📖

Probability Unit 1

Topic: *Counting*

- **Date:** @January 25, 2024
- **Lane:** *Probability*
- **What it covers**

Sampling, Probability trees, permutations, binomial distribution, pascal's rule Compare the binomial to the negative binomial — same distribution, different stopping rule.

Exchangeable items. It makes them countable.

- **Requires:**

Previous take home assignment on counting the throws of a pair of dice.

- **Required by**

Perturbations are used by *Matrix determinants*

Binomial coefficients are used by *Binomial distributions*

Probability: What is the question?

What is probability? — One definition: *Naive probability* as an observed frequency. How to count the number of "desired" events out of all possible.

Probability as a ratio of counted events per possible events

Outcome set. (or "Sample space") We start by defining a set of discrete possibilities. They can be enumerated, that means assigned labels, here assumed to be a finite set. The possible outcomes are *events* that are *mutually exclusive* — if one occurs, then no other can occur; and *exhaustive* - the cover all possibilities.
We say the events
Partition the sample space.
We designate some subset of the outcomes as events we are interested in — call them "successes" or desirable events.

The definition of such a situation is called an **Experiment**

$$P(desired\_events|all\_possible\_events) = \frac{\text{Count of desired outcom}}{\text{Count of all possible outc}}$$

Note the vertical bar "|" in the middle of the definition of P(). On the left is the definition of the set of outcomes whose probability we are *measuring*. On the right, the *conditioning* set. Hence this is called a *conditional probability.* In principle all probabilities are conditioned by assumptions about the set of outcomes they are drawn from. In practice we will often drop the conditioning argument if it is understood, and just write: $P(desired\_events)$.

Combining Independent probabilities

Probability Trees.
Multiplication Rule

Individual events that are "the same" when sampled are called *exchangable* events - those that would not (Later we'll talk about "IID" events, and why that term is less precise.) the "principle of indifference" (or "principle of insufficient reason" - attributed to Laplace. This is not a mathematical, but a physical argument.) assigns equal probabilities over items that are probabilistically indistinguishable.

Principle of indifference

The principle of indifference (also called principle of insufficient reason) is a rule for assigning epistemic probabilities. The principle of indifference states that in the absence of any relevant evidence, agents should distribute their credence (or 'degrees of belief') equally among all the

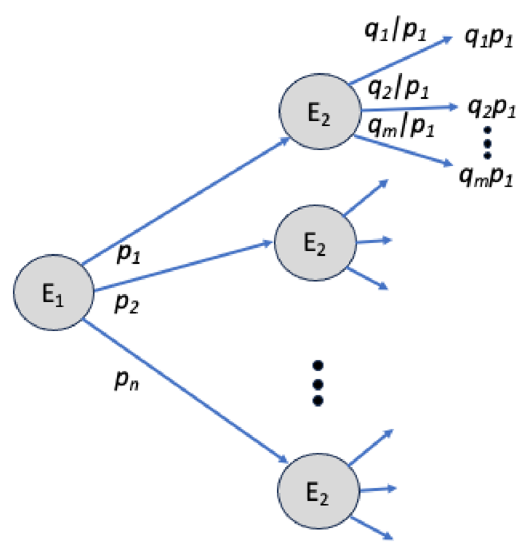
W https://en.wikipedia.org/wiki/Principle_of_indifference

Hence

$$P(one\_event|n\_indistinguishable\_events) = \frac{1}{n}$$

Repeated samples. Things get interesting when considering different ways of combining events E_i into the desired outcome. In short, the combination of two events has a frequency that is the product of the two, if the two events don't affect each other's probability (they are *probabilistically independent.* We will come across "independence" in another sense in vector spaces*).* But most often one event removes an item from the outcome set (think of drawing a card), so *conditioned* on the first event, the second event is drawn from a set of size one less. So the probability of the two events, *without replacement,* starting with n outcomes is $n(n-1)$. This can be generalized to k repetitions to give a count of $n(n-1)\ldots(n-k+1), k < n$ repetitions. (Note that by *exchangeability* these could be any k events, in any order.) When considering repeated events from a set of finite outcomes this is called *sampling without replacement,* in contrast to assuming an "infinite" pool of items that is equivalent to *sampling with replacement.*

A *probability tree* is a general way to visualize combinations of events. Each branching expresses the possibilities of one event. Subsequent branchings show conditioning of one event on another event's outcomes. (Events don't necessarily have to occur one after the other, but "upstream" events condition the probability of downstream ones.).



Computing the size of large samples

When the number of distinct objects gets large, it is awkward to enumerate all possible samples. We need computational ways to count the samples.

Sampling subsets of sets	Rules for distinct (e.g. labeled) objects, with and without replacement. The hypotheticals: How many possibilities are there if…? *Permutation operators - reorderings.*
Permutations	*Generating permutation orderings - by swaps.* A mathematical structure with just one operator (e.g "multiplication") known as a *Group.* Multiplication of group operations are <u>not</u> commutative.

How would you write a permutation function in python? Hint: an array indexed by another array.

Formally: $\pi : (1 \dots k) \rightarrow (1 \dots k)$ is a one-to-one, function over a fixed size domain - as such is both *injective, surjective.* This is called the "pigeon-hole principle"

So, for example a permutation that swapped the first two elements out of four is $\pi(1, 2, 3, 4) \rightarrow (2, 1, 3, 4)$.

> Theorem: Any permutation can be decomposed as a series of swaps, and the *parity* (if the number is odd or even) is the same for any decomposition.

Incidentally - *Why is the order of data interesting, or not?* We use permutations of data for tests - what properties (e.g. statistics) are invariant to permutations and which are not (e.g. association.)

Counting possible permutations of a list.

Factorials $n!$ and "trimmed" factorials.

Recursive def: $n! = (n-1)!n$

$$_nP_k = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

An example: Gertrude has enough paint to paint 3 rooms of her 5 room house each a different color. How many possibilities can she consider?

Gamma function: generalizes to continuous values: $n! = \Gamma(n+1)$. It interpolates values for positive real numbers, is undefined for non-positive integers, and has minimums at fractional "half integer" (-½, -3/2, ..) values.

Combinations

Removing Overcounting. Binomial Coefficients.
For certain combinations we want to throw away the labels that distinguish permutation set with the same items but in different orders. Consider the case above of Gertrude, but where she has only one color of paint. Loosely these are called "combinations." To remove this over-counting of permutations that are combinatorily the same, we divide the count of permutations by the number of permutations of a set of that length. This count is given by the *binomial coefficient:*

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

where $\binom{n}{0} = 1 \text{ and } \binom{n}{n} = 1$

Using binomial coefficients for exchangable subsets Example of "throwing away irrelevant labels" - (from an infinite deck) pick n cards - what is the probability of k red cards?

Binomial Theorem

$$(x+y)^n = \sum_k \binom{n}{k} x^k y^{n-k}$$

in the special case where y=*1,*

$$(x+1)^n = \sum_k \binom{n}{k} x^k$$

which can be proved using a Taylor expansion. With also *x=1* we obtain

$$2^n = \sum_k \binom{n}{k}$$

Finally, we obtain a *Binomial Distribution* if *x+y = 1:*

$$1 = \sum_k \binom{n}{k} x^k (1-x)^{n-k}$$

Assignment

1. Read Evans, *Probability,* Section 1.4
2. All problems in "Probability HW1"

References

- Youtube - First probability lecture

https://www.youtube.com/watch?v=53nZY5udKHQ

- D. Knuth "Fundamental Algorithms, Vol 1", (Addison Wesley, 1975)
  Chapter 1.2.5 Permutations and Factorials, & 1.2.6 Binomial Coefficients