Probability Unit 3

Topic: Random variables, Discrete Distributions, Expectation

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

What it covers

- A random variable - a function from $\Omega \to \mathbb{R}$
- Indicator variables
- Expectation $\sum xP(X = x)$
- An example of an iterated expectation

Requires

Definition of probability.

Required by

HW on histograms

Definition of continuous probability distributions.

Probability: Definition of a Random Variable

To make the theory useful we need to add one more thing; A way to assign numbers to events. These are called *random variables.* We've seen this intuitively for example when the throws of a die are labelled with the value on the face of the die. In general -

> 💡 A random variable is a mapping from an item in the state space to a number on the real line.

Technically its a *function*, not a “variable”. With the discrete events introduced so far, this is simple, almost trivial, since the state space and events have often been one-to-one; and the idea of labeling them with numeric values is obvious. Simply, we need both a probability and value associated with an event to do calculations. The purpose of random variables will be obvious when we consider the relationship of continuous distributions to sample space, events and random variables.

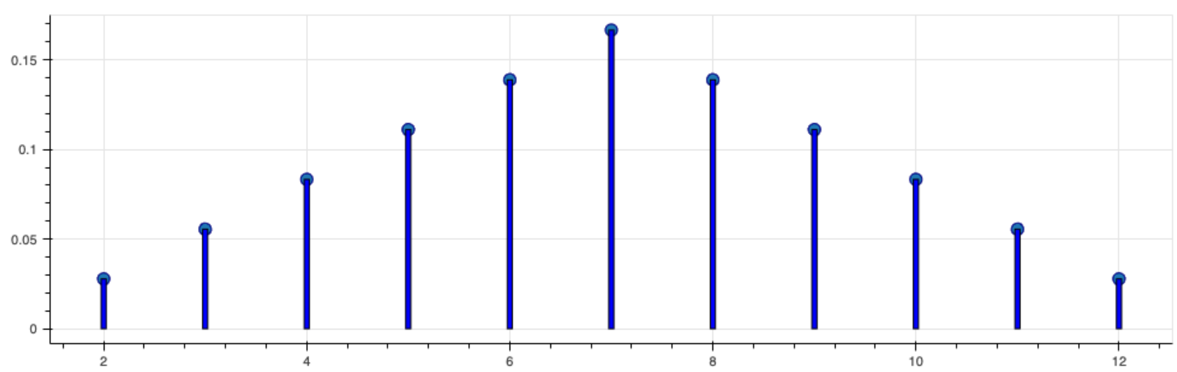
Incidentally, this notion of pairing two numbers, one intensive (probability) and one extensive (value) is a “duality” that appears in many sciences. Mass & velocity, Voltage & current, Force and distance.

We can be creative with how random variables are assigned to events, depending on what we are trying to model. The set algebra of events does _not_ imply a direct way to assign random variables to combinations of events. For instance, if we combine a set of coin flips, we can count the number of heads as a random variable, or whether a head appeared at all, or some other value to some pattern of flips.

The notation for probability carries over for random variables, so we can write $P(X)$ to mean the probability over the points that map into *X*. To be precise we should define the events using the random variable. For instance, to indicate the probability of the random variable having a certain value *x*, we create events from random variables like this: $P(X = x)$. Other events could be defined by algebraic expressions like $P(X > k), P(X > Y), P(X = x \text{ and } Y = y)$ for example. Conditioning on events defined by random variables are also possible.

Distribution of Random Variables

The expression $P(X)$ for all values of *X* is called the *distribution* of *X*. We can represent this by its *probability mass function (PMF)* $p_X(x)$. Since *X* is discrete (an integer) this PMF expresses a *discrete distribution.* For the sum of the throws of a pair of dice it should look familiar:



The distinction between $P(X)$ and $p_X(x)$ is subtle (the first is a *measure,* the second a *function,* and the distinction is often ignored. The fundamental properties of probability measures also apply to distributions. e.g. $\sum_x p_X(x) = 1$.

Computing with Random Variables

For any random variable distribution, we can take an "expectation" to create an equivalent certain value. This is also true for any function of random variables. This is similar to finding a fair price for a trade where there are unknowns. In probability the "probability-weighted sum" of the random variables define a number — the distribution's expectation:

$$E[X] = \sum_x x\, p_X(x) = \sum_x x\, P(X = x)$$

Expectation is a linear transformation, $E[aX + Y] = aE[X] + E[Y]$. Similarly a conditional expectation is defined from the conditional probability distribution. $E[X \mid A] = \sum_x x\, P(X = x \mid Y)$. Note this is a function; it has a value for each value of the random variable *Y*. Just to complete the toolkit, probability itself can be defined as the expectation of an indicator variable of an event. An *indicator function* returns 1 for "true" and 0 for "false." So, pedantically one can write $P(I(E) = 1) = P(E)$. The "fundamental bridge" between expectation and probability is made using indicator functions::

$$E[I(A)] = P(A = true)$$

Estimating Expectation

When working with data, given a sample $x_0, x_1, ...x_n$, the "plug in" estimate of the expectation is the average (also called the *mean)* of the sample. $\overline{x} = 1/n \sum_i x_i$. The bar above the *x* is shorthand for "estimated expectation." This is equivalent to counting the number of occurrences of each value of *x*, to estimate the distribution, by which to weight the sum.

Examples

"Rolling back" a Probability Tree.

Compute the expected cost of serving a person using the "vegan / non-vegan" tree from HW one. Assume $v = P(\text{vegan}), c = P(\text{cheesecake} \mid \text{non-vegan}), \text{cheesecake} = \$10, \text{sherbert} = \$1, \text{vegan} = \$12, \text{non-vegan} = \$20.$

$$E[a\ serving] = 21(1-v)(1-c) + 30(1-v)c + 13v$$

A "recurrent" tree

Compute the expected duration of an errant system process. In each second process can either die with $P(\text{die}) = p$,or continue, with probability $1 - p$, at which point the process lives for another second, and repeats the choice to live or die with the same probability. Eventually the process will die. The tree to express this continues to branch until it does, but to solve this we take advantage that the expectation at the root of the tree is the same as at the "live" leaf:

$$E[\text{duration}] = 1 * P(\text{die}) + (1 + E[\text{duration}]) * P(\text{live})$$

$$E[d] = p + (1 + E[d]) * (1-p), E[d] = \frac{(1-p)}{p}$$

The lifetime of this process follows a *geometric distribution.*

References

Evans, Ch 2.1, 2.2, 2.3