Ch 3 Random variables and Distributions

Def 3.1.1 Let \( \Omega \) be the sample space of an experiment. Let \( X \) be a real-valued function defined on \( \Omega \):

\[
X : \Omega \rightarrow \mathbb{R}
\]

\( X \) takes an outcome \( \omega \in \Omega \) and maps it to a unique number in \( \mathbb{R} \).

Then \( X \) is called a random variable (r.v.).

Eg. Flip a coin twice. \( \Omega = \{HH, HT, TH, TT\} \)

\[
X(\{HH\}) = 2 \\
X(\{HT\}) = 1 \\
X(\{TH\}) = 1 \\
X(\{TT\}) = 0
\]

\( X \) is a r.v. \( P(X=2) = 2 \times 1/2 \times 1/2 = \frac{1}{4} \)

In fact, \( X \) is the number of heads.

\[
X = 0, 1, 2 \rightarrow \text{support of the r.v. } X
\]

Eg. 3 donors enter a blood bank.

It is known that 9% population have 0-negative blood.

Assume the three donors are independent.

\( \Omega = \{O+, O-, O+, O-, O+, O-, O+, O-, O+, O-, O+, O-\} \)

Define \( X = \# \) of 0-

\[X(O-, O-, O-) = 3, X(\text{any others}) = 2, X(\text{all are 0+}) = 0\]

The largest value \( X \) can take is 3, the smallest.

\[
P(X=0) = P(\text{mi}, \text{mi}, \text{mi}) = 0.9 \times 0.9 \times 0.9 = 0.9^3 = \binom{3}{0}(0.07)^0(1-0.07)^3
\]

\[
P(X=1) = \binom{3}{1}(0.07)^1(0.93)^2
\]

\[
P(X=2) = \binom{3}{2}(0.07)^2(0.93)^1
\]

\[
P(X=3) = (0.07)^3 = \binom{3}{3}(0.07)^3 = (1-0.07)^0
\]

The set of possible values \( \{0, 1, 2, 3\} \) is called the support of \( X \).

In general, if there are \( n \) donors, let \( p = \text{prob a single donor is } 0^- \)

\( X = \# \) 0- donors

\[
P(X=k) = \binom{n}{k}p^k(1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n.
\]

Always arrange

\( n = \text{total donors} \)

\( k = \text{number of 0- donors} \)
e.g. Tossing a fair coin 10 times. Assume independence.

Let \( X \) denote the number of heads

\[
P(X=k)=\binom{10}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{10-k} = \binom{10}{k} \left(\frac{1}{2}\right)^{10}, \quad k=0,1,2,\ldots,10
\]

These random variables follow Binomial distributions. They will be discussed later.

**Types of random variables**

1. **Discrete**: \( X \) can take on only a finite or countably infinite number of possible values, i.e., the support of \( X \) is finite or countably infinite.
   
   We can list the possible values of \( X \) as \( x_1, x_2, \ldots \)

   e.g. \( X \) = # of blood donors with type AB in a sample of 3 donors
   
   \( X = 0, 1, 2, 3 \)
   
   \( X = \# \) of tosses of a fair coin until a head appears
   
   \( X = 1, 2, \ldots \)
   
   \( X = \# \) of taxi arrivals in an hour
   
   \( X = 0, 1, 2, \ldots \)

2. **Continuous**: The set of all possible values of \( X \) is an interval (or union of intervals) in \( \mathbb{R} \).

   e.g. \( X \) = height of a randomly selected woman
   
   \( X = (0, \infty) \)
   
   \( X \) = the waiting time until the first bus
   
   \( X = (0, \infty) \)

How to characterize the distribution of \( X \)?

1. **Cumulative distribution function (cdf)**
   
   For \( x \in \mathbb{R} \), \( F(x) \equiv P(X \leq x) \)

   the definition works for both discrete and continuous r.v.s

2. If \( X \) is discrete, one can define the probability mass function of \( X \):

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

3. If \( X \) is continuous, one can define the probability density function as the function \( f \) such that for any interval \((a, b)\)

   \[ P(a < X < b) = \int_a^b f(x) \, dx \]