Random Variable

A *random variable* is a function that maps the outcome of a random experiment to a real number. We use $\mathcal{C}$ to denote the set of all outcomes; c to denote a single outcome.

For example, let $C = \{\text{GPT}, \text{GAN}, \text{BERT}, \text{YOLO}\}$, then a random varialbe X can be

$$X(\text{GPT}) = 0, X(\text{GAN}) = 1, X(\text{BERT}) = 2, X(\text{YOLO}) = 3.$$

You can change to any number you want.

We cannot observe a random variable itself, i.e., the mapping X is unobservable. We can only define the mapping, and then observe the result of applying this mapping to an experiment outcome.

The *realization* of a random variable is the result of applying the random variable (i.e., mapping) to an observed outcome of a random experiment. This is what we actually observe.

Typically, we use lowercase to denote the realized number; uppercase to denote the random variable. e.g., x is a realization of X.

Probability Mass Function

A random variable X is said to be *descrete* if its space $\mathcal{D}$ is either finite or countable.

Let X be a discrete random variable with space $\mathcal{D}$. The *probability mass function* of X, $p_X(d_i)$, is defined by

$$p_X(d_i) = P[\{c : X(c) = d_i\}] = P[X = d_i],$$

for all $d_i \in \mathcal{D}$.

The induced probability distribution, $P_X(\cdot)$, of X is

$$P_X(D) = \sum_{d_i \in D} p_X(d_i) = \sum_{d_i \in D} P[\{c : X(c) = d_i\}] = \sum_{d_i \in D} P[X = d_i],\ D \subset \mathcal{D}$$

Note that the notation $P[X = d_i]$ is an abbreviation, since the outcome c is not actually important here.

Cumulative Distribution Function

The *cumulative distribution function*, $F_X(x)$, of X is defined by

$$F_X(x) = P_X((-\infty, x]) = P[\{c : X(c) \leq x\}] = P(X \leq x).$$

Probability Density Function

A random variable X is said to be *continuous* if its cdf $F_X(x)$ is continuous for all $x \in \mathbb{R}$.

Let X be a continuous random variable with interval $\mathcal{D} \subset \mathbb{R}$ as space. The *probability density function* of X, $f_X(x)$, is a function that satisfies

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt.$$

When there exits such a function $f_X(x)$, X is also called an *absolutely continuous* random variable.

If $f_X(x)$ is also continuous, we have

$$\frac{d}{dx}F_X(x) = f_X(x)$$

by the Fundamental Theorem of Calculus. Note that for any continuous random variable X, there are no points of discrete mass, hence

$$P(X = x) = 0,$$

for all $x \in \mathbb{R}$.

From this, we can also infer that

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b)$$

Expectation

The *expectation* of X is defined by

$$E[X] = \begin{cases} \sum x_i p(x_i) & \text{if } X \text{ is discrete with pmf } p(x) \text{, and } \sum |x| p(x) < \infty \\ \int x f(x) dx & \text{if } X \text{ is continuous with pdf } f(x) \text{, and } \int |x| f(x) dx < \infty \end{cases} \tag{1}$$

Expectation is also called *mean*, or *expected value*, and mostly denoted by μ.

The expection can reflect the transformation of random variable. Let $Y = g(X)$, then

$$E(Y) = E(g(X)) = \sum g(x)p(x)$$
$$E(Y) = E(g(X)) = \int g(x)f(x)dx$$

The expection is linear with respect to random variable,

$$E[k_1 g_1(X) + k_2 g_2(X)] = k_1 E[g_1(X)] + k_2 E[g_2(X)]$$

Variance and Standard Deviation

Let X be a random variable with finite mean μ and $E[(X-\mu)^2]$ is also finite. The variance of X is defined by

$$var[X] = E[(X-\mu)^2] \tag{2}$$

Variance is mostly denoted by σ^2. The single σ is called the *standard deviation*. The number σ is sometimes interpreted as a measure of the dispersion of the points of the space relative to the mean value μ.

Note that

$$\begin{aligned} \sigma^2 &= E[(X-\mu)^2] = E(X^2 - 2X\mu + \mu^2) \\ &= E[X^2] - 2\mu^2 + \mu^2 \\ &= E[X^2] - \mu^2 \end{aligned}$$

Random Vector

Consider two random variables X_1 and X_2 on the same sample space $\mathcal{C}$, that they assign each element c of $\mathcal{C}$ one and only one ordered pair of numbers $X_1(c) = x_1$, $X_2(c) = x_2$. Then we say that (X_1, X_2) is a random vector. The *space* of (X_1, X_2) is the set of ordered pairs $\mathcal{D} = \{(x_1, x_2) : x_1 = X_1(c), x_2 = X_2(c), c \in \mathcal{C}\}$.

Probability Mass Function

A discrete random vector (X_1, X_2) with finite or countable space $\mathcal{D}$. The *joint probability mass function* of (X_1, X_2), $p_{X_1,X_2}(x_1, x_2)$, is defined by

$$p_{X_1,X_2}(x_1, x_2) = P[X_1 = x_1, X_2 = x_2]$$

for all $(x_1, x_2) \in \mathcal{D}$.

Cumulative Distribution Function

The cumulative distribution function of (X_1, X_2), $F_{X_1,X_2}(x_1, x_2)$, is defined by

$$F_{X_1,X_2}(x_1, x_2) = P[\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\}],$$

for all $(x_1, x_2) \in \mathbb{R}$. This is also called *joint cumulative distribution function.*
We'll also abbreviate $P[\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\}]$ to $P[X_1 \leq x_1, X_2 \leq x_2]$.

Probability Density Function

A random vector (X_1, X_2) with space $\mathcal{D}$ is said to be continuous if

$$F_{X_1,X_2}(x_1, x_2) = P[\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\}]$$

is continuous.
The joint probability density function of (X_1, X_2), $f_{X_1,X_2}(x_1, x_2)$, is defined to satisfy

$$F_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{X_1,X_2}(w_1, w_2) dw_1 dw_2$$

for all $(x_1, x_2) \in \mathbb{R}$. Then

$$\frac{\partial^2 F_{X_1,X_2}(x_1, x_2)}{\partial x_1 \partial x_2} = f_{X_1,X_2}(x_1, x_2)$$

For an event $A \subset \mathcal{D}$

Random Sample