1.3. Probability

1. Let us begin with an informal definition of the probability function.

2. *Definition.* Probability is the measure of the likeliness that an event will occur.

   In the discrete case, the probability function is the function that assigns a probability to each sample point and each event.

3. *Definition.* If X is a discrete random variable, then the probability mass function of X is a function that assigns a probability to each value of the random variable X.

4. Now we look an example.

5. *Example.* Let the experiment be the toss of two coins in a row. Let X be the number of coins that turn out heads.

   The sample space for this experiment is

$$S = \{hh, ht, th, tt\} .$$

   The probability function P is a function that assigns a probability to each sample point and to each event. In this example, the four sample points are equally likely, so we have

$$P(hh) = P(ht) = P(th) = P(tt) = \frac{1}{4}.$$

   This way the probability function assigns a probability of $1/4$ to each sample point.

Next, the probability function assigns a probability to each event via the following mechanism:

$$\text{For each } A \subset S, P(A) = \sum_{\omega \in A} P(\omega).$$

For example, if $B = \{ht, th\}$, then $B \subset S$ and

$$P(B) = P(\{ht, th\}) = P(ht) + P(th) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

In particular, $P(\emptyset) = 0$, and $P(S) = 1$.

Recall that X is the number of coins that turn out heads. This random variable X has three possible values $0, 1, 2$. For each $i = 0, 1, 2$, we define

$$f(i) = P(X = i).$$

It is clear that, by this definition of f,

$$f(0) = P(X = 0) = P(\{tt\}) = \frac{1}{4},$$

$$f(1) = P(X = 1) = P(\{ht, th\}) = \frac{1}{2},$$

$$f(2) = P(X = 2) = P(\{hh\}) = \frac{1}{4}.$$

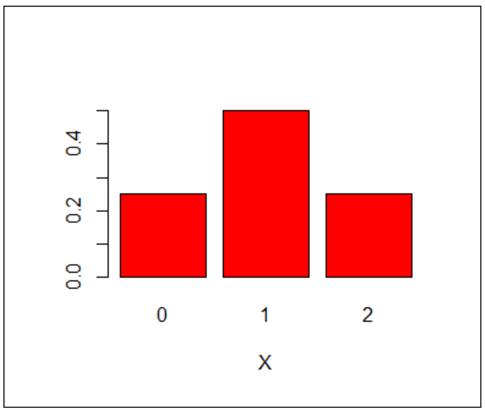
In summary,

$$f(0) = \frac{1}{4}, \quad f(1) = \frac{1}{2}, \quad f(2) = \frac{1}{4}.$$

We can use a table to represent the function f:

x	0	1	2
$f(x)$	$1/4$	$1/2$	$1/4$

This function f is the pmf (probability mass function) of the random variable X, because it assigns a probability to each possible value of X. We can also draw the bar graph of f:



6. Below is the formal definition of probability:

7. *Definition.* (Probability)

   Probability is a real valued set function P that assigns, to each event A in the sample space S, a number $P(A)$ such that

   (a) (non-negativity) $P(A) \geq 0$,

   (b) (unitarity) $P(S) = 1$,

   (c) (additivity) If $A_1, A_2, \cdots, A_k$ are mutually exclusive events, then

   $$P(A_1 \cup \cdots \cup A_k) = P(A_1) + \cdots + P(A_k).$$

8. It is clear that the domain of the probability function P is the collection of all events — the collection of all subsets of the sample space S.

9. The three properties in this formal definition of probability are called probability axioms — the nonnegativity axiom, the unitarity axiom, and the additivity axiom.

10. Next, we derive some more properties of the probability function from the axioms.

11. *Theorem.* For each event A, $P(A) = 1 - P(A')$.

    — Proof. Note that A and A' are disjoint. By additivity,

$$P(A) + P(A') = P(A \cup A') = P(S) = 1.$$

    It follows immediately that $P(A) = 1 - P(A')$.

12. *Theorem.* $P(\emptyset) = 0.$

— Proof. By the last theorem, we have

$$P(\emptyset) = 1 - P(\emptyset') = 1 - P(S) = 1 - 1 = 0.$$

13. *Theorem.* (monotonicity) If $A \subset B$, then $P(A) \leq P(B)$.

— Proof. By additivity and non-negativity,

$$P(B) = P(A) + P(B \backslash A) \geq P(A) + 0 = P(A).$$

14. *Theorem.* For each event A, $P(A) \leq 1.$

— Proof. By monotonicity,

$$P(A) \leq P(S) = 1.$$

15. *Theorem.* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

16. *Theorem.* (optional at this time) We have

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(B \cap C) - P(A \cap C) \\ &\quad + P(A \cap B \cap C). \end{aligned}$$

17. The last two theorems are special cases of the inclusion-exclusion principle.

18. Next, we introduce a method for finding probabilities.

19. *Classical Probability Model.* In a classical probability model, the sample space of the experiment is a finite set, and all sample points are equally likely.

    Suppose that

$$S = \{e_1, e_2, \cdots, e_m\}$$

and the m outcomes are equally likely. If $A \subset S$ and A contains k sample points, then

$$P(A) = \frac{k}{m}.$$

20. *Example.*

Let the experiment be the toss of three coins. The sample space is

$$S = \left\{ \begin{array}{cccc} hhh & hht & hth & thh \\ tth & tht & htt & ttt \end{array} \right\}.$$

Let X be the number of coins that turn up heads. Then

$$(X = 2) = \{hht, hth, thh\},$$

and

$$P(X = 2) = P(\{hht, hth, thh\}) = \frac{3}{8}.$$

21. *Example.*

Let the experiment be the toss of two dice, and let S be the sample space of the experiment.

Let X be the sum of the outcomes of the two dice, and let A be the event that $X \geq 9$. Find $P(A)$.

— Solution.

First of all, the sample space is

$$\begin{array}{cccccc}
(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\
(2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\
(3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\
(4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\
(5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\
(6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6)
\end{array}$$

It clear that the event A contains the ten sample points in the lower right corner of the sample space, therefore,

$$P(A) = \frac{10}{36}.$$

Here, $10 = |A|$ and $36 = |S|$.