Elements of Discrete Probability
2. Random Variables

1 Introduction

The term ”random variable” may be somewhat confusing. The modifier “random” distinguishes it from an ordinary variable in mathematics: in fact, the latter is any element of some set. A random variable, instead, is uniquely associated with a sample space, in fact it is a function $X : S \to D$ whose domain are the points of the sample space S and whose codomain is, in our considerations, the set D of the real numbers.

Consider this simple game of chance: A player rolls a fair six-sided die. If she rolls a number between 1 and 2 she receives \$3. If she rolls anything else she loses \$4. We can define a function that takes an outcome of the die roll (an event E) and maps it to the proper payoff X.

$$X(E) = \begin{cases} 3, & E \in \{1, 2\}, \\ -4 & E \in \{3, 4, 5, 6\}. \end{cases}$$

This prepares us for the following:

Definition 1 *A* random variable *is a function from the sample space of an experiment to the set of real numbers. That is, a random variable assigns a real number to each possible outcome (sample-space point).*

2 Properties of random variables

The assignment of probabilities to the points of a sample space induces an assignment of probabilities to the values of a random variable defined on that sample space. Consider a random variable X that has the value of the sum of two rolled six-sided fair dice. For example, $X(\{3, 2\}) = X(\{1, 4\}) = 5$. There are 6^2 possible ways of rolling two dice (the points of the sample space), but there are only eleven possible sum values, $2, 3, \ldots, 12$. Most sums, like 7, can be obtained in many different ways, while the sums 2 and 12 can only be obtained in just one way. Figure 1 shows a histogram of the sums (number of times each sum occurs) obtained by adding the outcomes of two dice rolls.

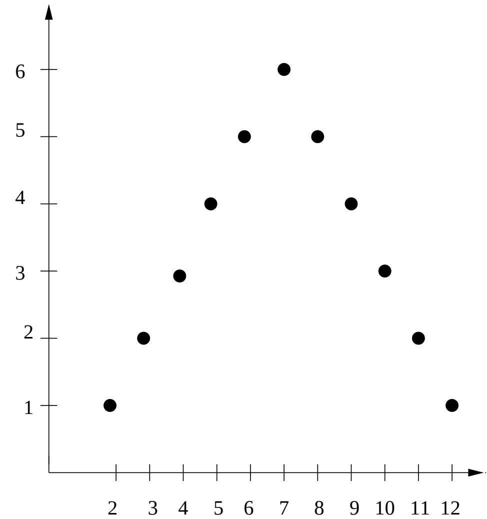

Figure 1: Frequency v. X

According to Figure 1, there are six pairs yielding the sum 7. We can say that $\Pr(X = 7) = 6/36 = 1/6$. Similarly, there is only one way to roll two dice to get $X = 2$, so $\Pr(X = 2) = 1/36$.

Example Given the experiment above, what is the probability that $4 \leq X \leq 10$? This is the sum of probabilities,

$$\begin{aligned}\Pr(4 \leq X \leq 10) &= \Pr(X = 4) + \Pr(X = 5) + \Pr(X = 6) \\ &\quad + \Pr(X = 7) + \Pr(X = 8) + \Pr(X = 9) + \ Pr(X = 10) \\ &= \frac{1}{36}(3 + 4 + 5 + 6 + 5 + 4 + 3) \\ &= \frac{5}{6}\end{aligned}$$

Let S be the sample space and X be a random variable defined on S. By computing the sum

$$\sum_{a \in S, X(a) = x} \Pr(a)$$

we obtain a function of x, the probability that the outcome of the experiment has value x for the random variable X. Formally:

Definition 2 *Let V be the set of the values of a random variable X. For any $x \in V$, the real-valued function f defined by*

$$f(x) = \Pr(X = x) = \sum_{a \in S, X(a) = x} \Pr(a) = \Pr(X^{-1}(x))$$

is called the density function *of X.*

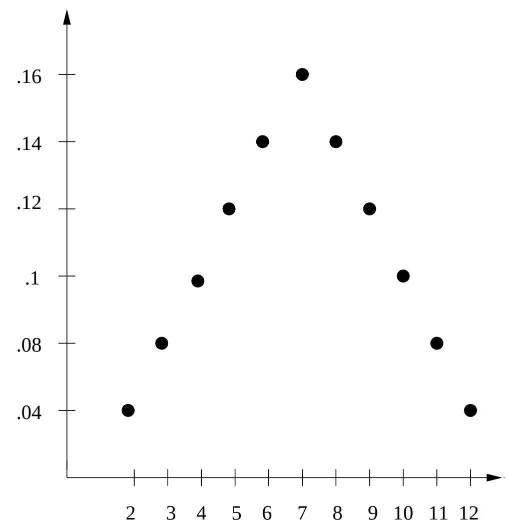

Figure 2: Density (mass) function of X

The term "density" is more appropriate for continuous random variables, that is when the set V of values is continuous. When V is countable (either finite or countably infinite) it is more appropriate to speak of "discrete density function" or "mass function". In the following, except when explicitly noted otherwise, we shall refer to **discrete** random variables; for uniformity, however, we shall use the terms "density funtions".

Figure 2 shows a graph of the density of the random variable X discussed earlier. Note the similarity between the shapes of the histogram in Fig 2.1 and of the graph of the density.

The density function f of a discrete random variable X has the following two properties:

1. $f(x) \geq 0, \, x \in R$
   Since $f(x)$ is a sum of probabilities, it cannot take on negative values.

2. $\sum_{x \in V} f(x) = 1$ Since each point of the sample space has a unique value of the random variable, the set $\{X^{-1}(x) | x \in V\}$ is a partition of the sample space S. By definition of probability over S, the property follows.

Figure 2 shows a graph of the density of the random variable X discussed earlier. Note the similarity between the the shapes of thehistogram in Fig 2.1 and the graph of the density.

Associated with each density function is a *distribution* function.

Definition 3 *The* distribution function $F(t), -\infty < t < \infty$ *(of a discrete random* *variable t) is defined by*

$$F(t) = Pr(X \leq t) = \sum_{x \leq t} f(x)$$

.

It is important to stress that the density function and the distribution of a random variable X carry the same information, since one is trivially obtainable from the other.

3 Moments of a distribution

We now study two important parameters of a density function. Consider a game of chance, characterized by a random variable X, the "payoff", which may assume the values $\{x_1, x_2, \dots, x_r\}$. Suppose we play the game n times, and let $X_1, X_2, \dots, X_n$ be the corresponding outcomes (payoffs). Although we are at a loss in predicting the outcome of each individual play, we can assume that the rules of the game are not changed while we play, so that the common density of the random variables $X_1, X_2, \dots, X_n$ is the same as the density $f(x)$ of X. Let $N_n(x_i)$ denote the number of plays in our n repetitions that yielded the value x_i. Then we can write

$$X_i + \dots + X_n = \sum_{i=1}^{r} x_i N_n(x_i)$$

and consider the *average* payoff (arithmetic mean0

$$\sum_{i=1}^{r} \frac{N_n(x_i)}{n} x_i$$

Based on our initial assimilation of relative frequencies and probabilities (to be formally substantiated later), we may replace $N_n(x_i)/n$ with the probability $f(x_i)$, and define the *expected value* of X.

Definition 4 *X is a discrete random variable with values $x_1, x_2, \dots$. The* (mathematical) expectation *$E[X]$ of X, also denoted μ and referred to also as* average *or* mean value*, is defined as*

$$E[X] = \sum_{j=1}^{\infty} f(x_j) x_j \tag{1}$$

Theorem 1 *If X and Y are two random variables on a space S, then $E[X+Y] = E[X] + E[Y]$.*

Proof:

$$
\begin{aligned}
E[X+Y] &= \sum_{s \in S} f(s)(X(s)+Y(s)) \\
&= \sum_{s \in S} f(s)X(s) + \sum_{s \in S} f(s)Y(s) \\
&= E[X] + E[Y]
\end{aligned}
$$

□

Theorem 2 *If X is a random variables on a space S and a a real number, then $E[aX] = aE[X]$.*

Proof

$$
\begin{aligned}
E[aX] &= \sum_{s \in S} f(s)aX(s) \\
&= a \sum_{s \in S} f(s)X(s) \\
&= aE[X]
\end{aligned}
$$

□

These important properties of the expectation, expressed by Theorems 1 and 2, are referred to as **linearity**.

To provide some intuitive underpinning to the notion of expectation, we note that if, in an extreme case, the density of random variable X is concentrated on a single value μ (i.e., $f(\mu)$=1), then trivially $E[X] = \mu$ tells us what we can"expect" when performing the experiment.

In general, the density is not so concentrated. However, if we imagine to slowly decrease the value of $f(\mu)$ and to distribute the decrement symmetrically around μ, then, while $E[X] = \mu$ remains as before, our uncertainty about an individual outcome of the experiment progressively increases as the "spread" increases. It appears therefore of great interest to have a measure of the spread of the density around its mean value. Such measure is provided by the parameter *variance*.

Definition 5 *The variance of X, denoted by $var[X]$ is defined by*

$$var[X] = E[(X - E[X])^2].$$

By expanding the right-hand and using the linearity properties of $E[x]$ and the fact that the expectation of a constant is the constant itself, we have

$$\begin{aligned} var[X] &= E[X^2 - 2XE[X] + E[X^2]] \\ &= E[X^2] - 2E[XE[X]] + E[X]^2 \\ &= E[X^2] - 2E[X]^2 + E[X]^2 \\ &= E[X^2] - E[X]^2 \end{aligned}$$

The variance is often denoted σ^2. The nonnegative number $\sigma = \sqrt{var[X]}$ is called the *standard deviation* of X. Clearly, the more X tends to deviate from its mean, μ, the larger $|X - \mu|$ tends to be, and hence the larger the variance becomes [1].

By their formal analogy with notions of mechanics, expectation and variance are referred to as *moments* of a random variable, in fact the *first* and *second* moments, respectively.

We conclude that the expectation provides an intuitive appreciation for the "bulk" of the probability, while the standard deviation provides a measure of the "spread" around the mean value.

4 Joint and conditional probabilities for random variables. Independence.

The notions of joint and conditional probabilities for events transfer in a straightforward manner to random variables. The key consideration is that the $\Pr(X = x)$ is the total probability of the subset $X^{-1}(x)$ of the sample space.

Suppose we have random variables X and Y defined on the same sample space S. What is the joint probability of $X = x$ and $Y = y$? Clearly, it is the total probability of all points $a \in S$ for which simultaneously $X(a) = x$ and $Y(a) = y$, that is (in the usual notation for random variables),

$$\begin{aligned} \Pr(X = x, Y = y) &= \Pr(\{a | a \in S, (X(a) = x) \wedge (Y(a) = y)\}) \\ &= \textstyle\sum_{a \in S, X(a) = x, Y(a) = y} \Pr(a) \\ &= \Pr(X = x) \Pr(Y = y | X = x) \end{aligned}$$

Again, two random variables defined on the same sample space are *independent* iff

$$\Pr(X = x, Y = y) = \Pr(X = x) \Pr(Y = y)$$

Example Consider the following random variables defined on the set of integers $S = \{1, 2, \ldots, 12n\}$:

[1] The spread, that is, the variance, should be insentitive to the sign of the deviation $X - \mu$; The choice of $(X - \mu)^2$ over $|X - \mu|$ is due to analytical advantages.

1. For $j \in S$, $X(j) = j \mod 3$.
2. For $j \in S$, $Y(j) = j \mod 4$

Are X and Y independent? To answer this question, we must evaluate the three terms

$$\Pr(X = x, Y = y), \Pr(X = x), \text{and} \Pr(Y = y)$$

. We construct the following array:

S	1	2	3	4	5	6	7	8	9	10	11	12	13	...	$12n$
$x = j \mod 3$	1	2	0	1	2	0	1	2	0	1	2	0	1	...	0
$y = j \mod 4$	1	2	3	0	1	2	3	0	1	2	3	0	1	...	0

We realize that that there exactly 3 distinct values of x with common probability 1/3, so that $\Pr(X = x) = 1/3$, 4 distinct values of y with probability 1/4,so that $\Pr(Y = y) = 1/4$ and 12 distinct pairs (x, y) with probability 1/12, so that $\Pr(X = x, Y = x) = 1/12$. Therefore,

$$\frac{1}{12} = \Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y) = \frac{1}{3}\frac{1}{4} = \frac{1}{12}$$

and we conclude that the two variables are independent. In what follows, we may replace the notations $\Pr(X = x)$,$\Pr(X = x, Y = y)$ with the shorter $\Pr(x)$,$\Pr(x, y)$.

We already know, from the linearity of expectation, that the expectation of a sum of random variables is the sum of they respective expectations (regardless of whether the variables are independent). What can we say about the product of two random variables X and Y? The answer is provided by the following theorem.

Theorem 3 *Let X and Y be independent random variables. Then*

$$E[XY] = E[X] \cdot E[Y]$$

Proof: From the definition of $E[]$

$$\begin{aligned} E[XY] &= \sum_{x,y} x \cdot y \Pr(x, y) \\ &= \sum_{x,y} x \cdot y \Pr(x)\Pr(y) \quad \text{(by independence)} \\ &= \sum_{x} x \Pr(x) \cdot \sum_{y} y \Pr(y) = E[X] \cdot E[Y] \end{aligned}$$

□

We also have:

Theorem 4 *Let X and Y be independent random variables. Then*

$$var[X+Y] = var[X] + var[Y]$$

Proof: From the expression

$$var[Z] = E[Z^2] - (E[Z])^2$$

we obtain

$$\begin{aligned}
var[X+Y] &= E[(X+Y)^2] - (E[X+Y])^2 \\
&= E[X^2+Y^2+2XY] - (E[X]+E[Y])^2 \quad \text{by linearity of } E[] \\
&= E[X^2]+E[Y^2]+2E[XY]-(E[X])^2-(E[Y])^2-2E[X]E[Y] \quad \text{by linearity of } E[] \\
&= (E[X^2]-(E[X])^2)+(E[Y^2]-(E[Y])^2)+2E[XY]-2E[X]E[Y] \quad \text{rearranging} \\
&= var[X]+var[Y]+2E[XY]-2E[X]E[Y]
\end{aligned}$$

Consider the rightmost two terms in the last expression. Since X and Y are independent by hypothesis, by Theorem 3 we have $E[XY] = E[X]E[Y]$, so that

$$var[X+Y] = var[X] + var[Y]$$

□