Probability Unit 2

Topic: Events, Set algebra, Conditioning

- Date: @February 1, 2024
- Lane: *Probability*

What it covers

- Events and sets as the domain of discrete probability; set algebra
- Discrete probability. Axiomatic approach 2 axioms, finite additivity and unit.
- Concept of a measure of a set (why a measure is not a function)
- Measures imply finite additivity
- Probabilities are “unit” measures. e.g. 0< P <1
- Algebra of probabilities over sets of events
- Events form a Boolean Algebra
- All probabilities are conditional
- Bayes Rule “flips” conditionings

Requires

Naive definition of probability. Set theory

Required by

All probability modules including definitions of Discrete and continuous probability distributions.

A better definition of Probability

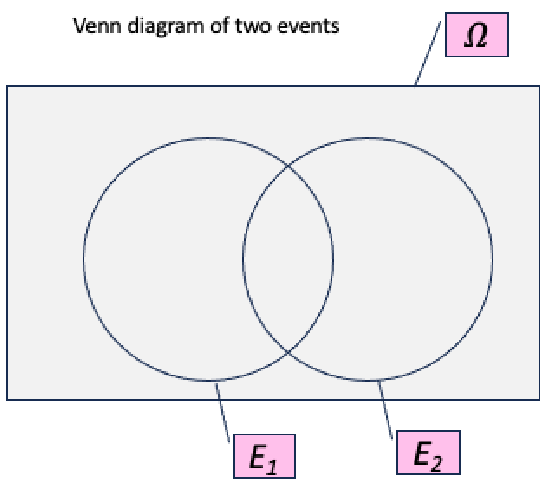
Here’s a case where naive probability doesn’t apply. Using records of activity, D, say customer visits on each day of the year for several past years, one creates a probability distribution over the 365 days of the year that vary from day to day. Then given this distribution *P(D),* one uses the probability of a sale given a visit to compute sales for the year. The “principle of indifference” does not apply to the distribution *P(D)* since there’s additional knowledge about how probabilities change day to day. $P(S = success|D)$ Conditioning one probability on another event naturally follows when drawing a probability tree. Computing the probability *P(S)* is equivalent to “rolling back” the tree.

$$P(S) = \sum_i P(S = success|D)P(D_i)$$

Events

Technically an *event* E is a subset of possibilities (often called "successes") out of all possibilities — the so called "sample space." Examples are "3 heads out of 5 coin-tosses", or "a team with only Californians chosen at random from the class." Events are combined using set union, intersection, set difference, and subset operations: $E_1 \cup E_2, E_1 \cap E_2, E_1 \setminus E_2, E_1 \subset E_2$. The certain event Ω equals the entire sample space, so the complement of an event $E^c = \Omega \setminus E$. The complement of the entire sample space is the null set. $\varnothing = \Omega^c$, which represents an impossible event. (Note that the "items" in the sample space may be infinite, but we restrict events to a countable collection.)

Venn Diagrams visualize these operations:



Technically a probability is a *measure* — that is a number applied to a set. Following our intuitions, $P(\Omega) = 1$, the fraction of times the certain event occurs is one. Also, for mutually exclusive events (also called "disjoint") probabilities are additive:

$$\text{if } E_1 \cap E_2 = \varnothing,\ P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

It follows that an impossible event has zero probability. $P(\varnothing) = 0$. This property, of *finite additivity* can be extended to series of mutually exclusive events, $E_1 \cup E_2 \cup \ldots$ to generate the complete calculus of probability. In short, it is all that's needed to build the entire theory. Just to note, "fuzzy logic" starts similarly, but does not include finite additivity, and loses much of the power of reasoning with probability.

Some Boolean Algebra

"Boolean Algebra" refers to the 2-valued algebra of {T,F} with logical operations of "and", "or" and "not". Basic proofs use *Truth Tables.* Events can be constructed from other events, to which a probability measure can be appllied - by the rules for assigning probabilities to the logical operators.

Events form a "discrete sample space" — the entities we assign probabilities to, or in other words "have a non-zero measure. An infinite sample space (e.g. the real number line) creates a dilemma — how to create a non-zero proportion out of infinitesimals of an infinite set? This is why we start with discrete spaces, then show how continuous spaces just reduce to them.

Conditioning

Referring to the definition of probability where the denominator counts all possible events, a *conditional probability* “rebases” the proportion by replacing the denominator with another event:

$$P(E_1 \mid E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

To make this work, the numerator needs to include only those parts of the sample space common to both events.

In concept all probabilities are conditioned, but by convention the conditioning is left out when the conditioning is the entire state space: $P(E \mid \Omega) = P(E)$.

Total probability Rule

Conditional probabilities are useful for “divide and conquer” estimates of probabilities. Given estimates for the probability of an event $E = e$ for each partition *i* of a conditioning event C an estimate of the “total probability of *E is*

$$P(E) = \sum_i P(E \mid C = c_i)P(C = c_i)$$

Note the equivalence of this formula to “rolling back the tree.”

Bayes Rule

Conditionings can be “flipped” as shown by a bit of algebra:

$$P(E_1 \mid E_2)P(E_2) = P(E_1 \cap E_2) = P(E_2 \mid E_1)P(E_1)$$

Or equivalently, in what is known as *Bayes’ Rule.*

$$P(E_1 \mid E_2) = \frac{P(E_2 \mid E_1)P(E_1)}{P(E_2)}$$

The school of Bayesian statistics bases inference on this relation, as a way to apply data to a hypothesis. In that usage, the terms in Bayes’ Rule are named,

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal}}$$

References

Evans, Ch 1.2 and 1.3