Probability vs Statistics

Probability is where you know this population and its makeup, and you want to understand things about a sample. Statistics is where you have access to a sample and you try to characterize the population from this sample.

Coin Toss Example

Omega = {H, T}
F = { {}, {H}, {T}, {H,T}}
P: *F* → [0, 1] — maps *F* to a real number from 0 to 1
(Omega, *F*, P) — probability space
where

- P(Omega) = 1 — 1st axiom
- P(A) ≥ 0, A in *F* — 2nd axiom, any event has to have a probability of at least 0
- P(U from i=1 to n) A_i = Sum from i=1 to n of P(A_i) if A_i intersect A_j = {} — 3rd axiom, probability of multiple things is the sum of their probabilities assuming their outcomes don't overlap

P(A^C) = 1 - P(A) — derived from above axioms
P(A) ≤ 1 — P(A^C) ≥ 0
P({}) = 0
P(A U B) = P(A) + P(B) - P(A intersect B)
P(A | B) = P(A intersect B) / P(B) = P(B|A) P(A) / P(B) — Baye's Rule
A and B are independent iff P(A interesect B) = P(A)

Based on coin toss distribution, what is the probability of a certain event or outcome?
What is a distribution? To characterize a population, we would like to know all the possible outcomes and their respective probabilities. This list is a distribution.

We need to switch from heads and tails to a numerical thing that we can represent better. We will use a random variable.

Random Variables

A random variable X is a function that takes in events in Omega and maps them to the real numbers. X: Omega → R

X(w) = {1 if w = {H}
{ 0 if w = {T}

Omega could be a group of people and X could be a function that returns their age or height etc.

$P_X(1) = P(X=1) = P(\{w \mid X(w) = 1\}) = P(\{H\})$

What is the distribution of our random variable X?

We said it's the list of all possible outcomes, which is 1 and 0 in this case.

We drew a graph for the distribution of our coin toss with horizontal axis as X(w) and vertical axis as probability. There is a point at (0, 1-p) and a point at (1, p).

Bernoulli Trial

Let's say we flip the coin multiple times such that we have a random variable X_1 for the first flip, a random variable X_2 for the second flip, and so on.

Geometric Distribution

What is the probability that it takes us k number of tosses to get the first heads?

Probability that it takes k tosses to get the first H:

$P_{X1}(0)P_{X2}(0)P_{X3}(0)...P_{Xk}(1)$

We are multiplying (1-p) k times, so the probability is $(1-p)^k$

This is called the Geometric distribution.

We are assuming independence between X_1, X_2, X_3, …

Binomial Distribution

What is the probability that we get k number of heads in the first n coin flips? This is called the Binomial Distribution. There are n choose k ways to put k objects into n bins. We want k successes, so we multiply by p^k, and n-k failures, so we multiply by $(1-p)^k$. We want n choose k ways of doing that.
$P_n(k)$ = (n choose k) p^k $(1-p)^k$

<u>Poisson Distribution</u>
Flip the coin over a large period of time, on average we get 60 heads per hour. What is the probability that we get k heads per minute? This is called the Poisson Distribution, which says that the probability of seeing k heads in a given minute is lambda $e^{-lambda}$ / k!
lambda is our rate, so it equals 1. The probability decreases at a factorial rate with k and is maximum when k = 1. The more we ask for, the lower the probability will be, and in an exponential way.

So far, we are still in the realm of discrete probability distributions. We can still list out all the possible outcomes and their possibilities.

<u>Nondiscrete Distribution</u>
What is the probability that the time between two consecutive heads is two minutes? There is an infinite number of values that can get a very close probability. We can't reasonably call this a probability if we assign it a value because then we would be violating the initial axioms. There is no way we can sum all these values. The question doesn't make sense from a probability standpoint. We should not ask for the probability that it is exactly two minutes. Instead, we should ask for the probability that it is at most two minutes or within a certain interval. This is because time is a continuum of values, so we cannot look at it discretely.

P(X < k) — cumulative distribution

P(time between consecutive Hs) = $e^{-lambda * t}$
where lambda is the rate (1 per minute) and t is the time

This is called the exponential distribution.

Probability Density Function

This is the instantaneous change in the probability at a given point. We can find an infinitesimal small range, find the probability for that range, P(A → B), what is the instantaneous probability at this time t? This is like taking derivatives.

F(t) = lambda $e^{\text{-lambda t}}$ — probability density function

To get the cumulative distribution function, we would integrate this function.

P(t in [t_1, t_2]) = integral from t_1 to t_2 of lambd $e^{\text{-lambda t}}$ dt = P(t ≥ t_1, t≤ t_2)

Normal Distribution

If we were to sample all these times between successive heads and take the average time between successes, what would be the distribution of this average after n samples of intervals? It would be a normal distribution. This is a phenomenon called the Central Limit Theorem.