STATISTICS-Sem 2-Gen-Statistical Definition of Probability and Conditional Probability-Saptarshi Mondal-28 Mar 2020

A statistical definition of probability (Due to *VON MISES*)

To do this we need to define the concept: relative frequency.
Relative frequency is the proportion of the sample space on which an event E occurs. In an experiment with 100 outcomes, and E occurs 81 times, the relative frequency is 81/100 or 0.81.

The frequentist approach is based on the notion of statistical regularity; i.e., in the long run, over replicates, the cumulative relative frequency of an event (E) stabilizes. The best way to illustrate this is with an example experiment that we run many times and measure the cumulative relative frequency (crf). The crf is simply the relative frequency computed cumulatively over some number of replicates of samples, each with a space S.

Let's take a look at an example of statistical regularity.

Suppose we have a treatment for high blood pressure. The event, E, we are interested in is successfully controlling the blood pressure. So, we want to be able to make a prediction about the probability that a patient treated in the future will have blood pressure under control, P(E). To estimate this probability we conduct an experiment that is replicated over time in months. The data are presented in the table below.

Month	**No. of subjects (S)**	**No. Controlled (E)**	**Cumulative S**	**Cumulative E**	**crf**
1	100	80	100	80	0.800
2	100	88	200	168	0.840
3	100	75	300	243	0.810
4	100	77	400	320	0.800
5	100	80	500	400	0.800
6	100	76	600	476	0.793
7	100	82	700	558	0.797
8	100	79	800	637	0.796
9	100	80	900	717	0.797
10	100	76	1000	793	0.793
11	100	77	1100	970	0.791
12	100	78	1200	948	0.790

The crf values down the right most column fluctuate the most in the beginning, but rapidly stabilize. Statistical regularity is the stabilization of the crf in the face of individual fluctuations form month to month in the relative frequency of E.

Finally, we are in a position where we can obtain a definition of probability. Here goes: In words, the probability of an event E, written as P(E), is the long run (cumulative) relative frequency of E. More formally we define P(E) as follows:

$$P(E) = \lim_{n \to \infty} crf\,(E)$$

In other words, If an experiment is performed repeatedly under essentially homogeneous & identical conditions, then the limiting value of the ratio of the number of times the event occurs to the number of trials, as the number of trials becomes indefinitely large, is called the **probability** of happening of the event, provided limit is finite and unique.

This definition also suffers from the following shortcomings :

ι.) The conditions of the experiment may not remain identical, particularly when the number of trials is sufficiently large.

ιι.) The relative frequency, may not attain a unique value no matter how large is the total number of trials.

ιιι.) It may not be possible to repeat an experiment a large number of times.

ιϖ.) Like the classical definition, this definition doesn't lead to any mathematical treatment of probability.

Axiomatic Probability: Definition, Kolmogorov's Three Axioms

During the XXth century, a Russian mathematician, Andrei Kolmogorov, proposed a definition of probability, which is the one that we keep on using nowadays. Axiomatic probability is a unifying probability theory. It sets down a set of axioms (rules) that apply to all of types of probability, including frequentist probability and classical probability. These rules, based on Kolmogorov's Three Axioms, set starting points for mathematical probability.

Kolmogorov's Three Axioms:

The three axioms are:

- For any event A, **P(A) ≥ 0.** In English, that's "For any event A, the probability of A is greater or equal to 0".
- When S is the sample space of an experiment; i.e., the set of all possible outcomes, **P(S) = 1**. In English, that's "The probability of any of the outcomes happening is one hundred percent", or—paraphrasing— "anytime this experiment is performed, something happens".
- If A and B are mutually exclusive outcomes, **P(A ∪ B) = P(A) + P(B).** Here ∪ stands for 'union'. We can read this by saying "If A and B are mutually exclusive outcomes, the probability of either A or B happening is the probability of A happening plus the probability of B happening"

Just because these axioms are universal, doesn't mean they provide all the answers. For example, any function that satisfies all three axioms is called a *probability function*. However, the axioms don't tell you which function to choose; it merely states that the probability function you choose must satisfy the rules.

Fine (2014) goes so far as to say the axioms lack "essential content". What these three axioms *don't* do:

- Tell us where and when to apply the rules,
- Give us guidelines or procedures for calculating probabilities,
- Any insights to the nature of random processes.

From axiom 3 it can be stated that $P(\phi) = 0$

If we need to prove this, let us take $B = \phi$ and make a note that

A and ϕ are disjoint events. Hence, from axiom 3 we can deduce that-

$P(A \cup \phi) = P(A) + P(\phi)$ or

$P(A)=P(A) + P(\phi)$

i.e. $P(\phi)=0$

Axiomatic Probability Example:

Now let us take a simple example to understand the axiomatic approach to probability.

On tossing a coin we say that the probability of occurrence of head and tail is ½ each. Basically here we are assigning the probability value of 1/2 for the occurrence of each event.

This condition basically satisfies both the conditions, i.e.

- Each value is neither less than zero nor greater than 1 and
- Sum of the probabilities of occurrence of head and tail is 1

Hence, for this case we can say that the probabilities of occurrence of head and tail are 1/2 each.
Now, say P(H) = 5/8 and P(T) = 3/8
Does this probability value satisfy the conditions of axiomatic approach?

For this, let us again check the basic initial conditions of the axiomatic approach of probability.

- Each value is neither less than zero nor greater than 1 and
- Sum of the probabilities of occurrence of head and tail is 1

Hence this sort of probability value assignment also satisfies the axiomatic approach of probability. Thus, we can conclude that there can be infinite ways to assign the probability to outcomes of an experiment.

Main properties of probability:

- $P(A)+P(\bar{A})=1$.

That is, the probabilities of complementary events add up to 1. Often we will use this property to calculate probability of the complementary set: $P(\bar{A})=1-P(A)$.

Let's see why. We know that, on the one hand, A and $\bar{A}$ are incompatible, and on the other that $A\cup\bar{A}=\Omega$, since one is the opposite of the other. This is another way of understanding what we already knew, i.e., that the event $A\cup\bar{A}$ is a sure event, and therefore, because of axiom 2, $P(A\cup\bar{A})=1$, it always happens. Then, for the axiom 3, $P(A\cup\bar{A})=P(A)+P(\bar{A})$. But $P(A\cup\bar{A})=P(\Omega)=1$, thus $P(A)+P(\bar{A})=1$.

This property, which turns out to be very useful, can be generalized:

If we have three or more events, two by two incompatible, and such that their union is the whole sample space, that is to say, A,B,C two by two incompatible so that $A\cup B\cup C=\Omega$, then $P(A)+P(B)+P(C)=1$, for axioms 2 and 3.

We say in this case that A,B,C form an events complete system. Let's observe that whenever we express Ω as a set of elementary events, in fact we are giving a complete system of events.

- If $A\subset B$, then $P(A)\leq P(B)$.

The notation "if $A\subset B$" reads "if the event A is included in event B" that is to say, if all the possible results that satisfy A also satisfy B.

This property is quite logical: if, after throwing a dice, we want to compare the probability of A="to extract 2" with B="to extract an even number", then, the probability of A has to be smaller or the same as that of B since if we extract 2, we are extracting an even number. In other words, when A is satisfied, B is also satisfied, therefore it should be more difficult to satisfy A than B. Namely $P(A)\leq P(B)$.

- $P(A\cup B)=P(A)+P(B)-P(A\cap B)$.

This result, which is very important to remember, is a consequence of something that you can see in the sets cell: given two sets, A and B, you can express its union as $A\cup B=(A-B)\cup(A\cap B)\cup(B-A)$, which are two by two incompatible. Then, for axiom 3, $P(A\cup B)=P(A-B)+P(A\cap B)+P(B-A)$.

In the Sets Theory we have that $A=(A-B)\cup(A\cap B)$, which are two incompatible events, and therefore, for axiom 3, $P(A)=P(A-B)+P(A\cap B)$, that is to say, $P(A-B)=P(A)-P(A\cap B)$.

Similarly,

$B=(B-A)\cup(B\cap A)=(B-A)\cup(A\cap B)$, by which $P(B-A)=P(B)-P(A\cap B)$.

Replacing these probabilities in the equality, we find

$P(A\cup B)=P(A-B)+P(A\cap B)+P(B-A)=P(A)-P(A\cap B)+P(A\cap B)+(P(B)-P(A\cap B))=P(A)+P(B)-P(A\cap B)$

Now, we can solve some problems.

Example

A dice of six faces is tailored so that the probability of getting every face is proportional to the number depicted on it.

1. What is the probability of extracting a 6?

In this case, we say that the probability of each face turning up is not the same. If we follow the statement, it says that the probability of each face turning up is proportional to the number of the face itself, and this means that, if we say that the probability of face 1 being turned up is k which we do not know, then:

$P(\{1\})=k$, $P(\{2\})=2k$, $P(\{3\})=3k$, $P(\{4\})=4k$,

$P(\{5\})=5k$,$P(\{6\})=6k$.

Now, since $\{1\},\{2\},\{3\},\{4\},\{5\},\{6\}$ form an events complete system , necessarily

$P(\{1\})+P(\{2\})+P(\{3\})+P(\{4\})+P(\{5\})+P(\{6\})=1$

Therefore $k+2k+3k+4k+5k+6k=1$which is an equation that we can already solve: $21k=1$ thus $k=1/21$

And so, the probability of extracting 6 is $P(\{6\})=6k=6\cdot 1/21=6/21$.

2. What is the probability of extracting an odd number?

The cases favourable to event A= "to extract an odd number" are: $\{1\},\{3\},\{5\}$. Therefore, since they are incompatible events,

$P(A)=P(\{1\})+P(\{3\})+P(\{5\})=k+3k+5k=9k=9 \cdot 1/21=9/21$

Example

Tomorrow there is an exam. Esther has studied really hard, and she only has 1/5 probability of not passing the exam.

David has studied less, and he has 1/3 probability of not passing the exam. We know that the probability of both not passing the exam is 1/8.

What is the probability that at least one of them does not pass the exam?

The first thing that we must do is express the problem as we know how, i.e., with events. We define the events A="Esther does not pass the exam", B="David does not pass the exam".

From the statement, we know that $P(A \cap B)=1/8$.

We might think that if Esther has probability 1/5 of not passing the exam, and David 1/3 of not passing the exam, then the probability of at least one of them not passing, that is to say $P(A \cup B)$, should be $1/5+1/3=8/15$, but this is false.

If we compute it this way, we are assuming that the events A and B are incompatible, that is to say, that they cannot happen simultaneously, when the statement says that they could both not pass (simultaneously).

Therefore, the correct way of calculating this probability is using the formula that we have seen before: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

by replacing with the results that we know, we get $P(A \cup B) = 1/5 + 1/3 - 1/8 = 24/120 + 40/120 - 15/120 = 49/120$.

Conditional Probability

The ***conditional probability*** of an event *B* is the probability that the event will occur given the knowledge that an event *A* has already occurred. This probability is written *P(B|A)*, notation for the *probability of B given A*. In the case where events *A* and *B* are *independent* (where event *A* has no effect on the probability of event *B*), the conditional probability of event *B* given event *A* is simply the probability of event *B*, that is *P(B)*.

If events *A* and *B* are not independent, then the probability of the *intersection of A and B* (the probability that both events occur) is defined by *P(A and B) = P(A)P(B|A).*

From this definition, the conditional probability *P(B|A)* is easily obtained by dividing by *P(A)*:

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

Examples

In a card game, suppose a player needs to draw two cards of the same suit in order to win. Of the 52 cards, there are 13 cards in each suit. Suppose first the player draws a heart. Now the player wishes to draw a second heart. Since one heart has already been chosen, there are now 12 hearts remaining in a deck of 51 cards. So the conditional probability *P(Draw second heart|First card a heart)* = 12/51.

Suppose an individual applying to a college determines that he has an 80% chance of being accepted, and he knows that dormitory housing will only be provided for 60% of all of the accepted students. The chance of the student being accepted *and* receiving dormitory housing is defined by *P(Accepted and Dormitory Housing) = P(Dormitory Housing|Accepted)P(Accepted)* = (0.60)*(0.80) = 0.48.

To calculate the probability of the intersection of more than two events, the conditional probabilities of *all* of the preceding events must be considered. In the case of three events, *A*, *B*, and *C*, the probability of the intersection *P(A and B and C) = P(A)P(B|A)P(C|A and B)*.

Consider the college applicant who has determined that he has 0.80 probability of acceptance and that only 60% of the accepted students will receive dormitory housing. Of the accepted students who receive dormitory housing, 80% will have at least one roommate. The probability of being accepted *and* receiving dormitory housing *and* having no roommates is calculated by: *P(Accepted and Dormitory Housing and No Roommates) = P(Accepted)P(Dormitory Housing|Accepted)P(No Roommates|Dormitory Housing and Accepted)* = (0.80)*(0.60)*(0.20) = 0.096. The student has about a 10% chance of receiving a single room at the college.