Learn Probability

MathopediaForAI Tutorial 4.1

This is a tutorial that develops a firm foundation of probability starting from basics such as sets and logical representations to other intermediate level topics such as principle of inclusion and exclusion, constructive counting, pigeonhole principle, fibonacci numbers and graph theory.

Sets and Logic

Probability itself is the analysis and application of mathematical logic to develop predictions based on the likeliness of a defined event. To understand this, we'll first define some universal notations and key words which will be used throughout this tutorial.

Let's start with *sets*. Sets are essentially groups of numbers which can have any nature - rational, irrational, real, imaginary, integers, decimals, fractions, etc. Some of the closely related ideas with respect to sets are as listed below.

- Subsets and supersets - subsets are sets in which all the elements are also part of a larger set, which is called the superset.
- The empty set - often referred to by the symbol $\emptyset$, the empty set (also known as the null or void set), is a set that contains no elements. It can also be written as $\{\}$
- Power set - A set that contains all the possible combinations of subsets that constitute the parent set. The power set of the set S is represented as $\mathcal{P}(S)$.
- Unions and intersections - Union of 2 sets is represented as $A \cup B$ and outputs the set which contains all elements from A and B individually; Intersections are represented as $A \cap B$ and outputs the set which contains elements that are common in A and B.
- Converse of an inference - For an implication relation $P \implies Q$, the converse of the statement is $Q \implies P$
- Contrapositive of an inference - For an implication relation $P \implies, Q$, the contrapositive of the statement is $\text{Not}B \implies \text{Not}A$

- Simple truth and logic statements - true, false, and, or not

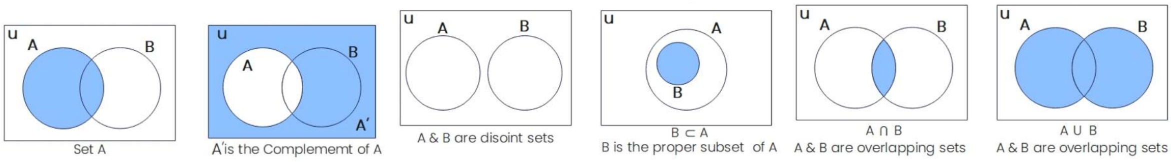

Image source: https://images.edrawmax.com/symbols/venn-diagram-symbols/venn-symbol-1.jpg

Principle of Inclusion and Exclusion (PIE)

Essential idea: The principle of inclusion and exclusion (often abbreviated to PIE) describes the elements that are a part of or excluded from the conditions posed on the union of two or more sets. Generally, this can be visualized as the number of items placed in a set that have one specific property.

To understand this better, let's start with a union of 2 sets, A and B. As per PIE for these 2 sets, we have that $\#(A \cup B) = \#(A) + \#(B) - \#(A \cap B)$. In simpler words, this means that the number of elements in the union of A and B will be equal to the sum number of elements in both the individual sets minus the number of elements that are common in both (since these will be added twice otherwise).

Applying the same concept to three sets, it can be said that

$$\begin{aligned}\#(A \cup B \cup C) = \#(A) + \#(B) + \#(C) \\ -\#(A \cap B) - \#(B \cap C) - \#(A \cap C) \\ +\#(A \cap B \cap C).\end{aligned}$$

Observing the pattern in the implementation of this principle, we can also generalize it to n number of sets and deduce that for union $A_I = \cap_i \in IA_i = A_{i1} \cap A_{i2} + \cap \ldots \cap A_{ir}$, we have that

$$\left| \cap_{i=1}^{n} \right| = \sum_{\emptyset \neq I \subseteq N_n} (-1)^{|I|-1} |A_I| .$$

Expanding this, we can rewrite the generalization as

$$\begin{aligned}\#(A_1 \cup A_2 + \ldots +_n) = \#(A_1) + \#(A_2) + \ldots + \#(A_n) \\ -\{\#(A_1 \cap A_2) + \#(A_2 \cap A_3) + \ldots + \#(A_{n-1} \cap A_n)\} \\ +\{\#(A_1 \cap A_2 \cap A_3) + \ldots + \#(A_{n-2} \cap A_{n-1} \cap A_n)\} \\ -\{\#(A_1 \cap A_2 \cap A_3 \cap A_4) + \ldots\}\end{aligned}$$

$$
\begin{aligned}
+ & \qquad\qquad \vdots \\
& +(-1)^{n-1}\#(A_1 \cap A_2 \cap \ldots \cap A_n).
\end{aligned}
$$

However, PIE is not always applicable in its raw form. In most cases, there are certain conditions imposed on the unions of the sets which means that directly applying PIE may not give an accurate answer. As such, the ‘statement’ of PIE is not as important as the idea itself: when two or more sets are joined in an union, the number of elements in their union can be counted by summing up their individual elements and making sure that no elements is added or subtracted twice.

Linking back PIE with the idea of complementary statements, it should be noted that while PIE is useful in identifying the number of properties that have a certain property, it can also be just as useful when trying to determining the number of elements that *don't* have a set of properties - just subtract the number of elements with those properties from the total number of elements!

Other Key Techniques and Concepts in Probability

Constructive Counting

Constructive counting deals with the idea of 1-1 correspondence between two sets to determine the number of elements in them. As the name suggests, constructive counting deals with the addition of elements to the final "list of elements" so that the 1-1 correspondence can be used to determine the size of the original set (which is difficult to analyze in itself) by first finding out the size of another set with an equal number of elements (which is easier to tackle). In most cases, determining whether 2 specific sets share a 1-1 correspondence is the main part of the problem and to prove that 2 sets share this kind of relation, we must show that for every element in A, B the exact same corresponding element and for every element in B, A has the same corresponding element. Sometimes, this can be done in an easier way by also splitting up sets A and B and then comparing these subsets to deduce the 1-1 correspondence.

Another key technique can be the use of complementary properties to determine the number of elements that *don't* have a certain set of properties by subtracting the number of elements that *do* have those properties from the total number of elements. Also, suitable diagrams can always be useful for understanding the problem, if not solving it, as they provide a visual explanation of the situation.

The Pigeonhole Principle

Essential idea: If n balls are placed into k boxes for $n > k$, then at least 1 box contains more than 1 ball.

Reading it loud, the pigeonhole principle seems rather intuitive and 'obvious'. But mathematically, it has several implications which can be useful in determining important relations and equations in probability. For instance, one key implication of this principle is that if n objects are placed into k boxes for $n > k$, then each box should contain minimum of $\frac{n-1}{k} + 1$ items. Of course, these boxes don't actually have to be tangible boxes but could simply represent a category as well. This means that the Pigeonhole principle can also be used for determining or classifying 'objects' (which can also be numbers) into multiple 'boxes' (or categories based on their individual properties.

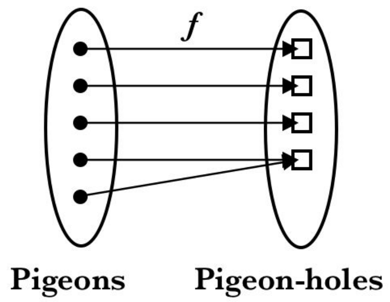

Image source: https://i.pinimg.com/736x/7e/00/ed/7e00ed579e584820 6bb88c68b4c55d58.jpg

Therefore, the pigeonhole principle can be used to determine the average number of items per category as well. However, this principle assumes independent assortment - which means that no event / factor can influence the sorting of the birds. Since this is almost never true in real life event based probabilities, we introduce another factor, namely constructive expectation to take each sub-event into account.

Constructive Expectation

As the name suggests, constructive expectation uses an additive approach (hence *constructive*) to determine the likeliness of an event occurring (hence *expectation*) based on the individual sub-parts of that event. Hence, the idea of constructive expectations can be written mathematically as

$$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n).$$

Intuitively, every event X can be seen as a set of small events which are $X_1, X_2, \ldots, X_n$. Thus, the equation above essentially gives us the expectation of the event X occurring by considering the expectations of each of its individual constituents (or sub-events). Despite this general idea though, it is also possible to have an event X which has sub-events that depend on each other - i.e. the expectation of event X_i may depend on the expectation of some other event X_{i-1}. In such cases, the concept of constructive expectation still holds true since it considers the individual expectations of the sub-events which already exist.

Distributions

So far, we have been looking at individual objects which can be put into different categories or boxes based on certain properties. Now, we will take a look at objects which are indistinguishable which are being put into distinguishable boxes. Since we are dealing with distinguishable boxes, let's consider a single big space wherein we can put in our own slots thereby having the privilege to add as many boxes as needed. If we have n items to be placed in k boxes, then we would have to add $k - 1$ dividers among $n - 1$ slots.

Using this logic, the number of possible distributions can be given as

$$\#\text{ of distributions } = \binom{n-1}{k-1}.$$

The general algorithm for creating these distributions can therefore be written as

$$\left\{\begin{array}{l}\text{Distributions of n} \\ \text{items to k boxes,} \\ \text{with some box(es)} \\ \text{possibly empty}\end{array}\right\} \leftrightarrow \left\{\begin{array}{l}\text{Distributions of n+k} \\ \text{items to k boxes,} \\ \text{with no boxes empty}\end{array}\right\}$$

$$\leftrightarrow \left\{\begin{array}{l}\text{Insertion of k-1} \\ \text{dividers into n+k-1} \\ \text{slots}\end{array}\right\}.$$

For distribution problems that impose certain conditions / restrictions on the properties of the distributed elements, apply the most restrictive condition(s) first as this would help you narrow down your options in the very beginning itself.

Mathematical Induction

The principle of mathematical induction allows us to prove a certain conjecture based on the intuitive idea of the domino effect - we just need to prove that assuming the conjecture holds true for a certain value of n, this very assumption automatically implies that the conjecture must hold true for $n + 1$ as well. Then, we show that the conjecture does hold true for the least possible value of n and therefore prove that the statement holds true for all n.

As anticipated though, this method of proof has some significant limitations as it can only be applied for proving statements based on a single variable and will only provide proof of its existence for integers.

Image source: https://www.chilimath.com/wp-content/uploads/2020/12/mathematical-induction-as-falling-dominoes.png

Thus, whenever a statement needs to be proven for a set of positive integers, mathematical induction is the way to go!

Fibonacci Numbers

As nature's most popular representatives, fibonacci numbers are a setform a set that has infinitely many elements, the first few of which are as listed below.

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, \ldots$$

In mathematical notations, the sequence is defined as $F_0 = 0$, $F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for all $n \geq 2$. This recursive definition of the fibonacci sequence can also be represented as the following closed form.

$$F_n = \frac{(1+\sqrt{5})^n - (1-\sqrt{5})^n}{2^n\sqrt{5}}.$$

As un-intuitive as it might appear, this formula can also be proven by mathematical induction (give it a shot!). It also has links to Binet's formula and utilizes the golden ratio to define F_n in terms of that.

Recursion

<u>Essential idea</u>: Recursion can be visualized as the complementary concept of constructive counting. In recursion, the n^{th} term is defined in terms of the previous few terms. An example of a recursive sequence is as shown below.

$$C_n = C_0C_{n-1} + C_1C_{n-2} + \cdots + C_{n-1}C_0 = \sum_{k=0}^{n-1} C_kC_{n-1-k}.$$

The sequence above shows the Catalan numbers sequence defined recursively.

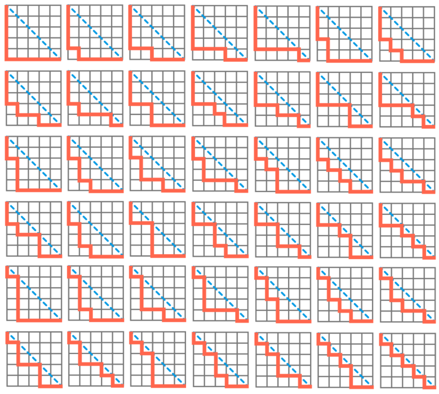

Image source: https://dso55uzetaobb.cloudfront.net/brioche/uploads/mzzJsNGyoR-group-2.png?width=1200

'Solving' recursions typically refers to the derivation of an expression for the n^{th} term in terms of n alone with no dependency on the previous terms. Thus, 'solving' the catalan sequence would give a general formula for the following expression of the n^{th} term.

$$\boxed{C_n = \frac{1}{n+1}\binom{2n}{n}}.$$

For developing a general formula, a conjecture can be tested with some small values - this won't prove the conjecture but can still be useful in determining whether the conjecture can possibly hold true and if needed, it can also be used to determine the necessary changes to the conjecture. Also, if two recursive series have the same generalization or solution, they might be sharing a one-one correspondence relation!

Conditional Probability

Essential idea: Classic probability is defined as the probability (or likeliness) of an event occuring. Thus, adding this concept to conditional probability, this conditional probability can be defined as the likeliness of an event occurring under a certain set of conditions or restrictions.

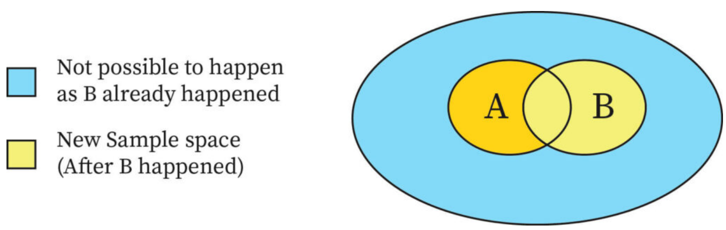

Image source:
https://cms-media.bartleby.com/wp-content/uploads/sites/2/2021/05/31143718/conditional-probability-1-1024x308.jpg

Thus, for an event X which consists of several equally likely sub-events, we can say that

$$P(X) = \frac{\text{Number of successful outcomes}}{\text{Number of possible outcomes}}.$$

Rewriting this mathematically, we can say that for an event X to occur under some condition Y, the probability can be given as

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

The main 'catch' in a conditional probability problem is usually the identification of the event and condition(s), so watch out!

Combinatorial Identities

Essential idea: Combinatorial identities often have an algebraic and geometric representation , therefore such identities can be proven in terms of 3 different methods : algebra (mathematical induction) , block walking argument , and committee forming argument.

The first two methods are relatively common while the third method is specific to combinatorial proofs only. This method has 3 principal steps to it.

1. Finding a subset that is easy to deal with
2. A single coefficient in terms of nCr
3. Developing Case works based on the possibilities of coefficients

Sometimes using common combinatorial identities can help prove new ones . Another tip is to look for coefficients or parts of coefficients which are repeated and can be replaced by specific variables.

Events with States

Essential idea: Stages are the sub events which constitute an event. Identification of states is an important step in solving a problem as this would determine the variables to be considered for calculating the conditional probability. Thus it is useful to define variables for each individual state.

A classic example of state based events could be a tournament played by 2 players. In this case, the individual state would correspond to whether a person wins or loses a particular round.

Some ways of dealing with such a problem include pattern identification by plugging in small values or constructing block diagrams with each block representing a state and arrows representing the flow of events . Both of these methods help in identifying the general relation between multiple states and are useful for developing conjectures.

Also, not every problem needs to have a solution which gives a general equation for the probability of the main event . Occasionally, problems based on states expect pattern identification and proofs for these patterns but they need not be expressed as mathematical expressions.

Generating Functions

Essential idea: Generating functions are representations of infinitely long recursive sequences in terms of polynomials of some degree n. Since they are polynomials, their general form can be written as $f(x) = a_0 + a_1x + a_2x^2 + \cdots$ .In these polynomials, the coefficient a_k will depend on k, where k is the exponent of x.

Most of the time, problems that deal with generation functions will require you to calculate the value of each individual coefficient based on which the generating function can be defined. Such an example can be given by the function $\frac{1}{(1-x)^n}$. For this, the coefficient of x^k will be given as shown below.

$$\text{Coefficient of } a^k = \binom{n-1+k}{k}$$

Similarly, the generating function for the fibonacci sequence is given as $f(x) = \frac{x}{1-x-x^2}$ and this can also be used to prove Binet's formula. As an additional tip, generating functions are typically applied in distribution problems in which a lot of casework is required.

Graph Theory

Essential idea: Graphs in probabilities are used for solving problems visually by describing the complex relationships between different entities.

Terms in planar graphs:

- Edges: the lines that connect pairs of vertices
- Vertices: The elements that are part of a set are each represented as a vertex

- Path: Finite sequence of edges which ultimately connect vertex vertex A to vertex B
    - Simple path - path with no repeated edges
    - Eulirean path- connects every edge on the graph exactly once
    - Hamiltonian Path - contains every vertex (except the starting and ending ones) exactly once

Properties of graphs:

- Elements of set represented as vertices
- Relationship between pairs of elements represented as edges
- For n vertices and d degree, a graph will have $\frac{nd}{2}$ edges.
- Bipartite graphs - Graph G is bipartite if vertices of G can be partitioned in two 2 sets S and T which can be connected to each other and no cycle of odd length
- Connected graphs: For every possible pair of vertices A and B, there exists a path to connect these two vertices
    - Tree - connected graph with no cycles, a tree with n vertices has a minimum of $n - 1$ edges
    - Leaves - every tree contains at least 2 edges of degree 1 and these are known as leaves

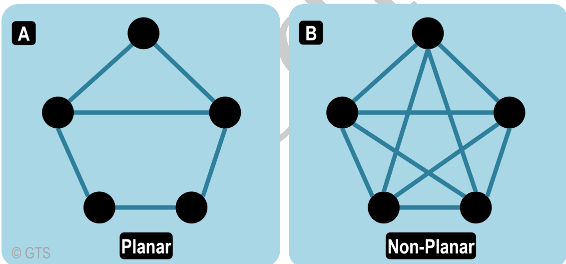

Image source: https://transportgeography.org/wp-content/uploads/planar_non_planar_graphs.png

Eluer's Formula for planar graphs:

$$\boxed{V + F = E + 2}$$

E - Number of Edges
F - Number of Faces
V - Number of Vertices

Another thing to remember about planar graphs is that if it has more than 1 edge, $E \leq 3V - 6$. Also, any graph for which $E > 3V - 6$ is not planar.