Lecture 6

With every random variable, we would like to associate a probability function that will represent the probability that X returns a specific value. Such functions are called distributions.

Definition 0.1 (Distribution)**.** *Let S be some set. A* distribution *over S is a function $\mu : S \to [0,1]$ such that the set $\{x \in S : \mu(x) \neq 0\}$ is finite or infinitely countable, and*

$$\sum_{x \in S} \mu(x) = 1.$$

The set $\{x \in S : \mu(x) \neq 0\}$ is called the support *of μ. We will also denote by*

$$\sum_{x} \mu(x),$$

the sum over all elements in the support of μ. This sum is well-defined since μ is a non-negative function.

Example 1:

1. Let $S = \mathbb{R}$, let $\mu(1) = \mu(0) = \frac{1}{2}$, and let $\mu(x) = 0$ for every $x \in \mathbb{R} \setminus \{0,1\}$.
2. Let $S = \mathbb{R}$, let $\mu(k) = \frac{1}{2^k}$ for every positive integer k, and let $\mu(k) = 0$ for every other real number k.
3. For every probability space $(\Omega, \mathbb{P})$, the function $\mathbb{P} : \Omega \to [0,1]$ is a distribution over $S = \Omega$.

Definition 0.2 (Distribution of a Random Variable)**.** *Let $(\Omega, \mathbb{P})$ be a probability space, and let $X : \Omega \to S$ be a random variable. The* distribution of X *is a distribution over S, which will be denoted by μ_X, and is defined as*

$$\mu_X(x) = \mathbb{P}(X = x) \text{ for every } x \in S.$$

Remark 0.3. *We will use $X = x$ to indicate the event $\{X = x\}$, which is in fact the event*

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

Example 2:

We revisit the cases listed in an example from Lecture 5.

1. A fair die is rolled twice, the two rolls being independent. Let X be the sum of the two outcomes. Then X is a random variable taking values in $\{2, 3, \ldots, 12\}$ with positive probability, that is, the set $\{2, 3, \ldots, 12\}$ is the *support* of μ_X. The distribution of X is then

$$\mu_X(2) = \mathbb{P}(X = 2) = \mathbb{P}(\{(1,1)\}) = \frac{1}{36},$$
$$\mu_X(3) = \mathbb{P}(X = 3) = \mathbb{P}(\{(1,2),(2,1)\}) = \frac{2}{36},$$

and so on. In general we have

$$\mu_X(j) = \begin{cases} \frac{j-1}{36} & \text{if } 2 \le j \le 7 \\ \frac{13-j}{36} & \text{if } 8 \le j \le 12 \\ 0 & \text{otherwise} \end{cases}$$

Indeed, let x_1 and x_2 denote the result of the first and second die rolls, respectively. Assume first that $x_1 + x_2 = j$ for some $2 \le j \le 7$. Since $x_2 \ge 1$, it follows that $1 \le x_1 \le j - 1$. Moreover, for every such value of x_1, there is exactly one possibility for the value of x_2, namely, $x_2 = j - x_1$. Assume now that $x_1 + x_2 = j$ for some $8 \le j \le 12$. Since $x_2 \le 6$, it follows that $j - 6 \le x_1 \le 6$; the number of choices for x_1 is thus $6 - (j - 6) + 1 = 13 - j$. Moreover, for every such value of x_1, there is exactly one possibility for the value of x_2, namely, $x_2 = j - x_1$.

We will sometimes write

$$X \sim \begin{cases} 2 & 1/36 \\ 3 & 2/36 \\ \vdots & \\ 12 & 1/36 \end{cases}$$

2. A coin with probability 1/3 for heads is tossed 4 times, all coin tosses being mutually independent. Let Y_1 be the number of heads in the first two tosses, and let Y_2 be the number of heads in the second and third tosses. Then the support of μ_{Y_1} is $\{0, 1, 2\}$ and its distribution is

$$Y_1 \sim \begin{cases} 0 & 4/9 \\ 1 & 4/9 \\ 2 & 1/9 \end{cases}$$

Similarly

$$Y_2 \sim \begin{cases} 0 & 4/9 \\ 1 & 4/9 \\ 2 & 1/9 \end{cases}$$

Observe that even though Y_1 and Y_2 are distributed exactly the same, i.e., $\mu_{Y_1} \equiv \mu_{Y_2}$, there is no equality between the two random variables, since as functions over Ω they return different values, e.g.,

$$Y_1\left((0,0,1,1)\right) = 0 \neq 1 = Y_2\left((0,0,1,1)\right).$$

In general, we say that two random variables, X and Y, *have the same distribution* if

$$\mathbb{P}\left(X = x\right) = \mathbb{P}\left(Y = x\right)$$

holds for every x.

Remark 0.4. *If $X = Y$ (as functions), then they have the same distribution.*

3. Let $(\Omega, \mathbb{P})$ be a probability space, and let $A \subseteq \Omega$ be an event. Then the *indicator of A* is distributed as follows

$$1_A \sim \begin{cases} 1 & \mathbb{P}\left(A\right) \\ 0 & 1 - \mathbb{P}\left(A\right) \end{cases}$$

We prove that the distribution of a random variable is indeed a distribution.

Claim 0.5. *The distribution of a random variable X over S is a distribution over S.*

Proof. Let $T = \{x \in S : \exists \omega \in \Omega \text{ such that } X(\omega) = x\}$ be the image of X. The support of μ_X is

$$\{x \in S : \mu_X(x) \neq 0\} = \{x \in S : \mathbb{P}\left(X = x\right) > 0\} \subseteq T,$$

where the inclusion holds since if $\mathbb{P}\left(X = x\right) > 0$, then $\{\omega \in \Omega : X(\omega) = x\} \neq \emptyset$. Since, moreover, $|T| \leq |\Omega|$ (as $X : \Omega \to T$ is onto T) and Ω is finite or infinitely countable (as $(\Omega, \mathbb{P})$ is a probability space), it follows that the support of μ_X is finite or infinitely countable. Since

$$\mu_X(x) = \mathbb{P}\left(X = x\right) \in [0, 1],$$

for every $x \in S$, it remains to verify that

$$\sum_x \mu_X(x) = 1.$$

Note that T is finite or infinitely countable and observe that

$$\sum_x \mu_X(x) = \sum_{x \in T} \mu_X(x) = \sum_{x \in T} \mathbb{P}\left(X = x\right).$$

Moreover

$$\bigcup_{x \in T} \{X = x\} = \bigcup_{x \in T} \{\omega \in \Omega : X(\omega) = x\} = \Omega,$$

where the last equality holds by the definition of T. Furthermore, these events are pairwise disjoint (as X is a function). Hence

$$\sum_{x \in T} \mathbb{P}\left(X = x\right) = \mathbb{P}\left(\bigcup_{x \in T} \{X = x\}\right) = \mathbb{P}\left(\Omega\right) = 1.$$

$\square$

0.1 Pairs of Random Variables

Definition 0.6 (Joint and Marginal Distribution)**.** *Let $X, Y : \Omega \to S$ be two random variables over the same probability space $(\Omega, \mathbb{P})$. Then we can view (X, Y) as a single random variable $(X, Y) : \Omega \to S^2$, defined as*

$$(X, Y)(\omega) = (X(\omega), Y(\omega)) \in S^2.$$

The distribution of this random variable is

$$\mu_{(X,Y)}(x, y) = \mathbb{P}\left(X = x, Y = y\right)$$

over S^2, and it will be called the joint distribution *of X and Y. The distributions of X and Y when considered separately, are called* marginal distributions*.*

Recall the following example from the previous lecture: a coin with probability 1/3 for heads is tossed 4 times, all coin tosses being mutually independent. Let Y_1 be the number of heads in the first two tosses, let Y_2 be the number of heads in the second and third tosses, and let Y_3 be the number of heads in the last two tosses. What is the *joint distribution* of Y_1 and Y_2? The support of both Y_1 and Y_2 is $\{0, 1, 2\}$. Hence the support of (Y_1, Y_2) is a subset of $\{0, 1, 2\}^2$. Since the support is small, it will be convenient in this case to represent the joint distribution in a table. The joint distribution of Y_1 and Y_2 is described in Table 1.

$Y_2 \backslash Y_1$	0	1	2
0	$\left(\frac{2}{3}\right)^3 = \frac{8}{27}$	$\frac{1}{3} \cdot \left(\frac{2}{3}\right)^2 = \frac{4}{27}$	0
1	$\frac{1}{3} \cdot \left(\frac{2}{3}\right)^2 = \frac{4}{27}$	$\left(\frac{1}{3}\right)^2 \cdot \frac{2}{3} + \frac{1}{3} \cdot \left(\frac{2}{3}\right)^2 = \frac{6}{27}$	$\left(\frac{1}{3}\right)^2 \cdot \frac{2}{3} = \frac{2}{27}$
2	0	$\left(\frac{1}{3}\right)^2 \cdot \frac{2}{3} = \frac{2}{27}$	$\left(\frac{1}{3}\right)^2 = \frac{1}{27}$

Table 1

We next compute the *joint distribution* of Y_1 and Y_3. Observe that for every x and y, the events $Y_1 = x$ and $Y_3 = y$ are independent, since they correspond to the outcomes of two pairs of independent coin tosses. Their joint distribution is described in Table 2.

Recall that the *marginal distribution* of Y_2 equals the *marginal distribution* of Y_3. However, Table 1 and Table 2 demonstrate that the *joint distribution* of Y_1 and Y_3 is different from the *joint distribution* of Y_1 and Y_2. In other words, the *marginal distributions* do not determine the *joint distribution*. On the other hand, the *marginal distributions* are determined by the *joint distribution*. This is done as follows

$$\{X = x\} = \bigcup_y (\{X = x\} \cap \{Y = y\}),$$

$Y_3\backslash Y_1$	0	1	2
0	$\left(\frac{4}{9}\right)^2$	$\left(\frac{4}{9}\right)^2$	$\frac{4}{9}\cdot\frac{1}{9}$
1	$\left(\frac{4}{9}\right)^2$	$\left(\frac{4}{9}\right)^2$	$\frac{4}{9}\cdot\frac{1}{9}$
2	$\frac{4}{9}\cdot\frac{1}{9}$	$\frac{4}{9}\cdot\frac{1}{9}$	$\frac{1}{9}\cdot\frac{1}{9}$

Table 2

and the events are pairwise disjoint. Therefore

$$\mathbb{P}(X=x)=\sum_{y}\mathbb{P}(X=x,Y=y).$$