the first roll of the die is a 6
B be the event that the number on the second roll of the die is a 6
and let C be the event that the sum of the numbers on both rolls of the dice is greater than or equal to 11.
We have $P(A) = 1/6 = P(B)$ and $P(C) = 3/36 = 1/12$.
Suppose now that we performed the first roll of the die and we see that the number on the uppermost face is a 6. In other words, we know that the event A has already happened.

(a) Suppose that we have not yet rolled the die a second time and we would like to calculate the probability that B will happen given that A has already happened, $P(B|A)$. That is we wish to calculate the probability that we will roll a six on the second given the prior information that we got a six on the first.

$$\begin{matrix} \{(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ \color{red}{(6,1)} & \color{red}{(6,2)} & \color{red}{(6,3)} & \color{red}{(6,4)} & \color{red}{(6,5)} & \underline{\color{red}{(6,6)}}\} \end{matrix}$$

The six outcomes in A are shown in red. We know that when the experiment is completed, the result will be one of those outcomes and all are equally likely. One of of these six outcomes exactly one is in the event B giving us that $P(B|A) = \frac{n(B \cap A)}{n(A)} = 1/6$.

(b) Suppose now that we have not yet rolled the die a second time and we would like to calculate the probability that C will happen given that A has already happened, $P(C|A)$. That is we wish to calculate the probability that the sum of the numbers on both rolls will be greater than or equal to 11 given the prior information that we got a six on the first.

$$P(C|A) = \frac{n(C \cap A)}{n(A)} = \_\_\_\_.$$

$$\begin{array}{llllll}
\{(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\
(2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\
(3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\
(4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\
(5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\
(6,1) & (6,2) & (6,3) & (6,4) & \underline{(6,5)} & \underline{(6,6)}\}
\end{array}$$

2.1.2. *Estimates from Data.* When we are **estimating conditional probabilities using relative frequencies**, as we did in the basketball example above, we can estimate $P(A|B)$ by restricting our attention to trials of the experiment for which B had occurred (or for which B was an underlying condition) and calculate the proportion of those trials for which A also occurred (or the relative frequency of A among the trials in which B occurred). The resulting formula is the same as above:

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

where $n(A \cap B)$ and $n(B)$ denotes the number of trials in the data where the events $A \cap B$ and B occurred respectively.

Activity 1: The Hot Hand and Conditional Probability Our initial question about the hot hand can be translated to a question about conditional probability. Consider a string of data of baskets and misses resulting from a (large) number of consecutive shots taken by a basketball player. For any given shot, let the event BP(basket previous) denote the event that the previous shot resulted in a basket. For any given shot, let B denote the event that the given shot results in a basket. Our question about the hot hand in this case would translate to "Is $P(B|BP) > P(B)$?"

We can estimate $P(B|BP)$ from the data as long as there are sufficiently many shots in the sample where the event BP occurs and if we are estimating $P(B|BP)$ from the string of data, it seems appropriate to estimate $P(B)$ from the same set of data.

Lets look at the following string of data showing consecutive shots collected from 8 consecutive games for Dion Waiters in regular season basketball games. We have colored the shots in the event PB in red.

MBBMBMBMB (9 shots)
MMMMMMBM (8 shots)
MBMBBMMMBB (10 shots)
BMMMMM (6 shots)
BMBMBMMMBMMBBM (14 shots)
BMBMBMBMBBBMBBM (15 shots)
MBBMMMBMMMMMMM (14 shots)
MMMMBBBMMM (10 shots)
MMBMBBMMMMMB (12 shots)
BMBMMBMMMBBMMMM (15 shots)
MMMMBBBM (8 shots)
BBMMBMMMMBMM (12 shots)
MMMBMMMMM (9 shots)
BBBBBMBBM (9 shots)
MBBMBMMBBMMMMBBB (16 shots)
MBMMMBMMMMBMBBMB (16 shots)
BBBBMMMMBMMBMB (14 shots)

MMMBMMMMM (9 shots)

(a) Estimate $P(B)$ from the entire data set by calculating the relative frequency of B in the data set.

(b) Estimate $P(B|BP)$ from the data by calculating the relative frequency of B in the red outcomes (the trials in which BP occurred) in the data set.

$$P(B|BP) \approx \frac{n(B \cap BP)}{n(BP)} =$$

---

3. Independence

Note that sometimes extra underlying conditions for an experiment (or the fact that another event has happened or is guaranteed to happen) change the probability of an event and sometimes they do not. In other words, sometimes $P(A|B) = P(A)$ and sometimes $P(A|B) \neq P(A)$.

Definition 3.1. *Two events A and B (where $P(B) \neq 0$) are said to be* **independent** *if $P(A|B) = P(A)$. This amounts to saying that the fact that B has happened or is guaranteed to happen does not change the probability that A will happen.*

Since $P(A|B) = \frac{P(A\cap B))}{P(B)}$ (assuming that $P(B) \neq 0$), saying that A and B are independent is equivalent to saying that $P(A) = \frac{P(A\cap B))}{P(B)}$ which in turn is equivalent to saying that

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

From this we can also deduce that if $P(A|B) = P(A)$, we must have $P(A \cap B) = P(A)P(B)$ and $P(B) = \frac{P(A\cap B)}{P(A)} = P(B|A)$. Thus if any three of the formulas below hold true for two events A and B, we can conclude that the events A **and** B **are independent** and vice versa: if the events are independent, we can conclude that all three formulas hold true.

$$\boxed{P(A|B) = P(A), \qquad P(A \cap B) = P(A)P(B), \qquad P(B|A) = P(B).}$$

Example 3.1. *Let us consider the experiment from above where we roll a fair six sided die twice and observe the number on the uppermost face for both rolls.*

Let
A be the event that the number on the first roll of the die is a 6,
B be the event that the number on the second roll of the die is a 6
and let C be the event that the sum of the numbers on both rolls of the dice is greater than or equal to 11. We have $P(A) = 1/6 = P(B)$ and $P(C) = 3/36 = 1/12$.

(a) Are A and B independent events?

(b) Are A and C independent events?

The Hot Hand The question about the hot hand can also be translated to a question about independence of events. Consider a string of data of baskets and misses resulting from a (large) number of consecutive shots taken by a basketball player. For any given shot, let the event BP(basket previous) denote the event that the previous shot resulted in a basket. For any given shot, let B denote the event that the given shot results in a basket. Our question about the hot hand in this case would translate to "Are B and BP independent events (and if not is $P(B|BP) > P(B)$)." (Note that it may be true that B and BP are not independent events and $P(B|BP) < P(B)$ for some players.)

Example 3.2. *Using your estimates from the data set for consecutive shots taken by Dion Waiters above, decide whether the events B and BP are independent for Dion Waiters.*

3.1. **Many Independent Events.** Several events $E_1, E_2, \ldots, E_n$ are independent if $P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}) = P(E_{i_1}) \cdot P(E_{i_2}) \cdot \cdots \cdot P(E_{i_k})$ for any subset $\{i_1, i_2, \ldots, i_k\}$ of $\{1, 2, \ldots, n\}$. In the cases of independent events we can multiply probabilities:
If $E_1, E_2, \ldots, E_n$ are independent events then

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_1) \cdot P(E_2) \cdot \cdots \cdot P(E_n).$$

Example 3.3. Independent trials of an experiment. *If I roll a die three times, I do not expect any combination of events that might happen on the first two rolls of the die to affect the probability of any event on that might happen on the third roll (I am using logic to deduce independence here). Thus I can use the formula above to calculate the probability that I get a "6" on the first roll, a "6" on the second roll and something less than 6 on the third roll.*

Example 3.4. Independent trials of an experiment. *If I have an unfair coin where the probability of heads is (.7) on each flip and the probability of tails is (.3) on each flip, what is the probability that I get the sequence HHHT if I flip the coin 4 times?*

Definition 3.2. *Experiments with only two outcomes called success (s) and failure (f) where the probability of success is p are called Bernoulli experiments. Note that since there are only two outcomes in the sample space, we must have that the probability of failure (f) on each trial is given by $P(f) = 1 - p$. Each trial of a Bernoulli experiment is called a Bernoulli Trial.*

4. Sequences of Baskets (B) and Misses(M) generated randomly with $P(B) \neq 1/2$.

In the last section we learned what distribution of runs we should expect in a sequence generated randomly by flipping a fair coin many times and used our results to compare our expectation to data for basketball players with a 50% chance of making every shot. We would naturally expect that the distribution of runs in a sequence of heads and tails generated by an unfair coin would differ from that of a fair coin. In this section we will learn what to expect in a randomly generated sequence of success' and failures generated from independent Bernoulli trials when the probability of success is not necessarily equal to 1/2. We will also determine what the expected length of the longest run of successes should be in such a sequence and compare it to some real data.

Example 4.1. *Using the* `applet`. *from the University of Alabama in Huntsville introduced in the last section, we simulated a sequence of 30 shots by a basketball player for whom the probability of getting a basket on each shot is* 0.75*, by setting* $p = 0.75$*,* $n = 1$ *and pressing the play button 30 times. The sequence is written below. Check if the longest run of baskets is longer than the longest run of heads expected if I flipped a fair coin 30 times.*

H H H H H T H H T H T H T H H H H T T T H H H H H H H H H T

4.1. Modelling Runs of Success'.

Consider now a basketball player who has a probability p of making a basket each time he/she takes a shot. (Note that we are implicitly assuming that the shots taken are independent, since we assume that the probability of a basket on any given shot does not depend in any way on what happened in previous shots.) Each time the player takes a shot, there are two possible outcomes B (basket) or M (miss). Since $P(B) = p$ and $P(B) + P(M) = 1$, we must have $P(M) = 1 - P(B)$. Lets Consider Experiment A from the last section, but this time, we will conduct the experiment withe the probability of a basket on each shot equal to p. In case of confusion, we will call it Experiment A_p.

Experiment A_p The player shoots until the first time he/she misses a shot and then the experiment is over. The outcomes observed from this experiment will be all possible strings of Baskets (B) followed by a single Miss (M). Thus the sample space looks like:

$$\{M,\ BM,\ BBM,\ BBBM,\ BBBBM,\ BBBBBM,\ \dots\}$$

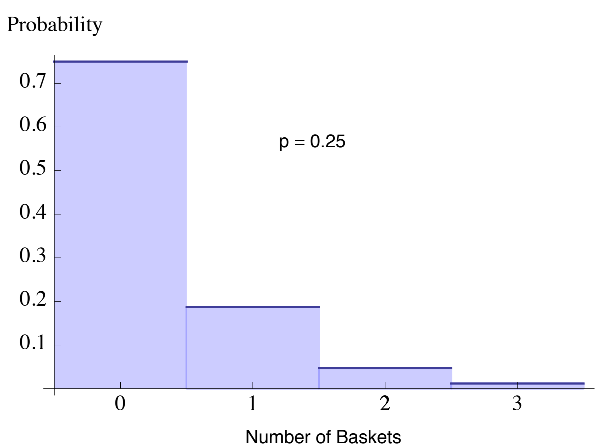
Using the assumption of independence, we can multiply probabilities to get

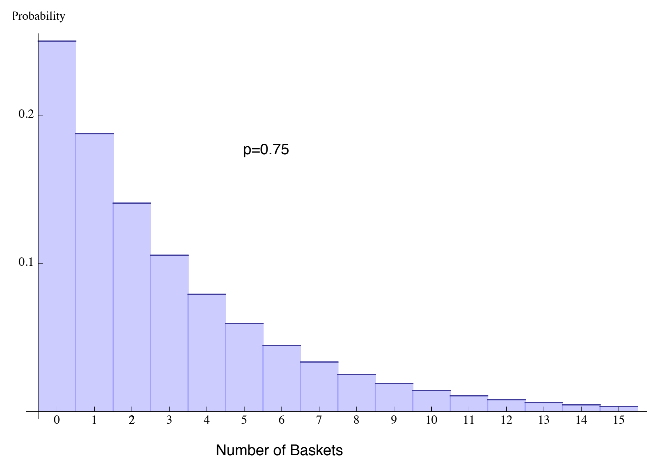
$$P(M) = 1 - p,\ P(BM) = p(1-p),\ P(BBM) = p^2(1-p), \dots,\ P(BBBM) = p^3(1-p).$$

The **probability distribution** for this experiment is shown below:

Length of run of Baskets	Outcome	Probability
0	M	$(1-p)$
1	BM	$p(1-p)$
2	BBM	$p^2(1-p)$
3	BBBM	$p^3(1-p)$
$\vdots$	$\vdots$	$\vdots$
N	$\underbrace{BB\ldots B}_{\text{N times}}M$	$p^N(1-p)$
$\vdots$	$\vdots$	$\vdots$

We show pictures of the distributions when $p = 0.25$ and $p = 0.75$ below. As you might expect, longer runs of success have a higher probability for greater values of p.





Example 4.2. *A basketball player with a probability p of getting a basket on each shot keeps shooting until he/she misses their first shot.*

(a) If $p = .3$, what is the probability that the outcome will be of the form $BBBBBBM$? (where B stands for a basket and M stands for a miss.)

(b) If $p = .8$, what is the probability that the outcome will be of the form $BBBBBBM$?

The Longest Run Note that if a player repeated Experiment A_p 1000 times, we expect that about $p1000$ outcomes to have 1 basket or more (since roughly $(1-p)1000$ of the outcomes should be basket runs of length 0 (M)). We expect about $p^2 1000$ outcomes to have 2 baskets or more, $p^3 1000$ outcomes to have 3 baskets or more, ... about $p^N 1000$ of the 1000 outcomes to have N baskets or more. At some point, we will find a final value of N for which $p^N \times 1000$ rounds to 1, indicating that we would expect only one run of length N or greater in the data. This gives us a method of estimating the length of the longest run of baskets that should appear in our data. We can work by trial and error to find the largest value of N for which $p^N \times 1000$ rounds to 1 or we can use logarithms to solve the equation $p^N(1000) = 1$ to get

$$N = \log_p(\frac{1}{1000}) = \frac{-\ln(1000)}{\ln(p)}$$

where the function ln denotes the natural logarithm and is available on most scientific calculators. We round the answer off to find an integer value for N.

The longest run in M trials of Experiment A_p
Similarly, if we run Experiment A_p R times, we expect the longest run of baskets to have length approximately equal to

$$log_p 1/R = \frac{-\ln(R)}{\ln(p)}$$

(where ln denotes the natural logarithm function) or the largest integer L for which $p^L \times R$ rounds off to 1.

Activity 3: Comparing Data to Expectations

Randomly generated data with $p = 0.75$ Lets look at some data generated randomly to see how it compares with our predictions. The following string of data was generated by the Excel file Trials1.xlsx. We used Excel to simulate 50 trials of this experiment which resulted in 186 shots. In this case, we would expect about $(0.25)50 = 12.5$ runs of baskets of length 0 (M), $(0.25)(0.75)50 \approx 9$ runs of baskets of length 1, $(0.25)(0.75)50 \approx 7$ runs of baskets of length 2 .... Keeping in mind that the law of large numbers just gives us a rough idea of what to expect in a relatively low number of trials, how well does this data fit our expectations for the number of runs of different length?

B B M B B B M M M M B B B M B B B B B M B B B B B B B B B B M M M B B B B M M M M M
B B B M B B B M B B M B M B M B B B B B B B B B B B B B B B B M B B B M B B B B B M B
B B M B M M B B M B M B B M B B M B B B B M B B B M B B B B B M B B M B B B M B B B B
B B B B B M M B B B B M B B B M B B B M B B M M B B M M B B B M B B B B B M B B B B B
M B B B B B B B B M M M

What is the length of the longest run of baskets in the above set of data and how does it compare to what you might expect in 50 trials of Experiment $A_{0.75}$?

4.2. **What to expect in a sequence of K shots by a player with prob. p of success on each?** As with Experiment A of the previous section, If a basketball player performs R trials of Experiment A_p above, he/she usually ends up with a lot more than R shots. We saw that 50 trials of Experiment $A_{.75}$ led to 186 shots in the randomly generated sample in the above Activity. If a player with a probability p of making each shot takes K shots, where K is very large, we would expect (by the law of large numbers) that roughly $(1-p)K$ of the outcomes would be Misses. Therefore the player will have run roughly $(1-p)K$ trials of Experiment A_p.

Below we show the distribution of runs of baskets we should expect from K Bernoulli trials where the probability of success (B) on each trial is p.

Experiment: K Bernoulli trials, $p = P(B)$, $1 - p = P(M)$

Length of run of Bs	Outcome	Expected Number
1	B	$p(1-p) \times (1-p)K$
2	BB	$p^2(1-p) \times (1-p)K$
3	BBB	$p^3(1-p) \times (1-p)K$
$\vdots$	$\vdots$	$\vdots$
N	$\underbrace{BB\ldots B}_{\text{N times}}$	$p^N(1-p) \times (1-p)K$
$\vdots$	$\vdots$	$\vdots$

Example 4.3. *If a basketball player with a 60% chance of success on every shot takes 100 shots in a row, how many runs of baskets of length 4 would you expect to see in the data?*

The longest run of success in K Bernoulli Trials with $P(S) = p$ (see [**?**] for further development of this topic.): Out of K trials of a Bernoulli Experiment with probability of success equal to p, we expect the longest run of success' to have length approximately equal to

$$log_p \frac{1}{(1-p)K} = \frac{-\ln((1-p)K)}{\ln(p)}$$

(where ln denotes the natural logarithm function) or the largest integer L for which $p^L \times (1-p)K$ rounds off to 1.

Example 4.4. *If a basketball player with a 60% chance of success on every shot he/she takes, what is the longest run of baskets you would expect to see in a sequence of 100 consecutive shots for that player?*

4.2.1. *What about Tails?* To convert the above formulas to formulas for probabilities of runs of tails, you should substitute $1 - p$ for p in the formulas.

Example 4.5. *The following data shows a string of consecutive shots taken by* `Dion Waiters` *who has a career FG% of* 0.417.
(spaces separate different games and empty lines indicate a rest).

(a) Is the longest run of baskets in the data consistent with what one would expect for 548 Bernoulli Trials with $p = 0.417$*?*

(b) Is the longest run of misses in the data consistent with what one would expect for 548 Bernoulli Trials with $p = 0.583$*?*

MBBMBMBMB MMMMMMBM MBMBBMMMBB BMMMMM BMBMBMMMBMMBBM BMBMBMBMBBBMBBM MBBMMMBMMMMMMM MMMMBBBMMM MMBMBBMMMMMB BMBMMBMMMBBMMMM MMMMBBBM BBMMBMMMMBMM MMMBMMMMM BBBBBMBBM MBBMBMMBBMMMMBBB MBMMMBMMMMBMBBMB BBBBMMMMBMMBMB MMMBMMMMM

MMMMMMMMBBMBMB MMMBBBMMBBMBMBMMBMMMM BBBMBBMMMM BMMMBMBMMM BBMBMMMM BMBMBBBMMBMMMM BMMBBMMM MMMMBBBBBMMM BMMBBBMBMMBBBBMM MBMM BBBMMBMMMBBMMBMBBM MMBBMMMBMM MMB BMBMMMBMBMBMM
MBBBBMBMBB MMBBMBBBMMMMBBBMMMBBB MMMBMMMB BMBMBM M MMMBM BBBMBMMMMBMMMMMBM MBMMBMB BMBMMMMBM MMBMMMM MMMBMMB BMMBMMMBBBMBBBMB MBBBMMMMBMM MBBMBMBM

BMBMBMMBMMMMBB MBMM MMBMMMBMBMM MBMBMBMMMM BBMMMMBBM