Axioms of probability

Math 477

September 8, 2014

1 Sample space and events

Consider an experiment where we toss a coin twice. All the possible outcomes are

$$\{TT\}, \{HH\}, \{TH\}, \{HT\}.$$

We call these (elementary) events. The events have the following properties:

a. The union of two events is an event:

$$\{TT\} \cup \{TH\} = \{\text{First toss is } T\}.$$

b. The intersection of two events is an event:

$$\{\text{First toss is } T\} \cap \{\text{Second toss is } T\} = \{TT\}.$$

c. The complement of an event is an event:

$$\{TT\}^c = \{ \text{ At least one of the toss is } H\}.$$

Note: In everyday language, union corresponds to OR, intersection corresponds to AND, complement corresponds to NOT.

Suppose we toss a coin n times. It is not difficult to see that that more generally we have the followings:

a'. The union of finitely many events is an event: The event $\{\text{First toss is } T\}$ is the union of finitely many events where each of them has the form $\{T \cdots\}$.

b'. The intersection of finitely many events is an event: The event $\{\text{All tosses are } T\}$ is the intersection of n events where each of them has the form $\{ \text{ The nth toss is } T\}$.

Suppose we toss a coin indefinitely. Then we have the followings:

a". The union of (countably) infinitely many events is an event: The event {We eventually see a T} is the (countable) union of events of the form { The nth toss is $T, n = 1, 2 \cdots$ }.

b". The intersection of (countably) infinitely many events is an event: The event {All the even toss is T} is the (countable) intersection of events of the form { The nth toss is $T, n = 2, 4, 6 \cdots$ }.

Terminology: When two events have nothing in common (their intersection is $\emptyset$, the empty set) we say they are *mutually exclusive*. For example, the two events {First toss is H} and {First toss is T} are mutally exclusive.

Abstractly, we use capital letters at the beginning of the alphabet: A, B or $E_1, E_2 \cdots$ to denote an event. We also see that in the examples above, an outcome (or an elementary event) is an event that has no sub-event contained in it (in other words, a smallest possible event). We will also denote the intersection of two events A, B as AB (instead of $A \cap B$).

The set of all possible outcomes of an experiment is referred to as *the sample space* and denoted as S. Another very popular notation for the sample space that you can find in other textbooks is Ω. We may use Ω for sample space later on in the lecture notes.

2 Elementary set operations

Since events are basically *subsets* of the sample space, one way to create or describe new events is to *apply set operation* on basic events. We have seen some simple examples above, for example the event A or B is created out of their union and the event A and B is created out of their intersection. But the operations can be more complicated, so we list some more elementary operations and their properties here. You should try to see why they are true.

a. Commutative laws:

$$E \cup F = F \cup E, EF = FE.$$

b. Associative law:

$$(E \cup F) \cup G = E \cup (F \cup G), (EF)G = E(FG).$$

c. Distributive law:

$$(E \cup F)G = EG \cup FG, EF \cup G = (E \cup G)(F \cup G).$$

d. DeMorgan's law:

$$\left(\cup_1^n E_i\right)^c = \cap_1^n E_i^c, \left(\cap_1^n E_i\right)^c = \cup_1^n E_i^c.$$

3 Three axioms of probability

3.1 Introduction

So far we have defined events and discussed their elementary properties. Now one basic thing we do when we observe events is that we associate some likelihood to those events. This allows us to say things like this event is as likely as (or more likely than, or less likely than) the other event. For example, if we believe a coin is fair, we say the event of getting a head is as likely as getting a tail. Or if we toss 2 dice (assuming fair) we say the event of getting a sum of 7 is more likely than getting a sum of 11. Or (this is a well-known quote from the movie Good Will Hunting, and it's true, statistically speaking) that the event of getting hit by lightning is more likely than the event of winning a lottery. Implicit in all these statements is a belief that we can *associate a quantity to the likelihood of each event*, and compare them. These quantities are what we called probabilities and denote the probability of an event A as $P(A)$. But this association *cannot be arbitrary*, that is they have to follow certain basic laws. To give a simple example, if an event A is contained in another event B, the probability of A must be smaller than the probability of B (the probability of getting a 2 is smaller than getting an even number when tossing a die). Another is a probability cannot be negative (it doesn't make sense to talk about negative likelihood). We call these *axiom of probabilities*, which are the most basic laws that proability of events have to follow. Then we derive other consequences that the probabilities have to obey, from the axioms, and verify that they indeed correspond to our intuition about how likelihood should be have.

3.2 The axioms

1. For any event E,

$$0 \leq P(E) \leq 1.$$

This basically limits the range of values that a probability can take to be from 0 to 1.

2.

$$P(S) = 1.$$

One may read this roughly as the probability that something happens is 1.

3. For any sequence of mutually exclusive events $E_1, E_2 \cdots$, that is $E_i E_j = \emptyset$ when $i \neq j$

$$P(\cup_1^\infty E_i) = \sum_{i=1}^{\infty} P(E_i).$$

3.3 Some remarks

The axioms of probability might strike you as missing something, since nowhere does it mention *what value* $P(E)$ you should assign for a particular event E. It only says that IF you assign probabilities to events in your sample space, then they have to follow 3 basic laws. So on the one hand the assignment cannot be too arbitrary, but on the other hand it seems like you have quite a bit of freedom to choose which particular probabilities to assign to which event. For example, in the experiment of tossing a coin, you can verify that there are infinitely many ways to assign probability to this sample space $\{H, T\}$, namely just assign probability p to the event $\{H\}$ and $1 - p$ to the event $\{T\}$, for any number $p \in [0, 1]$. But from your experience perhaps, you feel like one should choose $p = 1/2$ as the obvious value, unless one has reason to doubt the coin is fair. On what basis do you base this belief in? It would go something like this: if you toss the coin a large number of times, say n, and count how many heads you obtain, say $n(H)$, then you have

$$1/2 \approx \frac{n(H)}{n}.$$

In other words, you *propose to define*

$$P(H) = \lim_{n \to \infty} \frac{n(H)}{n}.$$

This may seem good at the beginning, but indeed this definition is problematic. The problem lies in the fact that how do we know the quantity $\lim_{n \to \infty} \frac{n(H)}{n}$ converges? You can indeed *assume* that it converges as a starting point of your theory, but this would certainly *impose* quite a bit of assumption on nature, and that's not what a good theory is about. So this is why from the basic axioms of probability, we do not say anything about what the particular value of probability of an event is.

Nevertheless, the intuition that we do obtain the probability of the event, as the ratio of the number of its outcomes over the number of the experiments, as the number of the experiments goes to infinity, is a sound intuition, *which can be proved* via the famous Law of Large Number, based on the axioms of probabilities. We will cover this Theorem later on in this class. Thus you see that the axioms of probability are indeed "good axioms," in the sense that we can derive theorems that correspond to our intuition via these axioms. The bottom line is that as far as Probability theory is concerned, one can assign any probability distribution on the events, as long as they obey the basic axioms. That is one can assign probability p to the event $\{H\}$ for any $p \in [0, 1]$. In this way p represents *our belief about the fairness of the coin.* The question whether or not they *correspond* to the "true" probability of the physical object (whether $p = 1/2$ if the coin is indeed fair) (a question that is studied by Statistics) can be settled by doing lots and lots of experiment and invoke the Law of Large Number.