Lecture Notes 4: Probability and Statistics

So far we have talked about random experiment, sample space, σ-field and probability function. Now we are going to introduce an interesting and important concept in probability theory, and it is known as random variables. Suppose, we have a random experiment, and based on the random experiment we have a sample space Ω. Now a random variable X is a real valued function defined on Ω. Note that Ω can have finite, countable or uncountable number of elements. Let us look at some example.

Example: Suppose we throw a dice twice and observe the faces which appear on the top. Clearly in this case the sample space Ω has 36 points, namely

$$\Omega = \{(i, j); i, j = 1, 2, \ldots, 6\}.$$

Suppose we define a function X on Ω as follows: $X(i, j) = i + j$. Clearly, X is a random variable and it can take values $2, 3, \ldots, 12$. If it is assumed that probability of appearing any face on the top is equally likely and the two dice are independently thrown, then

$$\begin{aligned}
P(X = 2) &= P(X = 12) = \frac{1}{36}, \quad P(X = 3) = P(X = 11) = \frac{2}{36}, \\
P(X = 4) &= P(x = 10) = \frac{3}{36}, \quad P(X = 5) = P(X = 9) = \frac{4}{36}, \\
P(X = 6) &= P(X = 8) = \frac{5}{36}, \quad P(X = 7) = \frac{6}{36}.
\end{aligned}$$

Example: Now let us look at the continuous case when the random variable X can take continuous values. Suppose we choose a number at random from $[0, 1]$, and let us denote that number as X. In this case $\Omega = [0, 1]$, and X is a random variable taking any values in $[0, 1]$. Since, it is assumed that the number is chosen at random, it means for and $0 < a < b < 1$,

$$P(a < X < b) = b - a.$$

A discrete random variable will be always characterized by the probability mass function (PMF). For a continuous random variable X if there exists a function $f(x)$, such that $f(x) \geq 0$, and for all $-\infty < a < b < \infty$,

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

then $f(x)$ is called the probability density function (PDF) of X. It is immediate that

$$\int_{-\infty}^{\infty} f(x)dx = P(-\infty < X < \infty) = 1.$$

From now on it is assumed that if $f(x)$ is a function satisfy

$$f(x) \geq 0, \quad \text{for all } x \quad \text{and} \int_{-\infty}^{\infty} f(x)dx = 1,$$

then there exists a random variable X, such that for all $-\infty < a < b < \infty$,

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

and $f(x)$ is the PDF of X.

Now we will introduce another important function, and it is known as the cumulative distribution function (CDF) or a distribution function (DF) of a random variable X. It is defined as

$$F_X(x) = P(X \leq x) = P(-\infty < X \leq x); \quad \text{for all } x.$$

Note that the CDF can be defined both for the discrete and continuous random variables. If X is a discrete random variable with the following PMF

$$P(X = a_i) = p_i; \quad i = 1, 2, 3 \ldots,$$

here a_i's are arbitrary real numbers, and p_i's satisfy the following continuous;

$$p_i \geq 0, \quad \text{for all } i, \quad \text{and} \quad \sum_{i=1}^{\infty} p_i = 1.$$

Then the CDF of X becomes

$$F_X(x) = \sum_{i: a_i \leq x} p_i.$$

It is clear that the distribution function of a discrete random variable is a step function, and it has jumps at $a_1, a_2, \ldots$.

Example: Suppose X has the following PMF

$$P(X=-1)=\frac{1}{4}, \quad P(X=0)=\frac{1}{2} \quad P(X=1)=\frac{1}{4}.$$

Then the CDF of X becomes

$$F_X(x)=\begin{cases} 0 & \text{if} \quad x<-1 \\ \frac{1}{4} & \text{if} \quad -1 \leq x<0 \\ \frac{3}{4} & \text{if} \quad 0 \leq x<1 \\ 1 & \text{if} \quad x \geq 1 \end{cases}$$

It can be seen that $F_X(x)$ is a step function, and it has jumps at $-1, 0, 1$.

If X is a continuous random variable with PDF $f(x)$, then the CDF of X becomes

$$F_X(x)=\int_{-\infty}^{x} f(u)du.$$

It is immediate that $F_X(x)$ is everywhere continuous.

Example: Suppose X is a continuous random variable with PDF

$$f(x)=e^{-2|x|}; \qquad -\infty<x<\infty,$$

then the CDF of X becomes

$$\begin{aligned} F_X(x)=\int_{-\infty}^{x} e^{-2|u|}du &= \begin{cases} \int_{-\infty}^{x} e^{2u}du & \text{if} \quad x<0 \\ \frac{1}{2}+\int_{0}^{x} e^{-2u}du & \text{if} \quad x \geq 0 \end{cases} \\ &= \begin{cases} \frac{1}{2}e^{2x} & \text{if} \quad x<0 \\ \frac{1}{2}+\frac{1}{2}(1-e^{-2x}) & \text{if} \quad x \geq 0 \end{cases} \end{aligned}$$

Now let us look at the some of the properties of a CDF. If $F(x)$ is a distribution function of a random variable, then it is clear

1. $0 \leq F(x) \leq 1$, for all x.
2. $\lim_{x \rightarrow -\infty} F(x)=0 \quad$ and $\quad \lim_{x \rightarrow \infty} F(x)=1$.

Now we will prove that $F_X(x)$ is right continuous for all x. Let a_n be a sequence of real numbers and $a_n \downarrow a_0$. Then let us consider the sets $A_n = \{X \in (-\infty, a_n]\}$, for $n = 1, 2, \ldots$. Note that $\bigcap_{n=1}^{\infty} A_n = A_0 = \{X \in (-\infty, a_0]\}$. Further,

$$\lim_{n\to\infty} F(a_n) = \lim_{n\to\infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right) = P(A_0) = F(a_0).$$

On the other hand let b_n be a sequence such that $b_n \uparrow a_0$, and if $B_n = \{X \in (-\infty, b_n]\}$, then $\bigcup_{n=1}^{\infty} B_n = B_0 = \{X \in (-\infty, a_0)\}$. Hence,

$$\lim_{n\to\infty} F(b_n) = \lim_{n\to\infty} P(B_n) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = P(B_0) = F(a_0-).$$

Therefore,

$$P(X = a_0) = P(-\infty < X \le a_0) - P(-\infty < X < a_0) = F(a_0) - F(a_0-).$$

From now on if a function $F(x)$ defined on the whole real line and satisfy the following properties

1. $0 \le F(x) \le 1$, for all x.
2. $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.
3. $F(x)$ is right continuous,

will be called a distribution function.

It easily follows from the definition of a distribution function:

$$P(a < X \le b) = F(b) - F(a), \quad P(a \le X \le b) = F(b) - F(a-)$$
$$P(a \le X < b) = F(b-) - F(a-), \quad P(a < X < b) = F(b-) - F(a).$$

Example Now consider the following function $F(x)$

$$F(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ 1 - \frac{1}{2}e^{-x} & \text{if} \quad 0 \le x < 2 \\ 1 & \text{if} \quad x \ge 2 \end{cases}$$

It is immediate that $F(x)$ satisfies all the above properties, hence, it is a proper CDF. It can be seen that it has jumps at 0 and 2. Here

$$P(X=0) = F(0) - F(0-) = \frac{1}{2}, \quad P(X=2) = F(2) - F(2-) = \frac{1}{2}e^{-2}.$$

It is clear that $F(x)$ is not a step function it is also not a continuous function. This kind of CDF is called a mixture distribution. We would like to write $F(x)$ in the following form

$$F(x) = \alpha F_c(x) + (1-\alpha)F_d(x).$$

Here $F_c(x)$ is a proper continuous CDF, $F_d(x)$ is a proper discrete distribution CDF and α is the mixing proportion. Let $g(x) = \frac{d}{dx}F(x)$. Therefore,

$$g(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \frac{1}{2}e^{-x} & \text{if} \quad 0 \le x < 2 \\ 0 & \text{if} \quad x \ge 2 \end{cases}$$

Therefore, $\alpha F_c(x) = \int_{-\infty}^{x} g(u)du$, and

$$\alpha F_c(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \frac{1}{2}(1-e^{-x}) & \text{if} \quad 0 \le x < 2 \\ \frac{1}{2}(1-e^{-2}) & \text{if} \quad x \ge 2. \end{cases}$$

Hence, $\alpha = \frac{1}{2}(1-e^{-2})$, $(1-\alpha) = \frac{1}{2}(1+e^{-2})$ and

$$F_c(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \frac{1-e^{-x}}{1-e^{-2}} & \text{if} \quad 0 \le x < 2 \\ 1 & \text{if} \quad x \ge 2. \end{cases}$$

Therefore, $(1-\alpha)F_d(x) = F(x) - \alpha F_c(x)$, and

$$(1-\alpha)F_d(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \frac{1}{2} & \text{if} \quad 0 \le x < 2 \\ \frac{1}{2}(1+e^{-2}) & \text{if} \quad x \ge 2. \end{cases}$$

and

$$F_d(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \frac{1}{1+e^{-2}} & \text{if} \quad 0 \le x < 2 \\ 1 & \text{if} \quad x \ge 2. \end{cases}$$