Discrete Random Variables & Discrete Distributions

A random variable is a function from the sample space \( \Omega \) to the set of real numbers \( \mathbb{R} \).

If the range of a random variable is discrete (as a subset of \( \mathbb{R} \)), then it is called a discrete random variable; otherwise it is a continuous random variable.

Remark: If the sample space \( \Omega \) is a finite set, then any random variable \( X : \Omega \rightarrow \mathbb{R} \) must be discrete. If \( \Omega \) is infinite, a random variable on \( \Omega \) may or may not be discrete.

Note: We will gradually extend our study from finite sample spaces to infinite ones.

Example: Toss a fair coin repeatedly. Let \( A \) be the event that the first time head appears. This sample space is infinite, since one can keep tossing the coin. And the event \( A \) also does contain
infinitely many results:

\[ A = \{ \text{the first toss was a head} \cup (H) \]
\[ \text{the first was a tail and the second was a head} \cup (TH) \]
\[ TTH, TTH, \ldots \]

We can still construct a discrete random variable on \( \Omega \).

Define it to be \( Y: \Omega \rightarrow \mathbb{R} \). Which counts the total number of tosses until the first heads appears.

So \( Y(H) = 1 \), \( Y(TH) = 2 \), \( Y(TTH) = 3 \), \ldots

---

**Probability vs. Random Variables**

Suppose \( X: \Omega \rightarrow \mathbb{R} \) is a random variable. Using this \( X \) we can talk about some events and their probabilities given out of this random variable. For example,

for some number \( a \), let \( A \) be the event consisting of all elements in \( \Omega \) that are mapped to \( a \) by \( X \).

i.e. \[ A = \{ e \in \Omega \mid X(e) = a \} \].

\[ \Box \]
The probability \( P(A) \) is usually denoted by

\[ P(X = a) \]

If we choose a different number \( b \), we will get another event \( \{X = b\} \) with a probability \( P(X = b) \).

So by varying this number we can define a function

for any real numbers, call it \( p \).

The \( p \) function is defined

\[ p(x) = P(X = x) \]

called a probability mass function.

E.g. Toss a coin 3 times. So \( \Omega \) consists of 8 outcomes.

Let \( X : \Omega \rightarrow \mathbb{R} \) be the function counting the number of heads in each outcome, then \( X \) takes values among 0, 1, 2, 3. (so it is discrete). Then the \( p \) function will be defined as follows:

\[ p(0) = \text{the probability of no heads} = P(\text{T T T T}) = \frac{1}{8} \]

\[ p(1) = \text{the probability of having exactly 1 head and 2 tails} = \frac{3}{8} \]
And similarly, \( p(2) = \frac{3}{8} \), \( p(3) = \frac{1}{8} \).

What else? if \( x \) is any number other than 0, 1, 2, 3, \( p(x) \) is simply zero.

Notice that \( p(0) + p(1) + p(2) + p(3) = 1 \) and \( p(x) \geq 0 \) for any \( x \).

**Example:** Tossing a coin until the first heads appears.

Let \( X : \mathcal{S} \rightarrow \mathbb{R} \) be the function counting the number of tosses. Now let's calculate the probability mass function \( p \).

First of all, what are the possible values under the random variable \( X \)?

- \( X(H) = 1 \)
- \( X(TH) = 2 \)
- \( X(TTH) = 3 \)

So \( X \) can be any positive integer.
Then \[ p(1) = P(X = 1) = P(\text{the first toss showed a head}) \]
\[ = \frac{1}{2} \]
\[ p(2) = P(X = 2) = P(\text{first head + second head}) \]
\[ = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \]
\[ p(3) = P(X = 3) = P(\{TTH\}) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} \]
\[ \vdots \]
\[ p(1) + p(2) + p(3) + \ldots = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots = 1 \]

Another related function is the (cumulative) distribution function of a random variable \( X \):
\[ F(x) = P(X \leq x) \]

Notice that since \( P \) is non-negative, \( F \) is a non-decreasing function.

In the previous example,
\[ X \leq 1 : \text{the first toss was head} \]
\[ X \leq 2 : H \text{ or TH} \]
\[ X \leq 3 : H, \text{ or TH, or TTH} \]
Useful and important random variables and their distributions

1. The binomial distribution.

Example: Toss a coin 10 times.

Random variable: the number of tails

So the random variable $X$ takes values from 0 to 10.

Calculate its probability mass function:

- If $X=0$, all are heads, $\Rightarrow p(0) = \left(\frac{1}{2}\right)^{10}$
- If $X=10$, all are tails, $\Rightarrow p(10) = \left(\frac{1}{2}\right)^{10}$
- If $X=1$, there is one tail and 9 heads, in which one is the tail? there are 10 positions that the tail can occur, so $p(1) = \binom{10}{1}\left(\frac{1}{2}\right)^{10}$
- If $X=2$, there are 2 tails, where are they? There are $\binom{10}{2}$ possible positions for the 2 tails to occur, $\Rightarrow p(2) = \binom{10}{2}\left(\frac{1}{2}\right)^{10}$
What if the coin is unfair: 0.3 — head
0.7 — tail

In this case

\[ p(0) = P(\text{all heads}) = (0.3)^{10} \]

\[ p(1) = P(1 \text{ tail}) = \binom{10}{1}(0.7)(0.3)^9 \]

Which one?

\[ p(2) = P(2 \text{ tail}) = \binom{10}{2}(0.7)^2(0.3)^8 \]

**Binomial distribution**. All trials are independent. Each trial has 2 possible outcomes: successful and unsuccessful with the probability of success \( p \).

(So failure has probability \( 1 - p \)).

If we experiment the trial \( n \) times. Let \( S_n \) be the random variable that counts the number of successes.

Then we can calculate

\[ P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \]

and \( S_n \) is said to be binomially distributed, with parameters \( n \) and \( p \). \( S_n \) is called a binomial random variable and its distribution is called the binomial distribution.
2. Geometric distribution

Experiment: toss a coin until the heads appear for the first time.

Random variable: the number of tosses to stop the experiment.

We saw that the probability mass function looks like:

\[ p(1) = \frac{1}{2} \]
\[ p(2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \]
\[ p(3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} \]

What if the coin is unfair? head \(\uparrow\) tail \(\downarrow\) \(1-p\).

Then:

\[ p(1) = p \]
\[ p(2) = (1-p) p \]
\[ p(3) = (1-p)^2 p \]
\[ p(4) = (1-p)^3 p \]