Models from Introductory Probability

Kien Lau, Andrew Zeng

April 2021

Abstract

This paper will cover the introductory ideas of counting principles that will lead to the fundamentals of probability as sets, exploring operations in set theory such as unions and intersections. Using these ideas, we will apply them to the Statistical field, summarizing the Binomial and Poisson distributions respectively. We show a simple binomial model used to predict weather model falls apart through unrealistic assumptions about our world.

1 Introduction

Let's say you and a friend play a game with three coins. When the three coins are flipped, the two of you decide that if all three of the coins land on heads, then your friend wins, and that if two of the coins land on tails, you win. Who is more likely to win? How many total outcomes are there? How likely is it that any of you will win at all?

The applications of probability surround our natural world, whether we realize it or not. In this section, we will pose questions relating to the field of probability, and in section two, essential building blocks to answer those questions will be explained. In the third section, these ideas will be applied and extended to more complicated topics such as the Binomial and Poisson distribution respectively, followed by the Method of Moments to accurately approximate, and altogether culminate to predict a weather model.

2 Preliminaries

To answer those questions, we'll have to break it down into the basics. The basic idea of probability is the field of mathematics that finds out how likely something is to happen, like how often will a six-sided die land on three, or how likely will you win at the lottery. For a start, here are the basic terms you need to know, closely following [1].

Definition 1. *An* ***experiment*** *or* ***trial*** *is the process of testing probability such as trials of flipping a coin or rolling a dice.*

Definition 2. *An* ***outcome*** *is a result of the trial, for example, heads or tails.*

Definition 3. *An* ***event*** *is a set of outcomes, for example, the event of rolling a dice and ending up with a 3 or less would include the outcomes of rolling a 1, 2, 3.*

Definition 4. *A* ***sample space*** *is a collection of all possible outcomes in a trial, for example in a coin flip, it would be heads or tails.*

2.1 Counting Principles

The fundamentals of probability is based on counting principles, like how many ways things can be arranged and counting all the possibilities.

Starting off, one important thing to understand is what a *factorial* is, usually denoted as (x!) and calculated by x $\cdot(x-1)\cdot(x-2)...\cdot 1$

In many ways, this is similar to a **sumation** $\sum_{i=1}^{4} i$, a function which inputs a changing variable that increases by one into an overall sum. Using the summation above as an example, expanding it will hold 1 + 2 + 3 + 4, which would be equivalent to 10.

How does this relate to probability? Consider this: you have three fruits–an apple, an orange, and a pear–and you are tasked with calculating all the possible ways they can be ordered. Since there are initially three fruits, you have three possibilities of what to put down first. Once that first fruit is taken out and put down, you have two fruit left to put down. What fruit you have left depends on the first fruit you put down; for example, if you put down the pear, you have the orange and the apple left. For each of the three previous possibilities, there are two distinct possibilities for the next fruit, or a total of $3\cdot 2$ possibilities. When the second fruit is put down, there is only one fruit left to put down–one possibility for every two of the previous possibilities. This results in $3\cdot 2\cdot 1$ total possibilities, or 3! (three factorial) possibilities.

Expanding on this, suppose that the elements are distinct, but the displayed result would not show the distinctness in the elements of the permutation. Take for example, the arrangement of the nine letters that make up the word "happiness". From the last example, intuitively, one might think it will just be 9! possibilities, however since there are two "P"s and two "S"es in the word, the arrangement would appear something like this: $HAP_1P_2INES_1S_2$. These similar letters can be shuffled around to form arrangements that look the same, even though they have different letters in different places, like this:

$HAP_2P_1INES_2S_1$ is no different then
$HAP_2P_1INES_1S_2$ is no different then
$HAP_1P_2INES_2S_1$ is no different then....

We must remove the arrangements with elements that are alike and repeat. There are two pairs of alike elements in this set P_1 and P_2 as well as S_1 and S_2.The solution to this problem would be

$$\frac{9!}{2!2!} = 90,720$$

The 2!2! represent the different combinations of the nine letters where the only difference is which of the two P's and S's are, and dividing those repeats from the total 9! takes away those repetitive arrangements from the results.

What if someone needed to find the possible arrangement of a specific subset of items within a set? Obviously, you can't just find that with factorials, as that method finds all possible arrangements of all items in the group. If you have a set of six objects, and you need to find out all the ways a pair of objects can be selected from them, the solution ends up being $6\cdot 5$, or the six possible objects initially available for selection, then the remaining five objects not initially selected. This can't be expressed as 6!, though, as that would include $4\cdot 3\cdot 2\cdot 1$ as well. So one might ask, where do the rest of the numbers go? If you divide 6! by 4!, it comes out to be this:

$$\frac{6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}{4\cdot 3\cdot 2\cdot 1}$$

The numbers shared by both numerator and denominator can be canceled out, leaving us with the desired $6\cdot 5$. By dividing out the items we don't need, we can find how many possible arrangements there are for the two items we need. The formula for this is $n!/(n-x)!$, where n is the total number of items to choose from, and x is the items picked from n. By subtracting x from n, we find the leftover numbers that are unneeded to calculate the total number of arrangements, called permutations.

The only problem with this is that different orders of the same objects are counted as different arrangements. For example, if the items are numbered 1-6, then the pairs (1, 2) and (2, 1) would be counted as two separate selections, even though they contain the same thing. Enter: choose notation.

Choose notation is read as "n choose k", and written as $\binom{n}{k}$. This notation gives the number of different combinations of k elements chosen from an n-element set, answering how many different ways of a size k group can be made from a set that's n size. This is combinatorial, so the ordering of the elements in each group does not matter, and the order of each group does not matter.

Choosing k among the n elements in some order can be expressed as $n!(n-k+1)$, but to compensate for the subsets being over counted, the $\frac{n!(n-k+1)}{k!}$ will give the true number of ways to create combinatorial subsets with k from n elements, as dividing by $k!$ removes the repeats from the equation. Through algebraic manipulation, we get the mathematical formula expressed as

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

This represents ways of choosing a combination subset of k elements from n elements. To illustrate, there's a set with four elements, (1, 2, 3, 4). In total there would be six or $\binom{4}{2}$ ways of making combinations of subsets with two elements, meaning (1, 2), (1, 3), (1, 4), (2, 3), (2,4) and (3, 4). The ordering of the subsets don't matter as well as the ordering of the elements in each subset doesn't matter, for example, subset (1, 2) is the same as (2, 1).

A binomial is a polynomial which is the sum of the monomial terms, for example, $(x+y)^1$. The Binomial theorem takes advantage of counting principles and Number theory resembling the patterns of pascal's triangle when the powers of a binomial are algebraically expanded to its full form [4]. A few examples to illustrate are below,

$$(x+y)^2 = x^2 + 2xy + y^2,$$

$$(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3,$$

$$(x+y)^4 = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4,$$

$$(x+y)^5 = x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 + y^5$$

Every term of an expanded binomial follows the rule of ax^by^c where the coefficient a is the **binomial coefficient** calculated as $\binom{n}{b}$ or $\binom{n}{c}$, both will yield the same result. Note that for every term in the expanded binomial, the structure is ax^by^c with every exponent increase in the compact binomial causing a to have the pattern of Pascal's triangle. The following formula will generalize $(x+y)^n$ to any non-negative n.

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k}y^k = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Using counting principles, the coefficients in a binomial with exponent 3 will be

$$(x+y)^3 =$$

$$(x+y)(x+y)(x+y) =$$

$$xxx + xxy + xyx + \underline{xyy} + yxx + \underline{yxy} + \underline{yyx} + yyy =$$

$$x^3 + 3x^2y + \underline{3xy^2} + y^3$$

When the binomial is multiplied by itself for an n amount of times, there are n^2 permutative terms that will be created. Remember back to the example of $HAP_1P_2INES_1S_2$ and combinations, where the distinctness of each letter does not matter, multiplication works the same way because it has the commutative property, thus the terms of permutations can be simplified down into combinations, obtaining the coefficient a of the expanded term.

2.2 Set operations

Events in probability can be converted into sets, and used to calculate the probability of certain events within that set. Set theory is incredibly useful in visualizing probability. There are two main basic set operations that are used in probability, **unions** and **intersections**.

Definition 5. *Given two sets A and B, we define the* ***union*** $A \cup B$ *to be the set of elements in A or B (or both).*

Definition 6. *Given two sets A and B we define the* ***intersection*** $A \cap B$ *to be the set of elements in A and B.*

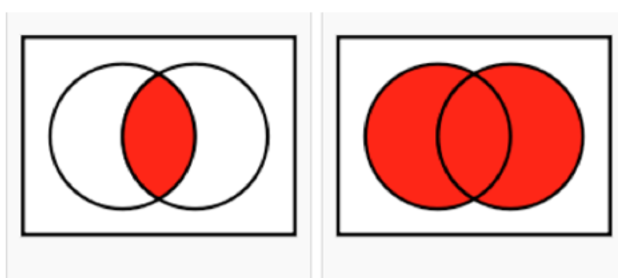
For example, say two dice were rolled. The probability of getting any one combination of two outcomes, if the order of the dice matters, will always be $\frac{1}{36}$. However, say one wanted to find the probability of one dice landing a 6, or a dice landing on a 3. The events would have to be calculated separately and combined together in order for that event to happen. This creates what is called a union, describing the probability of either a six *or* a 3 occur.

If two dice were rolled, the probability of a dice landing a 6 and a dice landing a 3 at the same time would be called an intersection, which is a section within a pair of events that overlaps each other. This could be thought of as an *and* statement, when both events happen. Often, the probability gets smaller because there are more requirements to fulfill, expressed as $(x \cap y)$, in which event x is overlapping with event y or vice versa.

Probability finds the chance of a certain event happening out of a sample space of all possible events. When more than one event is spotlighted within the sample space, these events can interact in ways like unionizing and intersecting that can affect the probability of finding those events within a sample space.

One thing to note is that when discussing probability, there should be a distinction between the sum of the unions versus the sum of the probabilities. Union notation is referring to sets which can overlap and complicate things as we describe later on, while the numerical probability is referring to the true value or "size" of the set.

The inclusion-exclusion principle is how to calculate the union of non-mutually exclusive events. Intuitively, one might think adding the values of the events would be the union of those events, however this will only work for mutually exclusive events because the intersection(s) are overcounted.



Intersection of two sets $A \cap B$ | Union of two sets $A \cup B$

The Inclusion-Exclusion principle is critical in calculating the union of non-mutually exclusive sets. Simply put, this principle over-includes or overcounts, and then takes away at the intersecting areas, excluding those areas from being considered multiple times. With two sets, the premise can be generalised into

$$|A \cup B| = |A| + |B| - |A \cap B|$$

For the union of three non-mutually exclusive events, the inclusion-exclusion principle will be

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$

2.3 Separation of events

There will be two ways that events can be categorized, and this affects or does not affect their probabilities in respect to other events in a sample space!

Definition 7. *When one event's probability remains unaffected by the occurrence of the other event, they are considered **independent**.*

This could be thought of an experiment randomly picking one card from a deck of cards and replacing that card, then shuffling the deck, and taking a second card. The knowledge of what the first card is doesn't change the probability of knowing what the second card is. Replacement is usually the keyword for independent events in a sample space, but this should not be confused with independent trials!

In order for an event to be independent, the equation must be true

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

Independence could be extended to more than two events. All permutations of two of the events would have to be independent themselves, for a set of a certain size to be independent of each other. Essentially, independent events are events where knowledge of the probability of one doesn't change the probability of the other. If the events are not independent, then they are said to be dependent.

On the other hand,

Definition 8. *Events are **mutually exclusive** if they cannot occur at the same time.*

For example, when tossing a coin, Heads and Tails are mutually exclusive since a coin flip cannot be both a head and a tail. There are two possible outcomes in the sample space, but both cannot occur at the same time in one flip. To mathematically satisfy mutual exclusivity, the equation below must be true.

$$P(A \cap B) = 0$$

Ultimately, independence and mutual exclusivity are opposite features, if A and B are independent then knowledge that A happened doesn't change the probability of B, however, if A and B are mutually exclusive, if A happened, B would have a probability 0 of occurring because mutual exclusivity can be thought of as an A **OR** B scenario. Thus from this, events cannot be independent and mutually exclusive at the same time.

This fact is circular because A and B being mutually exclusive will also ruin the fact that the events are independent because if $A \to B = 0$, if $B \to A = 0$, likewise if the probability of A, if B has occurred, is still A, then knowing that B occurred with mutually exclusive events will collapse the probability of A into 0, and ruins the mathematical qualification for events to be considered independent.

To illustrate this in example, suppose that an experiment consists of randomly picking a marble out of a bag of six, so that once a marble selected, the marble cannot be selected again. This is mutual exclusivity breaking independence of certain probabilities because now the bag would only have five marbles left, and the following random selection would have to abide by that new sample space.

Thinking about probability in terms of sets, conditional probability is the idea that a given event(s) becomes the entire sample space or set, and the Desired outcome is now a subset or intersecting with the given set. Using our idea from naïve probability of desired outcomes divided by the events in the sample space, the formula below could be thought of as the same way.

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

A subset is a set that's part of a set that's greater or equal in size. Of course, this would mean a subset is always an element of the original set. Conditionals could generally be thought of as the subset becoming the sample space of the conditional probability. In simple terms, conditionals are zooming into a specified set, and determining the probability of the desired set which lies inside of the specified set. This means that if conditional $P(E \mid F) = 0$ then we can conclude that the events are mutually exclusive since E does not lie in F; $E \cap F = 0$, so that they do not intersect anywhere.

Notice that for independent events, $P(E \cap F) = P(E) \cdot P(F)$, however if E and F are dependent of each other, $P(E \cap F) = P(E) \cdot P(F|E)$. Using the general idea of $\frac{P(E \cap F)}{P(F)}$. One might ask, why are the approaches to calculating the intersections of independent and dependent events different? Firstly, we could prove the supposition that the intersection of independent events are just the probabilities multiplied by rewriting the mathematical requirements for independent conditionals as

$$P(E \mid F) = \frac{P(E) \cdot P(F)}{P(F)}$$

Notice that the numerator and denominator, $P(F)$ cancel out, leaving $P(E \mid F) = P(E)$ as the conditional for independent events.

2.4 Two approaches to Probability

Axioms are the most basic assumptions that one must make in Mathematics, and they often build off of one another to create complex fields such as probability. An axiomatic system such as modern probability uses a set of three axioms which can be used in conjunction to logically derive theorems.

Axiom 1. $0 \leq P(E) \leq 1$
Axiom 2. $P(S) = 1$
Axiom 3. For mutually exclusive events E_1, E_2, ...

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

Understanding the Axioms Axiom 1 states that the probability of any given event must lie between the values of 0 and. The minimum and maximum could be completely arbitrary, but for simplicity's sake as we'll see with the next axiom, 0 and 1 is easier to quantify and rationally think about.

Axiom 2 states that the probability of the entire sample space must equal the maximum possible value, 1, and this would be the maximum for any arbitrary value that axiom 1 has.

Axiom 3 states that the probability of mutually exclusive sets added together is the same as the probability of the union of each event.

From these axioms we obtain the complement rule, which states

$$P(E^c) = 1 - P(E)$$

Intuitively, this is logical, saying E complement (everything that's not E, expressed as E^c) equals the sample space, 1 minus E. How do we prove this using the axioms?

Since E and E^c are mutually exclusive,

$$P(E \cup E^c) = P(E) + P(E^c) \text{ From axiom 3}$$
$$P(E \cup E^c = P(S) = 1 \text{ From axiom 2}$$
$$P(E) + P(E^c) = 1$$
$$P(E^c) = 1 - P(E)$$

Naive probability is the simple idea that all events are equally likely to occur. Counting principles are extremely useful when calculating naïve probability because the count of certain permutations/combinations will be the desired outcomes, and the count of all possible permutations/combinations will be the entire sample space. Before the axiomatic system of probability, Naïve probability was the prevalent approach with the probability of an event being calculated as number of times a desired event happened divided by the total number of trials, or

$$\text{Probability(E)} = \frac{\text{Desired outcomes}}{\text{Total outcomes}}$$

However, this simple and often intuitive model of likelihood breaks down when comparing non-equally likely events. This principle could wrongly be applied to absurd ”yes” or ”no” questions, and yield a fifty-fifty chance for either one to occur. Indeed, we see that when this approach to calculating probability doesn’t always necessarily yield the same result when the calculation is done again, the system breaks down and becomes illogical. It’s critical that the reader recognize that in order for naïve probability to hold true, many assumptions about the events in that sample space have to be made.

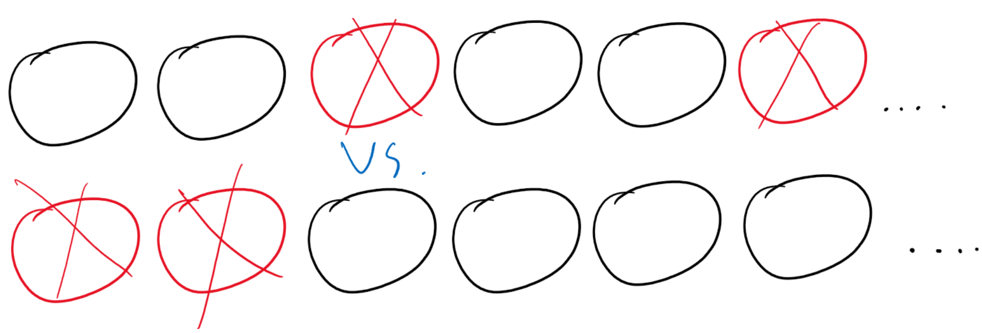
For most of probability, mathematicians often assume naive probability or events having equally likely outcomes, but in real life, not all possibilities have an equally likelihood of occurring. This intersects with other fields such as Psychology, for example, a certain flavor at an ice cream parlor might be chosen significantly more than the other options so to make assumptions like all flavors and colors are equally appealing, no one has allergies/health conditions preventing them from eating any flavor, or every customer’s favorite flavor is equally split, etc, would be a large assumption to make.

The third axiom of modern probability covers this topic with the probability of the unions of each event always being equivalent to the summation of the probability of each event. This means that for mutually exclusive events, the unions of each event equivalently yields the individual probabilities of each event added together, meaning the corresponding set of a particular event is equal to the numerical probability of that particular event. Simply, this axiom allows for mutually exclusive non-equally likely events to be calculated.

2.5 The “Paradox” of Counting

Imagine that three balls were added infinitely to a line, (a) how many balls remain if the third ball was removed every cycle of addition? (b) How many would remain if the first ball was removed every cycle of addition?

The answers for each hypothetical are starkly contrasting with outcome (a) resulting in infinitely many balls while (b) would result in zero balls leftover. The same idea of (a) could be applied to when the second ball was removed every cycle, because the left side would infinitely accumulate balls on the left hand side, meanwhile if the first ball was removed every cycle as in (b), subtracting one infinitely many times from an infinitely large count would result in zero left over.



Georg Cantor developed a concept that’s extremely similar to this; the set of all even numbers is equivalent to the set of all integers. Intuitively, one might think that the set of even numbers are half of all integers because the set of even numbers is a subset of all integers. Relating back to our thought experiment, one ball is a subset of the three ball addition every cycle, but notice that one ball taken away infinitely many times will equal any non-negative number added infinitely many times. Ultimately, probability gets weird when using infinity, and the answer might not always be intuitive!

3 Applications

Probability is the measure of certainty, a look into a "crystal ball" of sorts for plausible futures. Being able to truly understand the likelihoods, graphs, and ideas from the previous section is critical in truly grasping this power through the use of mathematics.

The following section will cover applied probability, and use cases in Statistics, a closely related field of manipulating uniformly distributed probabilities to yield interesting and useful results.

3.1 Law of Large Numbers

Suppose that you flipped a coin and in the first flip, the result is a head, would you say that a coin flip will always 100 percent result in a head? No, intuitively, we consider a coin flip using naïve probabiltiy, saying that heads has probability .5 and tails has probability .5, so as the number of trails approaches infinity, we should expect that the number of heads and tails are roughly equal with either outcome having a slight difference than the other. This however, is negligible because the slight difference in applied outcomes is nothing compared to the total number of trials. Gambling could be thought of as the same way with each trial being a pull of a slot machine. Initially one might have a large winning streak, thinking they beat the odds, however, as more and more pulls are done, the probability levels out to roughly the expected probability.

From day to day life, the law of large numbers is a subtle yet constant reminder of the applications of probability. This law stipulates that in probability, the more times a trial or experiment is done, the more accurate the data is to representing the expected probabilities of the events in the sample space. Anomalies of any data set are dampened out through the expected probabilities balancing the unrepresentative beginning trials after a large number of trials, hence the law of "large numbers".

3.2 Statistical distributions

A fun history The Bernoulli family is famous throughout history for their works in mathematics and one Jacobs Bernoulli (1655-1705) develops a rivalry with his brother, Johann after collaborating together on various applications of calculus. As their mathematical genius matured, both of them began attacking each other in print, and posing difficult challenges to test the others' skill [6].

In Jacob Bernoulli's most original work, *Ars Conjectandi*, he reviews the work of others on probability, and most notably, the term Bernoulli trial emerges. This trial describes an independent random experiment with exactly two possible outcomes as its sample space, "success" and "failure". Note that "success" and "failure" is all arbitrary, and represents the desired outcome that the mathematician is analyzing. The probability p represents "success" and probability p^c represents "failure" because remember, there are only two outcomes in a Bernoulli trial, so the following equation must be true if the experiment is a Bernoulli trial. We could conclude the same result using the complement rule.

One thing to keep in mind, independence in events should not be confused as independent trials. Independent trials could be thought of as doing that specific experiment under the same conditions as the first trial, meaning the outcomes of previous trials has no effect on future trials and everything is reset. Independent trials can have mutually exclusive events.

Closely related to the ideas of Bernoulli trials is the Binomial distribution, where the Bernoulli trial is done n times, each with a probability of success p. Using our proof that E and it's complement must add up to the total sample space from our introduction to the three axioms of probability, we can treat the $P(E)$ of success and $P(E^c)$ of failure the same way, since there will only be two events in the sample space.

$$P(E) + P(E^c) = 1$$

The probability of exactly k successes of any order in the experiment is given by $B(n, p)$ or probability mass function below [3].

$$f(k,n,p) = Pr(k;n,p) = Pr(X=h) = \binom{n}{k} p^k (1-p)^{n-k}$$

Note that the probability of obtaining k successes will never be 1, or certain, unless the probability for a success is 1.

The Poisson distribution is very similar to the Binomial distribution as it takes in one parameter lambda or λ, which is n, the number of "successes" and p, the probability of the "success" occurring. The differing characteristic is that a Poisson distribution is discrete and measures the probability of k, successes occurring with the parameter λ, a constant mean rate within a certain time frame. λ = np when n is large and p is small. Note that the trials for k successes must be independent of each other.

While the requirements for a Poisson distribution seems strict, there's a work around! The Poisson paradigm states that for a distribution to be approximately of Poisson distribution, it's not required that all k have the same probability of occurring, but only that all of the probabilities are near negligibly different, and the trials to be "weakly dependent".

The probability mass function for a Poisson distribution is calculated as shown below

$$f(k;\lambda) = Pr(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Note that for any Poisson distribution this will never guarantee that a large number of k successes won't happen, however it will drop off logarithmically, approaching the limit of zero.

3.3 Method of Moments

In theory, if one could sample infinitely many points from a distribution, the parameters of the distribution could be determined to arbitrary precision. However, in real life, this process is not so straightforward. There can be errors in the sample data, and for small sample sizes, the estimated parameters may not be very accurate, depending on the method used to calculate them, such as *maximum a posteriori*, or "MAP," estimation. Note that the first moment of any distribution, whether Binomial or Poisson, is always the mean or μ of the distribution. [5]

Generally the μ is calculated as follows

$$\widehat{\mu}_j = \frac{1}{n} \sum_{i=1}^{n} w_i^j$$

The equation above translates to summing up all of the values and dividing by the number of times that a value was added. Essentially, μ is another word for mean.

To test the method of moments, the following steps will demonstrate how to use the method of moments, and how accurate this method is to approximating the parameters of a certain Poisson distribution with the unknown parameter λ. We sample this Poisson distribution, and find the following first two moments.

$$E|X| = 1.372, \quad E|X^2| = 3.246$$

Remember, the first moment is always the μ, with the second moment being put into a moment function generator yielding,

$$E|X| = \lambda$$
$$E|X^2| = \lambda^2 + \lambda$$

To solve using the method of moments, the values of $E|X|$ and $E|X^2|$ are set equal to their moments, and solved respectively. The values are slightly different, but the estimates are not too far apart from each other and the true parameter of λ (1.475) in this Poisson distribution. The benefits of this approximation method is that it's really quick and can be done by hand, on top of that it's surprisingly accurate! This method gets more accurate as there are more variables to "check" the true value of the other variables.

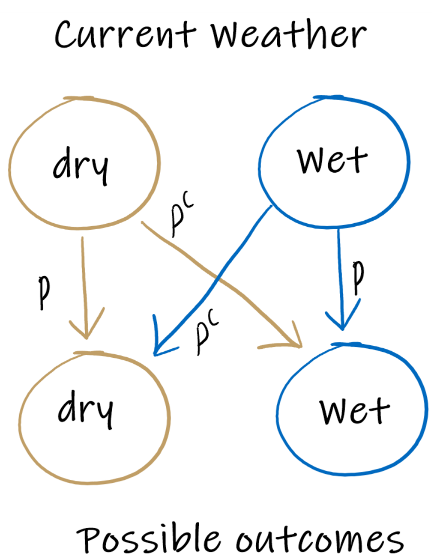
4 A simple weather model

Suppose that a weather forecasting algorithm is able to predict tomorrow's weather using a probability p, the **probability that it will remain the same as the current** weather. In this problem, there will only be two outcomes, the weather is dry or the weather is wet, both are mutually exclusive. The events in this sample space however, are probability p and p^c because the only outcomes are that the weather stays the same or changes. Given day zero was dry, the question remains how would someone solve for P_n the total probability that the weather on day n is dry?

In Computer Science the idea of working backwards through an equation is an idea called Recursion. Since the total probability of dry for any given day rests on the total probability on dry the day before that, Recursion could be used to solve this problem.

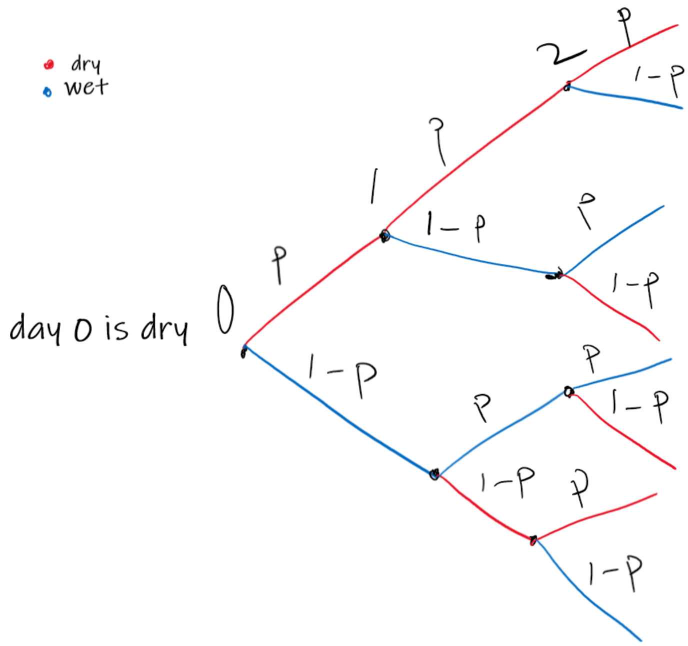
4.1 Overview of the Model

In our problem, the state of the weather is a *boolean*, where a true or false defines the outcome. In Boolean Algebra and in this problem, the true/false are mutually exclusive meaning it cannot be both dry/wet on the same day. Notice that the outcome of the weather tomorrow solely depends on the current weather. From this, we get the model shown below



Using this model, a matrix can be created, where the column represents the current weather state and row representing future possible weather states.

$$\begin{array}{cc} & \begin{array}{cc}\text{dry} & \text{wet}\end{array} \\ \begin{array}{c}\text{dry}\\ \text{wet}\end{array} & \begin{pmatrix} p & p^c \\ p^c & p \end{pmatrix}\end{array}$$



Using the model above, intuitively, one might solve this problem using the Binomial theorem, calculating the probability for each branch that ends on day n in dry weather and summing the probability of other branches that had the same result. This approach gets overwhelmingly complicated for a large n days because probability p measures that the weather stays the same, not the probability that the following day is dry. Looking at day three, notice that there are 2^n or eight total possible outcomes of the weather be wet or dry, or permutations of p and p^c. From our ideas of combinatorial proof of the Binomial theorem, realize that the three permutations of dry weather through two p^c's and one p could be simplified into $3(p \cdot [1-p]^2)$.

Let today be dry, and $p = .8$, thus $p^c = .2$, and $n = 3$

$$(.8^3 + 3(.8 \cdot .2^2))$$

Realize the pattern as n increases is similar to how the exponent of a binomial expands! The above equation could be generalized to

$$\text{Given } P_0 = 1 \text{ ,} \quad n \geq 1$$
$$P_n = (2p-1)P_n - 1 + (1-p)$$

Alternatively, we can conclude the same answer below using the Master Theorem because the map has a self-similar and recursive structure.

$$\text{For } n \geq 1$$
$$P_n = \frac{1}{2} + \frac{1}{2}(2p-1)^n$$

4.2 Real Data Testing

Let's test the culmination of the ideas in this subsection with actual data! The city of Milwaukee's daily weather data from Jan. 1st, 2000 to Jan. 1st, 2005, will be analyzed using code [2]. Assuming the five year period is a large enough number of trials, thus accurate of the true value for p in Milwaukee. In our problem, there will only be two possible outcomes, dry or wet, so to keep things simple, if the value of precipitation in any day is zero, then the weather is dry, however, if the value of precipitation in any day is greater then zero, then the weather is wet. Value p in this example will be calculated when tomorrow's weather stays the same as today. From the visualization model there will only be four scenarios, dry $\rightarrow$ dry or wet $\rightarrow$ wet, to calculate p, and dry $\rightarrow$ wet or wet $\rightarrow$ wet, to calculate p^c respectively.

Ideas applied- Using code to analyze real data, we see that between Jan. 1st, 2000 to Jan. 1st, 2005, there are 1098 days where the weather stayed the same in the following day, and 730 days where the weather switched! Let's now put the total desired outcomes over total sample space. $\frac{\text{p}}{\text{total sample space}}$ or $\frac{1098}{1098+730}$, equivalently, $p = 0.60065$. To test out another idea above, stating, $p + p^c = 1$, we could input p and algebraically solve for p^c, but since the total number of times the weather switched is 730, we could similarly repeat the step above, and solve $\frac{730}{1098+730}$, both equations roughly yielding $p^c = 0.39934$.

In the following test, an experiment will be ran on the same data set comparing the current weather state to the weather state up to five days ahead. For example, the code will compare the weather state of Jan. 1st, 2000 $\rightarrow$ weather state of Jan. 1st, 2000 + five days later. This process will be repeated until the "current" is Jan. 1st, 2005.

As described using the model above, the initially unequal probabilities even out to naïve probability, that is the total probability of dry weather in a large n is fifty percent. This answer to a large n days might seem unintuitive since the probability of the weather staying the same/changing is known, but realize that to guess if the weather stays the same 100, 500, 1000, or 10,000 days later from today is really inaccurate! The algebraic solution using the Master Theorem to this problem also concludes the same point. Indeed, as n, infinitely increases, the $(2p-1)^n$ portion of the equation will approach closer and closer to a value of zero, making solution evaluate to $P_n = \frac{1}{2} + 0$.

When comparing the current weather to n days ahead, we find that as n increases, P_n remains roughly the same. Realize that this will affect the algebraic solution and conclusion to this weather model.

n=1	$P_1 = 0.600656455142232$
n=1	$P_2 = 0.5897155361050328$
n=2	$P_3 = 0.6061269146608315$
n=3	$P_4 = 0.5946389496717724$
n=4	$P_5 = 0.612144420131291$

When we input P_n into the equation and solve for p in the right hand side of $P_n = \frac{1}{2} + \frac{1}{2}(2p-1)^n$ we find that the value of p exponentially increases as n increases. P_n stays roughly the same on the left hand side of the equation, with the differences in P_n's being nearly negligible, while on the right hand side of the equation, the exponent of n increases, causing the value of p to increase exponentially.

4.3 Takeaways

There are a multitude of ways to tackle this problem, but from introductory probability and under some assumptions, we've just demonstrated that for a large number of days after, the weather being dry or wet will be roughly fifty-fifty, regardless of the probability less than 1 that the weather stays the same. Contrastingly, from our real data testing of Milwaukee, we discover that real life weather does not fall in such a way that this model works. After just comparing the "current" day to two days later, the algebraic model breaks apart, in particular we see that the value of p exponentially increases than the true value. Ultimately in our algebraic model, even as the probability of the weather staying the same approaches infinity, the accumulating, and increasing assumptions that have to be made will level out the probability to be fifty-fifty. In our complex world, this theoretical situation and its solution does not become so blankly true because the perfect assumptions of this problem wouldn't fit the data to the model. The core idea about probability is that the less assumptions have to be made, the more certain one is about how true their odds are.

References

[1] Sheldon Ross (2018) *A First Course in Probability, Tenth Edition*, Pearson.

[2] Yuchuan Lai, David Dzombak "Compiled historical daily temperature and precipitation data for selected 210 US cities", 08/01/2021, https://kilthub.cmu.edu/articles/dataset/Compiled_daily_temperature_and_precipitation_data_for_the_U_S_cities/7890488?file=25951535

[3] Wikipedia "Binomial distribution", https://en.wikipedia.org/wiki/Binomial_distribution

[4] Wikipedia, "Binomial theorem", https://en.wikipedia.org/wiki/Binomial_theorem

[5] Wikipedia, "Method of Moments (statistics)" https://en.wikipedia.org/wiki/Method_of_moments_(statistics)

[6] Wikipedia, "Jacob Bernoulli", https://en.wikipedia.org/wiki/Jacob_Bernoulli