Probability Measures

Intuitively:

Probability measure assigns a number between 0 and 1 that measure the certainty or likelihood that an event occurs when a random experiment is done.

* For instance, if we flip a coin twice and ask that, what is the probability that at least one head (H) occurs ?

Mathematically: Probability measure is a set function that map an event to any number between 0 and 1. In short,

$$P : \underbrace{\mathcal{F}(S)}_{\text{event space}} \rightarrow \mathbb{R}[0, 1]$$

satisfying the Axioms of Probability. The above set function can also be termed as the probability law.

Precisely, A probability law is a function $P : \mathcal{F} \rightarrow [0, 1]$ that maps an event to a real number in between 0 and 1.

Axioms of Probability:

1. $P(A) \geq 0, \ \forall A \in \mathcal{F}$
2. $P(S) = 1$
3. If $A_1, A_2 \ldots \ldots A_n \in \mathcal{F}$, are disjoint, then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

4. If $A_1, A_2, \dots A_n \dots \dots \in \mathcal{F}$ and they are disjoint then,

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

Intuitively,

Axiom 1 is about the non-negativity of probability. That is, probability of any event cannot be negative

Axiom 2 is the normalization axiom. It assures that probability of observing all possible outcome is 1.

Axiom 3 is the additivity axiom. It is the most crucial one – the additivity axiom suggests that, if we have disjoint events, the probabilities can be added.

example:

Consider rolling a die: $S = \{1, 2, 3, 4, 5, 6\}$

Define an event: $A = \{1, 6\}$

So, probability that the event A occurs is:

$$P(\{1, 6\}) = P(\{1\}) \cup P(\{6\}) = P(\{1\}) + P(\{6\})$$

$$= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

As the events $\{1\}$ and $\{6\}$ are disjoint, based on the axiom 3 we can add the probability.

Corollaries from axioms:

Let's assume that A be an event in F

That is: A , Then

a) $P[A] = 1 - P[A]$ b) $P[A] \leq 1$ c) $P[\emptyset] = 0$

Prove that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

We can write $A \cup B$ as a partion of three disjoint subsets

$$A \cup B = (A-B) \cup (A \cap B) \cup (B-A)$$

$$\Rightarrow P(A \cup B) = P(A-B) + P(A \cap B) + P(B-A)$$

$$\Rightarrow P(A \cup B) = P(A-B) + P(A \cap B) + P(B-A) P(A \cap B) - P(A \cap B)$$

$$\Rightarrow P(A \cup B) = P(A \cap \bar{B}) + P(A \cap B) + P(B \cap \bar{A}) + P(A \cap B) - P(A \cap B)$$

$$\Rightarrow P(A \cup B) = P\{(A \cap \bar{B}) \cup (A \cap B)\} + P\{(B \cap \bar{A}) \cup (B \cap A)\} - P(A \cap B)$$

$$\Rightarrow P(A \cup B) = P(A \cap (\bar{B} \cup B)) + P(B \cap (\bar{A} \cup A)) - P(A \cap B)$$

$$\Rightarrow P(A \cup B) = P(A \cap S) + P(B \cap S) - P(A \cap B)$$

$$= P(A) + P(B) - P(A \cap B)$$

Proved

Definitions:

Assume that $A_1, A_2, A_3 \dots A_n \dots$ is a sequence of sets.

The sequence is increasing if

$$A_1 \subset A_2 \subset A_3 \dots \subset A_n \subset \dots \dots$$

Here, every set is the subset of the mediate next set.

The sequence is decreasing if

$$A_1 \supset A_2 \supset \dots \supset A_n \supset \dots$$

every subsequent set is the subset of the previous set.

Applying Union and Intersection

If $A_1, A_2, \dots A_n$ is an increasing sequence of sets, then

$$\lim_{n \to \infty} A_n = \bigcup_{i=1}^{\infty} A_i$$

Because, A_n contains all the previous elements

follows that

$$A_n = \bigcup_{i=1}^{n} A_i$$

If $A_1, A_2, \dots \dots A_n \dots$ is a decreasing sequence of sets, then

$$\lim_{n \to \infty} A_n = \bigcap_{i=1}^{n} A_i$$

As, A_n is the smallest set and other sets contain A_n

Important Fact: If a given sequence $A_1, A_2, \dots, A_n \dots$ is either increasing or decreasing sequence of sets then

$$P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$$

Known as, sequential continuity of the probability measure.

Here, $\left(\lim_{n\to\infty} A_n\right)$ is a set and $\left(\lim_{n\to\infty} P(A_n)\right)$ is a real number, and the real number is easier to find.

Example of Probability Spaces

Example 1: Let's consider that S is a sample space and $\mathcal{F}(S)$ is the event space of S. Consider, power set of S

Let's assume that we have a function $P(w)$ such that

$$P(\omega) : S \to \mathbb{R}$$ such that

Two condition to be satisfied

(i) $P(\omega) \geq 0$ (ii) $\sum_{\omega \in S} P(\omega) = 1$

This type of function is called probability mass function (pmf)

makes it easy to assign probabilities.

Assigning probability for all events as n goes large is doable but hectic task. Instead, we have pmf

We find "n" numbers to map each element of sample space S

Here, $P(\omega)$ is not a probability measure as it is not defined on the event space.

Instead, It is a set function that maps element of the sample space to the real line in such a way that —

- Values to map are positive, and
- if the values are summed together they are equal to 1

So, this set function, defined as the pmf, can be used to define a probability measure

$$P : \mathcal{F}(S) \rightarrow \mathbb{R} \text{ as,}$$

$$P(A) = \sum_{\omega \in A} P(\omega), \quad \forall A \in \mathcal{F} \qquad \ldots \ldots (1)$$

The set function satisfies $P(\omega) \geq 0$ and $\sum_{\omega \in S} P(\omega) = 1$, and hence, the equation (1) satisfies the axioms of probability. So, the definition in (1) is a valid probability measure.

Note: $P(\omega) = P(\underbrace{\{\omega\}}_{\text{singleton set}}), \quad \forall \omega \in S$

So, any event can be represented using countable and finite union of singleton set of members that form the specific event of interest.

Example: Uniform pmf

Let's consider "n" possible outcomes from a random experiment. The sample space S is

$$S = \{\omega_1, \omega_2, \omega_3 \dots \dots \omega_n\} \quad \text{Finite}$$

$$\mathcal{F}(S) = \text{event space} = \mathcal{P}(S) = \text{Power set of } S$$

So, $|\mathcal{F}(S)| = \text{cardinality of } \mathcal{F}(S) = 2^n$

As the sample space has finite number of element, we can take power set of S directly as the event space.

So, probability of any event A_K can be defined as:

Any of the possible 2^n events

$$P(A_K) = \sum_{\omega \in A_K} P(\omega)$$

Let's define

$$P(\omega) = \frac{1}{n} \equiv \text{uniform}$$

$$= \sum_{\omega \in A_K} \frac{1}{n}$$

$\forall A_K \in \mathcal{F}(S)$

$$= |A_K| \cdot \frac{1}{n} = \frac{|A_K|}{|S|} \equiv \text{Classical Notion of Probability Measure}$$

In classical probability

- Structure appears easy, but the counting is difficult.
- Finite number of outcomes
- All outcomes are equally likely

Consider, we are choosing 5 cards out of a deck of 52

3 one type, 2 another type

Counting the favorable 5 card hand would be difficult.

pmf example 3 : Binomial pmf

Let's consider the case of finite sample space which is made up of $n+1$ element.

$$S = \{0, 1, 2, \dots \dots n\}$$

Finite set, so, power set is okay.

$$\mathcal{F}(S) = \mathcal{P}(S) = \text{Power set of } S$$

So, cardinality $|\mathcal{F}(S)| = 2^{n+1}$, $|S| = n+1$

We define the pmf

$$P(K) = \binom{n}{K} a^K (1-a)^{n-K}, \quad a \in [0, 1]$$

$$K = 0, 1, 2 \dots n$$

where, $\binom{n}{K} = \frac{n!}{K!\,(n-K)!}$ $\quad 0! = 1$

So, $P(A) = \sum_{\omega \in A} P(\omega) = \sum_{K \in A} \binom{n}{K} a^K (1-a)^{n-K}$

where, $A \in \mathcal{F}(S)$

Now, Is $P(K)$ a valid form of pmf ?

$$P(K) \geq 0$$

$$\sum_{K=0}^{n} \binom{n}{K} a^K (1-a)^{n-K} = 1$$

$P(K) = \binom{n}{K} a^K (1-a)^{n-K}$

$\left.\begin{array}{l} \binom{n}{K} > 0, \quad a^K \geq 0 \text{ as } a \in [0, 1] \\ (1-a)^{n-K} \geq 0 \end{array}\right\}$ So, $P(K) \geq 0 \quad \forall K$

For the second condition, Let's consider the Binomial theorem

$$(a+b)^n = \binom{n}{0} a^n b^0 + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 \dots$$

$$\dots \dots \binom{n}{n} a^0 b^n$$

Assume $a = a$, $b = 1-a$. Then,

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^n b^{n-k} = (a+(1-a))^k$$

$$= 1^k = 1$$

$$\Rightarrow \sum_{k=0}^{n} \binom{n}{k} a^n (1-a)^{n-k} = 1 \quad \text{Proved}$$

pmf example 4: Think about probability measure

probability distribution of the number of trials needed to get one success. For instance, we can ask, what is the of trials necessary until we hit 6 in the random experiment.

We can consider geometric pmf for this case:

Let's consider S = sample space

$= \{0, 1, 2, 3, \dots \dots\}$ Countable

$$\mathcal{F}(S) = \mathcal{P}(S)$$

pmf: $P(k) = (1-a)\,a^k$, where $a \in (0,1)$

So, $P(A) = \sum_{k \in A} P(k) = \sum_{k \in A} (1-a)\,a^k$, $A \in \mathcal{F}(S)$

Is this valid pmf ?

1. $P(k) \geq 0$ as $k = \;,2,3\dots\dots$

2. We know that $\sum_{k=0}^{\infty} a^k = \frac{1}{1-a}$, $|a|<1$

So, $$\sum_{k=0}^{\infty} P(k) = \sum_{k=0}^{\infty} (1-a)\,a^k$$ As we know that $$\sum_{k=0}^{\infty} a^k = \frac{1}{1-a}$$

$$= \sum_{k=0}^{\infty} (1-a)\,.\,a^k$$

$$= (1-a)\sum_{k=0}^{\infty} a^k = (1-a)\left(\frac{1}{1-a}\right) = 1$$ Proved

Example 5: Poisson pmf

The sample space $S = \{0, 1, 2 \ldots \ldots\}$ countable

Event space $= \mathcal{F}(S) = \mathcal{P}(S)$

pmf: $$P(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, 3 \ldots \ldots \ \& \ \lambda > 0$$

So, $$P(A) = \sum_{k \in A} P(k), \quad \forall A \in \mathcal{F}$$

Is it a valid pmf? We need to check both the condition.

1. As $\lambda > 0$, $e^{-\lambda} > 0$, $k \geq 0$, we can say that $P(k) \geq 0$ for all $k \in S$.

2. We know that $$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

So, $$\sum_{k=0}^{\infty} P(k), \forall k = P(S) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda}$$

$$= e^0 = 1$$

Proved

All the pmf we have seen so far are for sample spaces that are countable. For an uncountable sample space, for instance $S = \mathbb{R}$, we can not take the $\mathcal{F}(S)$ as the power set of $\mathbb{R}$. Instead, we consider Borel Field.

So, $S = \mathbb{R}$

$\mathcal{F} = B(\mathbb{R})$

$P(\cdot)$: Probability measure for every event in the Borel Field.

But, How to define the probability measure?

For finite or uncountable sample space, we had pmfs defined that satisfy the axioms of probability and act as the probability measure for discrete case.

But for uncountable sample space pmf definition is not sufficient and hence, we need a new function for uncountable sample space. Precisely,

we introduce probability density function (pdf) to assign probabilities $P(A)$ where, $A \in B(\mathbb{R})$

Properties of the pdf:

1. $f(x) \geq 0, \quad \forall x \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} f(x) = 1$

Given a valid pdf, we can get a valid probability measure $P(\cdot)$ of any event A in $\mathcal{B}(\mathbb{R})$ as follows:

$$P(A) = \int_A f(r)\,dr = \int_{-\infty}^{\infty} f(r) \cdot 1_A(r)\,dr$$

where, $1_A(r) \triangleq \begin{cases} 1, & r \in A \\ 0, & r \notin A \end{cases}$ known as indicator function of set A.

So, given a valid pdf $f(r)$ and any set $A \in \mathcal{B}(\mathbb{R})$, we have $P(A) = \int_A f(r)\,dr = \int_{-\infty}^{\infty} f(r) \cdot 1_A(r)\,dr$

However, question can be asked as

Does $P(A) = \int_{-\infty}^{\infty} f(r) \cdot 1_A(r)\,dr$ give a valid probability for every $A \in \mathcal{B}(\mathbb{R})$? Answer is both YES & NO

⊞ Consider the Rieman integral

$$\int_{\mathbb{R}} f(r) \cdot 1_A(r)\,dr$$

- well defined for any A that is an interval such as $A = (a, b)$ or $[a, b]$ where $a < b$

But for some Borel set Rieman integral will be undefined. In those cases, we can use another form of integral.

- It is also well defined for A equal to a finite union of intervals.