MAT 3701: Axioms of Probability Theory

January 21, 2013

In order for us to apply the rigorous methods of deductive logic, every mathematical subject must be placed on a firm foundation of axioms. Probability is no exception. Having gained some intuition through experience with probabilistic problems, we are now ready to discuss the axiomatic foundation of probability.

First let us recall how this is done for other subjects with which you are familiar. If you have studied geometry, you know that a theory of geometry requires some undefined terms, such as "point", "line", "incidence", "betweenness", and "congruence". All of these undefined terms are sets of some form or another, set theory being the basis of the categorization that underlies all of mathematics. Points are the elements of the set of points; lines are elements of the set of lines. Incidence is a relation between points and lines, which, as you may know, can be written as a subset of the Cartesian product of the set of points with the set of lines; an ordered pair (P, l) consisting of a point and a line is in this subset if and only if the point P is incident with (that is, lies on) the line l.

We don't know what the elements of these sets are - that's why they are undefined, but we assume they satisfy certain axioms. For example, we assume that given any two points, there is a unique line incident with both of them.

We also add definitions that are based on these undefined terms. For example, we define segment AB to be the set containing point A, point B, and all of the points between A and B. The axioms may involve both defined and undefined terms. Congruence, for example, is a relation between segments, so it cannot even be discussed until segments are defined.

Finally, we can *apply* the theory to sets of known elements that satisfy its axioms. For example, we can take ordered pairs of real numbers to be the points and the solution sets to equations of the form $ax + by + c = 0$ to be the lines; thus, lines are specific sets of points. A given point is considered to be incident with a given line if it is an element of that set. We can similarly define what it means for one point to be between two other points and what it means for two segments or angles to be congruent, in such a way that all of the axioms of plane geometry are satisfied. The virtue of this is that we now know that every theorem proved about plane geometry is true about the (x, y)-plane with these definitions. We don't have to prove them again in this special situation.

We can even add axioms to the basic ones to create special types of geometry. For example, if we postulate that given a line and a point not on that line, there is at most one line through the given point that is parallel to the given line, then we have Euclidean geometry. (Note: "parallel" is a defined term. What is its definition? Geometry without a postulate on parallelism is called *neutral* geometry.) Alternatively, we can remove some of the axioms to get more general "geometries," such as *incidence geometry*, in which incidence is the only relation and only the axioms pertaining to incidence must be satisfied.

Another example of a theory is *group theory*, which you study in abstract algebra. A *group* consists of a set together with a binary operation and a specific element called a *unit*. Since neither the set whose elements make up the group, the group operation, nor the unit are specifically defined, this is just another way of saying that group theory has three undefined terms. Using these, we can define some other useful concepts, such as *inverse* and *subgroup*. Given an element g of the group, and denoting the unit of the group by e, an *inverse* for g is an element h such $gh = hg = e$.

The set underlying the group, its operation, and its unit satisfy certain axioms. That is, the properties in the definition of a group are really the axioms of group theory. Actually, there aren't very many: the set must be closed under the operation; the operation must be associative; the unit, e, has the property that for every element g of the group, $eg = ge = g$; and every element g must have an inverse. (The associativity of the group operation allows us to prove that an element has only one inverse, and also that there is only one unit. Try it!) Models of groups include the integers with addition as the operation and 0 as the unit; the set of powers of 2 with multiplication the operation and 1 as the unit; or the set of invertible 2×2 real-valued matrices with matrix multiplication as the operation and the identity matrix (the one whose first row is 1 0 and whose second row is 0 1) as the unit.

What makes abstract algebra "abstract" - and powerful - is that it describes the properties of undefined, "general" groups, which can then be applied to all the myriad examples. The reason that we define "a" group, but don't usually speak of "a" geometry (or, for that matter, "abstract geometry"), is probably that there are only two models of (plane) geometry that are truly different [1], and only one if we specify an axiom of parallelism, whereas there are infinitely many models of a group. So we talk simply of "geometry," or "Euclidean" geometry, or "hyperbolic" geometry, not "a" geometry. But this difference is only one of style; the idea behind how the theory is constructed is exactly the same.

Just as with geometry, we can add additional axioms to group theory. If we postulate that the operation is commutative, for example, then our group is a commutative group. The first two examples given above are commutative; the third is not. Even though group theory has very few axioms, we can also take some away. For example, if we dispense with requiring inverses, we have a *monoid*.

Now that we've reviewed the notion and purpose of an axiomatic theory, we are ready to precisely define a *probability space*, which is the fundamental object of study in probability theory. It consists of a set, called the *sample space*, whose elements are referred to as "outcomes," and a *probability measure* on some of its subsets, which satisfies a few axioms reminiscent of measuring lengths, areas, or volumes. As we will see from one of the homework problems, if the sample space is uncountably infinite (that is, its elements cannot be listed in even an infinite sequence), it is not reasonable to expect that every subset of the sample space can be measured. But if a set can be measured, it's complement certainly can, and the union or intersection of two measurable sets should be measurable. Moreover, countable unions of measurable sets should be measurable, as they are in the case of the coin flipped until a head arises. Hence we formulate the following definition in set theory:

Definition. *Let S be a set, and denote its power set by $\mathcal{P}(S)$. A* σ-algebra *of subsets of S is a subset Σ of $\mathcal{P}(S)$ satisfying the following properties:*

1. *If $A \in \Sigma$, then $A^c \in \Sigma$ (where A^c denotes the complement of A).*

[1] By truly different, we mean not *isomorphic*. An isomorphism between models is a 1-1 correspondence between their elements that preserves all the operations and relations. You probably recall that an isomorphism between two groups, G and G', is a 1-1 and onto function $\phi : G \to G'$ (as sets) such that $\phi(gh) = \phi(g)\phi(h)$. This is what is meant by preserving the operation: if you perform the operation on a pair of elements in the first group and see what the result corresponds to, or if you first map the two elements to the second group and perform the operation on their images, it doesn't matter, you get the same answer. (Note that the map $n \to 2^n$ is an isomorphism between the additive integers and the multiplicative powers of two.) There are infinitely many non-isorphic groups, but only two non-isomorphic neutral geometries.

2. *If* $A_n \in \Sigma$ *for* $n = 1, 2, 3, \ldots$, *then* $\cup_{n=1}^{\infty} A_n \in \Sigma$, *and* $\cap_{n=1}^{\infty} A_n \in \Sigma$.

Remark. Observe that finite unions and intersections are covered by property (2), since if $A_n = \emptyset$ for $n > k$, then $\cup_{n=1}^{\infty} A_n = A_1 \cup A_2 \cup \ldots \cup A_k$, and if $A_n = S$ for $n > k$, then $\cap_{n=1}^{\infty} A_n = A_1 \cap A_2 \cap \ldots \cap A_k$.

Remark. Observe as well that it is not necessary to assume closure under both unions and intersections in Property (2); either is sufficient, by DeMorgan's Laws.

Note on terminology. The symbol σ is used to refer to countably infinite set operations. A collection of subsets that is closed under complementation and finite unions and intersections is called just an *algebra*. Sometimes the word "field" is used instead of "algebra". The terminology is pretty inconsistent, so when reading a text, check the definitions given.

Definition. *Let* Σ *be a* σ*-algebra of subsets of a set* S. *A* probability measure *on* Σ *is a function* $\rho : \Sigma \to \mathbb{R}$ *such that:*

1. $\rho(S) = 1$.
2. *For every* $A \in \Sigma$, $\rho(A) \geq 0$.
3. $A_1, A_2, A_3, \ldots$ *are pairwise disjoint sets in* Σ, *then* $\rho(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \rho(A_n)$.

We can now formally define a probability space:

Definition. *A* probability space *consists of a set* S, *called the sample space, a* σ*-algebra,* Σ, *of subsets of* S, *and a probability measure* ρ *on* Σ. *The elements of* Σ *(in other words, the subsets of* S *whose probabilities can be measured) are called* events.

We can now make one further definition. Note that all of our definitions, including the following one, precisely define concepts that we have been using intuitively.

Definition. *Let* A *and* B *be events such that* $\rho(A) > 0$. *The* conditional probability of B given A, *denoted* $\rho(B|A)$ *is defined to be*

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

.