Probability

Christopher A. Sims
Princeton University
sims@princeton.edu

September 2, 2020

What it is

- It's a mathematical construct: a function mapping sets into the [0,1] interval.
- As we'll discuss, it can have many interpretations and applications.
- But to start we need to define it precisely as math.

Finite state space case

- We specify a set S containing a finite number of points $s_i, i = 1, \ldots, N$.
- The points are called "states", and S is called the "state space".
- A probability defined on S is a function P that maps subsets of S into [0,1].
- It must in addition have these properties:
  1. $P(S) = 1$.
  2. If $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$.

What sets $A \subset S$ is P defined for?

- In the finite state space case, most commonly we just assume that every subset of S has a well-defined probability.

- But there are cases where P is defined only for certain subsets of S.

- In that case, if we call the collection of subsets for which P is defined $\mathcal{F}$, we require:

  – If A and B are both in $\mathcal{F}$, then so are $A \cup B$ and $A \cap B$.
  – If A is in $\mathcal{F}$, then so is A^c.

  This makes $\mathcal{F}$ a **field** of sets.

Example of $\mathcal{F} \neq 2^S$

$$S: \quad \{1, 2, 3, 4\}$$

$$\mathcal{F}: \quad \{\{1, 2, 3\}, \{3, 4\}, \{3\}, \{4\}, \{1, 2\}, \{1, 2, 4\}, \emptyset, \{1, 2, 3, 4\}\}$$

Random variables and expectations

- A random variable is a function that maps points in S to the real line $\mathbb{R}$.
- We can also consider random vectors (functions that map into $\mathbb{R}^n$), random sets (functions that map into collections of sets), random sequences, etc.
- Random variables have **expectations**. If X is a random variable defined on S, and S is a finite set, with $\mathcal{F}$ simply 2^S, the expectation is defined as

$$E[X] = \sum_{j=1}^{N} P[s_j] X(s_j) \, .$$

Expectation when $\mathcal{F}$ is not 2^S

- This means that $P[s_j]$ is not defined for every point s_j in S.
- In that case $E[X]$ is not defined for every possible random variable X.
- Let ξ_i, $i = 1, \ldots m$ be the finite set of values taken on by $X(s)$ as s ranges over S.
- Another way to write the expectation is then

$$E[X] = \sum_{i=1}^{m} \xi_i P[X = \xi_i] \, .$$

- Here $P[X = \xi_i]$ is short for $P[\{s \in S \mid X(s) = \xi_i\}]$.

- This version of the definition makes it clear that the expectation is defined only if every set of the form $\{s \in S \mid X(s) = \xi_i\}$ is in $\mathcal{F}$.

- If X does not satisfy this condition, we say X is not **measurable** with respect to P, or its associated $\mathcal{F}$.

- Notice: $E[\ ]$ is linear. For two random variables X and Y and real number constants a and b, $E[aX + bY] = aE[X] + bE[Y]$.

Conditional probability

The conditional probability of the set B given the set A is

$$P[B \mid A] = \frac{P[B \cap A]}{P[A]}\,.$$

Example:

Tomorrow ╲ Today	rain	sun
rain	.15	.35
sun	.10	.40

$$\therefore P[\text{rain tomorrow} \mid \text{rain today}] = \frac{.15}{.15 + .10} = .6\,,$$

Though unconditionally $P[\text{rain tomoorrow}] = .5$

Conditional expectation

- For any set A with $P[A] > 0$, $P[\ \mid A]$ as a function on the sets in $\mathcal{F}$ is itself a probability function.

- For a random variable X measurable with respect to $\mathcal{F}$, the conditional expectation of X given the set A is $E[X \mid A]$, defined just as we defined ordinary expectation, only now with $P[\ \mid A]$ being used in place of $P[\ ]$.

Conditional expectation of one random variable given another

- Suppose X and Y are two random variables, both defined relative to the same $S, \mathcal{F}, P$ system. We write $E[Y \mid X]$ for the expectation of Y given X.

- This notation is potentially confusing, since X is not a set, and above we have defined conditional expectation only given sets. $E[Y \mid X]$ is a function of X, and hence itself a random variable. $E[Y \mid X]$ is the function that, when evaluated at $X = x$, is

$$E[Y \mid X = x] = E[Y \mid \{s \in S \mid X(s) = x\}]$$

The law of iterated expectations

- Once we consider $E[Y \mid X]$ as a random variable, we can conclude that

$$E[E[Y \mid X]] = E[Y] \,.$$

- To see this, let $\{\zeta_k\}$ be the set of values taken on by Y and (as before) $\{\xi_j\}$ be the set of values taken on by X. Then

$$E[E[Y \mid X]] = \sum_j P[X = \xi_j] E[Y \mid X = \xi_j]$$

$$= \sum_j P[X = \xi_j] \sum_k \zeta_k P[Y = \zeta_k \mid X = \xi_j]$$

$$= \sum_k \zeta_k \sum_j P[X = \xi_j] \frac{P[Y = \zeta_j \text{ and } X = \xi_j]}{P[X = \xi_j]}$$

$$= \sum_k \zeta_k \sum_j P[X = \xi_j \text{ and } Y = \zeta_k]$$

$$= \sum_k \zeta_k j P[Y = \zeta_j] = E[Y]$$

A defining property of expectation conditional on a random variable

- For any function $f(X)$, so long as $E[Yf(X)]$ exists,

$$E[Yf(X)] = E[E[Yf(X) \mid X]] = E[f(X)E[Y \mid X]]$$

- In fact, one can use this property to define conditional expectation and conditional probability: $E[Y \mid X]$ is the function $g()$ of X such that for any random variable Y and any function $f(X)$ such that $E[Yf(X)]$ exists, $E[Yf(X)] = E[f(X)g(X)]$. Then $P[A \mid X] = E[\mathbf{1}_A \mid X]$, where $\mathbf{1}_A$ is the indicator function for the set A.

Bayes' Rule

- This applies to a situation where we know $P[Y \mid X = \xi]$ for every possible value X, we get to observe the value of Y, but we don't see X. We would like to form a probability distribution for Y based on our observation of Y.

- To do this, we need to know the distribution of X. If we do know that, it is a trivial application of the definition of conditional probability that

$$P[X = \xi \mid Y = \zeta] = \frac{P[X = \xi]P[Y = \zeta \mid X = \xi]}{P[Y = \zeta]}\,.$$

- Where do we get $P[Y = \zeta]$? The law of iterated expectations tells us that

$$P[Y = \zeta] = E[P[Y = \zeta \mid X] \,,$$

and we've assumed we know the $P[Y \mid X]$ function and the distribution of X.

Currently relevant example of Bayes' rule

- X is 1 if a person is sick, zero if not.
- Y is a test outcome, with $Y = 1$ a "positive" test and $Y = 0$ "negative.
- We might know that $P[Y = 1 \mid X = 1] = .96$ and $P[Y = 0 \mid X = 0] = .98]$.
- Suppose the person in question tests positive. What is the probability she is sick, given the test result?

The “base rate fallacy”

- The test is guaranteed to give the correct answer with probability at least .96 whether the person is sick or well.

- It might seem then, that if the test is positive the person must be sick with probability at least .96.

The right answer

- We need to know $P[X = 1]$, the probability she is sick before we see the test outcome. That is, if she is drawn randomly from a population, the proportion of that population that is sick.

- So if 3% of the population she is drawn from are sick, the probability that she is sick given the test outcome is

$$\frac{.03 \times .96}{.03 \times .96 + .97 \times .02} = .60\,.$$

- The test does give the right answer with probability $.97 \times .98 + .03 \times .96 = .9794$, It is wrong with propability .4 **when it gives a positive result**, but these positive results are unlikely. Well people are so much more common than sick people, that false positives occur nearly as often as true positives.

Independence

- Two random variables X and Y are independent iff $E[f(Y) \mid X] = E[f(Y)]$ for any function of Y with well defined expectation.

- This means also that $P[Y \in A \mid X \in B] = P[Y \in A]$. I.e. the conditional distribution of Y given X does not depend on X.

- This means also that independence implies $E[f(X)g(Y)] = E[f(X)]E[g(Y)]$, so long as the expectations exist.

- Note that $P[Y \in A \mid X \in B] = P[Y \in A]$ implies $P[X \in B \mid Y \in A] = P[X \in B]$, so the definition of independence is symmetric in Y and X.

Mutual independence

- A collection $\{X_1, \ldots, X_n\}$ of random variables is mutually independent if for every j and every function $f(X_j)$ with finite expectation, $E[f(X_j) \mid X_{-j}] = E[f(X)]$, where X_{-j} is all the X_i's with $i \neq j$.
- That X_i is independent of X_j, pairwise, for every $i \neq j$, does not in general imply mutual independence of the whole collection.
- For the most commonly encountered joint distribution, the multivariate normal, pairwise independence does imply mutual independence, but this is not true in general.
- Simplest example: Y and X independent, each equal to 1 or -1 with probability .5. $Z = XY$.

Exchangeability

- A collection of random variables $X_1 \ldots, X_n$ is **exchangeable** iff their joint distribution does not depend on how they are ordered.

- If they all have the same distribution ($P[X_i \in A]$ does not depend on i, for any set $A \in \mathcal{F}$) and are mutually independent, they are exchangeable.

- If there is a random variable Z such that the conditional distribution of $\{X_1, \ldots, X_n\}$ given Z makes the X's mutually independent and identically distributed, then the X's are exchangeable.