

# Toss a Coin!

## What is Probability?

Classically, probability is defined as a relative frequency, wherein the probability of an event  $E$  is the (limiting) proportion of time  $E$  occurs.

$$P(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n}$$

One major drawback of this description of probability is that, the convergence of  $\frac{n(E)}{n}$  to a constant limiting value, is taken as an axiom of the system. However, it is not at all obvious why this would be the case. It would be better if our set of axioms were more fundamental, from which we could establish this fact.

## Axiomatic Framework

An alternate description entails assuming that for any event  $E$  of sample space  $S$  there exists a number  $P(E)$ , with the following properties:

1.  $0 \leq P(E) \leq 1$
2.  $P(S) = 1$
3. For any sequence of mutually exclusive events  $E_1, E_2, \dots$ ,

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

$P(E)$  is referred to as the probability of the event  $E$ . It tells us the degree of certainty associated with the event  $E$  occurring.

Using set theory results and the above axioms, we can prove many properties of  $P(E)$ , such as  $P(A^c) = 1 - P(A)$ .

Along with the above formulation, if we also assume that all outcomes of an experiment are equally likely, then the probability of any event  $E$  equals the proportion of outcomes in the sample space that are a part of  $E$ . This assumption allows us to compute values of  $P(E)$  in certain scenarios.

Eg 1. Probability of getting heads on a coin toss =  $P(H) = \frac{1}{2} = 0.5$ . This means that on tossing a coin we can say that the chance of it landing on heads is 50%.

Eg 2. Consider a biased coin, whose bias is unknown. Under this axiomatic description,  $P(H)$  is the number that gives the chance that the coin will land on heads upon tossing. Due to lack of information about the coin, we do not know this number.

Under this axiomatic probability framework, we can arrive at the **Law of Large Numbers**, which states: As the number of trials  $n \rightarrow \infty$ , the relative frequency of occurrence of  $E$  converges to  $P(E)$ .

In Eg 1. this means, with many coin tosses, the relative frequency of landing on heads converges to 0.5.

In Eg 2, the LLN implies that we can estimate  $P(H)$  by performing many trials, and finding the relative frequency of landing on heads. The estimate will converge to the actual value for larger sample sizes.

What this means, is that the axiomatic description of probability is a generalized version of the classical probability.

## Measure of Belief

Another interpretation of probability, is the measure of degree of *one's belief* of the occurrence of an event. This is the subjective interpretation of probability, which is useful in situations where there is uncertainty, but a frequency-based definition is not helpful. These probability values are also consistent with the axioms of probability.

## Random Variable

In a random experiment, the sample space lists all the possible outcomes of the experiment. But often, we are interested in some property of the outcome, not the outcome itself. For example, if the random experiment is two dice rolls, the sample space is

$$\Omega = \{(1, 1), (1, 2), (1, 3), \dots, (6, 6)\}$$

but we might be only interested in the sum of the outcomes of the two throws, i.e. value of the function

$$S(\omega) : \Omega \rightarrow \mathbb{R} = \omega_1 + \omega_2$$

We denote the event that  $S$  attains the value  $k$

$$\{S = k\} = \{\omega : S(\omega) = k\}$$

Then we denote the probability of  $S$  is  $k$

$$P(S = k) = P(\{\omega : S(\omega) = k\})$$

## Select 2 goldcoins?

Consider 3 boxes with 2 coins each: Box 1 has 2 silver coins, Box 2 has 1 gold and one silver coins, Box 3 has 2 Gold coins. If a randomly drawn coin from a random box is a gold coin, what is the probability of the other coin in that box also being gold?

Let  $G = 1$ , and  $S = 0$  for a moment. Let  $X$  = random variable denoting the other coin in the box.

Then,

$$\begin{aligned} E[X] &= G \times P(X = G) + S \times P(X = S) \\ &= G \times (?) + S \times (?) \end{aligned}$$

Consider,

$$X \sim \text{Binomial}(n, p)$$

Then,

$$\begin{aligned}
 p_M(x) &= \binom{n}{x} p^x (1-p)^{n-x} \\
 &= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}
 \end{aligned}$$

Put  $p \rightarrow 0$ ,  $n \rightarrow \infty$  such that  $np = \lambda$  stays constant.

$$\begin{aligned}
 p_M(x) &= \lim_{n \rightarrow \infty} \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
 &= \lim_{n \rightarrow \infty} \frac{\cancel{n(n-1)\dots(n-x+1)}}{x! \cancel{n^n (n-\lambda)^x}} \lambda^x (n-\lambda)^n \\
 &= \frac{\lambda^x}{x!} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = \frac{\lambda^x}{x!} e^{-\lambda}
 \end{aligned}$$

This gives us the Poisson's distribution.

What does this process actually mean? We took a binomial and made its success chances very low, and number of trials very high.  $\lambda$  is actually the expected value, i.e. number of successes on average for the experiment.

Note that the random variable still denotes the same quantity, i.e. the number of successful trials.

Consider the following random variables:

$X$ : Number of heads in  $n$  trials of coin tosses, each having a probability  $p$  of landing heads.

$Y$ : Number of white balls drawn (without replacement) in  $n$  trials from a bag having  $m$  white balls out of  $N$  total. ( $p = \frac{m}{N}$ )

Notice,  $X \sim \text{Binomial}(n, p)$ .

Let  $X_i$  and  $Y_i$  indicate success in the  $i$ th trial, i.e.

$$X_i = \begin{cases} 1 & \text{success} \\ 0 & \text{otherwise} \end{cases}$$

Then,  $X_i$  and  $Y_i$  are binomial distributions.

Note that,

$$\begin{aligned}
 E[X] &= \sum_i^n E[X_i] && \text{(why?)} \\
 &= np
 \end{aligned}$$

and similarly,

$$E[Y] = \sum_i^n E[Y_i] = np$$

## Expected Value - Definitions

What do the two definitions of expected value mean? How are they different?

$$E[X] = \sum_x x p_M(x)$$

“The expected value of  $X$  is the weighted average of each value of  $x$  weighted by the probability of  $X$  taking that value.”

$$E[X] = \sum_{s \in S} X(s) P(s)$$

“The expected value of  $X$  is the weighted average on each outcome, the value of  $X$  for that outcome, weighted by the probability of that outcome.”

Aside: Given that  $X = \sum_i^n X_i$ , show that  $E[X] = \sum_i^n E[X_i]$ .

From definition of  $E[X]$ ,

$$E[X] = \sum_x x p_M(x)$$

Now, note that  $X$  taking values  $x$  means  $\sum_i^n X_i$  taking values  $x$ . If  $X_i$  takes values  $x_i$ , then  $x = \sum_i^n x_i$ . (We are not asserting here anything, we are establishing a relation between the supports. Note that the  $x$ s are support variables, they do not hold some value.)

Also note,  $p_M(x) = P(X = x) = P(\sum_i^n X_i = x)$ .

Note that events  $\{X = x\}$  for different  $x$  are mutually exclusive (and exhaustive).

So,

$$E[X] = \sum_x \sum_i^n x_i p_M(x)$$

Before going there, let's just show if  $X = X_1 + X_2$  then  $E[X] = E[X_1] + E[X_2]$ . The general case follows directly.

$$\begin{aligned}
 E[X] &= \sum_x x p_M(x) \\
 &= \sum_x x \sum_{x_1, x_2} P(X_1 = x_1, X_2 = x_2) \\
 &= \sum_{x_1, x_2(x)} \sum_x (x_1 + x_2) P(x_1, x_2)
 \end{aligned}$$

Note that summing over all  $x_1 + x_2 = x$  and summing over all  $x$  is equivalent to summing over every pair  $x_1, x_2$ !

Then,

$$\begin{aligned}
 &= \sum_{x_1, x_2} x_1 P(x_1) + \sum_{x_1, x_2} x_2 P(x_1, x_2) \\
 &= E[X_1] + E[X_2]
 \end{aligned}$$

Continuing on, we noticed that expectation values of both  $X$  and  $Y$  are same.

Also note, probability of success in first try in both cases is  $p$ .

But, probability of first success and second success is:

$$p^2$$

and

$$\frac{m}{N} \times \frac{m-1}{N-1}$$

The second distribution ( $Y$ ) is  $Y \sim \text{Hypergeometric}(N, m, n)$ .

- For  $m = 1$ , hypergeometric distribution behaves as Bernoulli( $\frac{n}{N}$ ).
- For  $m \rightarrow \infty, N \rightarrow \infty, \frac{m}{N} \rightarrow p$ , hypergeometric distribution behaves as Binomial( $n, p$ )

## Theoretical Examples

1. Each time a coupon is obtained, it is of type  $i$  with probability  $P_i$ , for  $i = 1, \dots, N$ .  
 $T$  = number of coupons to be selected after which one has at least one coupon of each type.

Note that  $P(T = n) = 0$  for  $n < N$ .

Let:  $A_j = j$ th coupon was not obtained after  $n$  tries.

Then,

$$P(T > n) = P\left(\bigcup_j A_j\right)$$

Use exclusion-inclusion.

$$P(T > n) = \sum_j P(A_j) - \sum_{j_1 < j_2} P(A_{j_1} \cap A_{j_2}) + \dots + (-1)^{N+1} P(A_1 A_2 \dots A_N)$$

Now,

$$\begin{aligned} P(A_j) &= P(\text{In } n \text{ tries, could not find coupon } j) \\ &= (1 - P_j)^n \end{aligned}$$

Similarly,

$$\begin{aligned} P(A_{j_1} \cap A_{j_2}) &= P(\text{In } n \text{ tries, could not find coupon } j_1 \text{ and } j_2) \\ &= (1 - P_{j_1} - P_{j_2})^n \end{aligned}$$

( $\because$  getting coupon  $j_1$  and getting coupon  $j_2$  for a given try are mutually exclusive)

And,

$$P(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_k}) = \left(1 - \sum_{p=1}^k P_{j_p}\right)^n$$

So we can write: (for  $n > 0$ )

$$\begin{aligned} P(T > n) &= \sum_j (1 - P_j)^n - \sum_{j_1 < j_2} (1 - P_{j_1} - P_{j_2})^n + \dots + 0 \\ &= (1 - P_1)^n + (1 - P_2)^n + \dots + (1 - P_N)^n \\ &\quad - (1 - P_1 - P_2)^n - (1 - P_1 - P_3)^n - \dots - (1 - P_{N-1} - P_N)^n \\ &\quad + (1 - P_1 - P_2 - P_3)^n + \dots + (1 - P_{N-2} - P_{N-1} - P_N)^n \\ &\quad \dots \\ &\quad + 0 \end{aligned}$$

Similarly  $P(T > n - 1)$  will have all the same terms, with the power  $n - 1$ .

$$P(T = n) = P(T > n - 1) - P(T > n)$$