Note: Random Variables and Distributions

Author: Hao Chung *last revised: May 20, 2020*

1 Random Variables

Definition 1 (random variable)**.** Consider an experiment with a sample space Ω. A *random variable* is a function from the sample space to the real numbers. That is,

$$X : \Omega \to \mathbb{R}.$$

In other words, a random variable is a real-valued function of the outcome of the experiment.

Example 2. Suppose we conduct a random experiment of tossing a coin 3 times. Let $X : \{H, T\}^3 \to \mathbb{R}$ be the random variable indicating the number of heads in the random experiment. Then we have

$$X((H,H,H)) = 3, X((H,H,T)) = 2, X((H,T,H)) = 2, X((T,H,H)) = 2,$$
$$X((H,T,T)) = 1, X((T,H,T)) = 1, X((T,T,H)) = 1, X((T,T,T)) = 0.$$

Remark 3. Note that the type of a random variable is a "function" but not a "variable." The reason why we call it random variable is that we often consider the functions of it.

We usually write $\Pr(X = x)$ to denote the probability of the event that $X = x$, which is formally defined by

$$\Pr(X = x) = \Pr(\{\omega \in \Omega : X(\omega) = x\}).$$

We also usually write $\Pr(X \leq x)$ (or $\geq, <, >$) to denote

$$\Pr(X \leq x) = \Pr(\{\omega \in \Omega : X(\omega) \leq x\}).$$

If A is a set, we write $\Pr(X \in A)$ to denote

$$\Pr(X \in A) = \Pr(\{\omega \in \Omega : X(\omega) \in A\}).$$

In general, we may care about the probability that some property P holds, where P can reflect a set of outcomes

$$\Pr(\text{property } P \text{ holds}) = \Pr(\{\omega \in \Omega : P(\omega) \text{ holds}\}).$$

Definition 4 (statistical distance)**.** Let X and Y be two random variables defined on the same probability space and with the same range D. The *statistical distance* between X and Y, denoted as $\delta(X, Y)$, is defined by

$$\delta(X, Y) = \frac{1}{2} \sum_{d \in D} \left| \Pr(X = d) - \Pr(Y = d) \right|.$$

2 Distribution Functions

Definition 5 (cumulative distribution function)**.** The *cumulative distribution function (CDF)* of a random variable X, denoted by $F_X(x)$, is defined by

$$F_X(x) = \Pr(X \leq x), \text{ for all } x.$$

The CDF of X is also called the *distribution function* of X.

We say a random variable is *continuous* if $F_X(x)$ is a continuous function of x; similarly, a random variable is *discrete* if $F_X(x)$ is a discrete function of x.

The random variable X has the distribution F_X is denoted as $X \sim F_X(x)$, where the symbol $\sim$ is read as "is distributed as."

We say two random variables X and Y have the same distribution if and only if they have the same distribution function; that is, $\Pr(X \leq x) = \Pr(Y \leq x)$ for all x.

Definition 6 (probability mass function)**.** The *probability mass function (PMF)* p_X of a discrete random variable X is defined by

$$p_X(x) = \Pr(X = x).$$

Definition 7 (probability density function)**.** Suppose $F_X(x)$ is the distribution function of a continuous random variable X. The *probability density function (PDF)* f_X of X is the function that satisfies

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \text{ for all } x.$$

Given a function $g : \mathbb{R} \to \mathbb{R}$, if $Y = g(X)$ is a function of a random variable X, then Y is also a random variable, because it provides a numerical value for each possible outcome. The probabilistic behavior of Y can be expressed in terms of X, which is defined by

$$\Pr(Y \in A) = \Pr(g(x) \in A), \text{ for any set } A.$$

With the definition above, the PMF p_Y of Y can be calculated by

$$p_Y(y) = \Pr(Y = y) = \Pr(g(X) = y) = \sum_{\{x:g(x)=y\}} \Pr(X = x) = \sum_{\{x:g(x)=y\}} p_X(x).$$

3 Expectation and Variance

Definition 8 (expectation)**.** The *expectation* $\mathbb{E}[X]$ (also called *expected value* or *mean*) of a random variable X is defined by

$$\mathbb{E}[X] = \begin{cases} \sum_{x \in \text{support}(X)} x \cdot p_X(x) & \text{, if } X \text{ is discrete r.v. with PMF } p_X \\ \int_{-\infty}^{\infty} x \cdot f_X(x)dx & \text{, if } X \text{ is continuous r.v. with PDF } f_X. \end{cases}$$

Let $Y = g(X)$. Notice that the expectation of Y should be calculated by $\mathbb{E}[Y] = \sum_{y \in \text{support}(Y)} y \cdot p_Y(y)$. However, the following proposition provides a convenient way to calculate $\mathbb{E}[Y]$ without knowing p_Y.

Proposition 9.

$$\mathbb{E}[Y] = \sum_{x \in support(X)} g(x) \cdot p_X(x).$$

Proof. Because $p_Y(y) = \sum_{\{x:g(x)=y\}} p_X(x)$, we have

$$\mathbb{E}[Y] = \sum_{y \in \text{support}(Y)} y \cdot p_Y(y) = \mathbb{E}[Y] = \sum_{y \in \text{support}(Y)} y \sum_{\{x:g(x)=y\}} p_X(x).$$

By direct calculation, we have

$$\sum_{y \in \text{support}(Y)} y \sum_{\{x:g(x)=y\}} p_X(x) = \sum_{y \in \text{support}(Y)} \sum_{\{x:g(x)=y\}} y \cdot p_X(x) = \sum_{y \in \text{support}(Y)} \sum_{\{x:g(x)=y\}} g(x) \cdot p_X(x).$$

Because

$$\bigcup_{y \in \text{support}(Y)} \{x : g(x) = y\} = \text{support}(X),$$

we have

$$\sum_{y \in \text{support}(Y)} \sum_{\{x:g(x)=y\}} g(x) \cdot p_X(x) = \sum_{x \in \text{support}(X)} g(x) \cdot p_X(x).$$

□

Definition 10 (variance and standard deviation)**.** The *variance* $\text{var}(X)$ of a random variable X is defined by

$$\text{var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

The *standard deviation* $\sigma(X)$ of a random variable X is defined by $\sigma(X) = \sqrt{\text{var}(X)}$. The variance $\text{var}(X)$ is often denoted by $\sigma^2(X)$.

Proposition 11 (variance in terms of moments expression)**.**

$$\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Proof.

$$\begin{aligned}
\text{var}(X) &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\
&= \sum_x (x - \mathbb{E}[X])^2 \cdot p_X(x) \\
&= \sum_x \left(x^2 - 2x\mathbb{E}[X] + (\mathbb{E}[X])^2\right) \cdot p_X(x) \\
&= \sum_x x^2 \cdot p_X(x) - 2\mathbb{E}[X] \sum_x x \cdot p_X(x) + (\mathbb{E}[X])^2 \sum_x p_X(x) \\
&= \mathbb{E}[X^2] - 2(\mathbb{E}[X])^2 + (\mathbb{E}[X])^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.
\end{aligned}$$

□