RANDOM VARIABLES AND THEIR DISTRIBUTIONS

Wiebe R. Pestman
Centre for Biostatistics, University of Utrecht, Holland

Keywords: Random variable, Probability distribution, Distribution function, Independence, Conditional probability distribution.

Contents

Summary

Today the idea of a random variable is basic in many branches of science. In this article it is explained how the concept of a random variable can be captured in proper mathematics. The most important feature of a random variable is the so-called probability distribution associated with it.

In the case of a real-valued variable the probability distribution is completely pinned down by the distribution function. Probability distributions are traditionally classified into (absolutely) continuous and discrete ones. In basic courses on probability there often seem to be two theories of probability: One for absolutely continuous and one for discrete random variables. In this chapter an effort has been made to get around this artificial 'dichotomy'.

In terms of (joint) distribution functions the fundamental concept of independence is explained. Related to independence is the idea of conditional probability distributions, a concept playing an important role in many applications. In the last section it is explained how the probability distribution of a random variable can naturally be presented as a Borel measure.

1. Introduction

A *random variable* or *stochastic variable* is an observation or measurement of some aspect of some probability experiment. Otherwise stated, a random variable is a measurement which, when repeated under exactly the same conditions, does not necessarily result in the same outcome. Very often such measurements can be captured or summarised in the form of a number. Random variables are nowadays usually denoted by capital letters such as X, Y, Z and then their outcomes by the corresponding small letters x, y, z respectively. So X stands for a measurement that is still to be carried out and x stands for the result of the measurement once the experiment is finished. In X there are elements of probability whereas x is a definite object, very often some real number.

Example 1: In a game of chance a die is thrown twice. Conclusive factor in the game is the total number of pips thrown, which will be denoted by X. This X is evidently a random variable: When repeating the experiment, X does not necessarily assume the same value. In X there are elements of probability. For example, one could ask oneself about the probability that X is going to assume an outcome higher than ten. If the first throw shows three and the second four pips, then the outcome x of X is equal to $3+4=7$. In this outcome there is no element of probability anymore.

A probability experiment can mathematically be captured in terms of a Kolmogorov triplet $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the set of all possible outcomes, $\mathcal{F}$ the collection of all selected events, and $\mathbb{P}$ a function that assigns to every selected event the probability that it is going to occur. In the following the elements of Ω, the possible outcomes of the experiment, will be denoted by the letter ω. Events can be identified with subsets of Ω and will be denoted by the letter *E*. In this identification an event *E* occurs if the outcome of the experiment is an element of *E*. In this setting a real-valued random variable X, showing some aspect of the probability experiment $(\Omega, \mathcal{F}, \mathbb{P})$, can be defined as a function that assigns, in some prescribed way, to every outcome ω of the experiment a real number $X(\omega)$. Briefly, a real-valued random variable can in this setting be defined as a function $X:\Omega \to \mathbb{R}$, where $\mathbb{R}$ denotes the set of all real numbers.

Example 2: In the probability experiment of throwing a die twice the possible outcomes can be characterised by ordered pairs of numbers (i, j), where *i* and *j* are the numbers 1,2,3,4,5,6. Thus a specific outcome such as $\omega = (2,5)$ indicates that the first throw showed two pips and the second five. The set Ω is here given by

$$\Omega = \{(1,1),(1,2),\,\ldots\,,(6,6)\} = \{(i,j) \mid i,j = 1,2,3,4,5,6\}.$$

With every subset *E* one can associate the event that the outcome ω is going to be an element of *E*. Thus the subset

$$E := \{(2,1),(2,3),(2,5),(4,1),(4,3),(4,5),(6,1),(6,3),(6,5)\}$$

corresponds to the event that the first throw is going to be even and the second odd. For every subset E of Ω one can talk about the probability that the corresponding event will occur. Namely, this probability is, when all possible outcomes are equally probable, generally given by

$$\mathbb{P}(E) = \frac{\text{number of elements in } E}{\text{number of elements in } \Omega} = \frac{\text{number of elements in } E}{36}.$$

The probability that the first throw will be even and the second odd can thus easily be computed to be

$$\mathbb{P}(E) = \tfrac{9}{36} = \tfrac{1}{4}$$

where E is the 9-element subset defined above. The total amount X of pips thrown is one aspect of this experiment. It can be captured as a function $X : \Omega \to \mathbb{R}$ by setting $X((i, j)) := i + j$. So, if the outcome is given by $\omega = (3,4)$ then the outcome x of X is: $x = X((3,4)) = 3+4$. The event that one is going to observe that $X \geq 10$ corresponds to the 6-element subset

$$E = \{\, (4,6), (6,4), (5,5), (5,6), (6,5), (6,6) \,\}.$$

The probability that this will occur is

$$\mathbb{P}(E) = \frac{\text{number of elements in } E}{36} = \tfrac{6}{36} = \tfrac{1}{6}.$$

The probability above is often briefly denoted as $\mathbb{P}(X \geq 10)$.

2. The Distribution Function of a Random Variable

Given some real-valued random variable X and some arbitrarily chosen numerical value v one could ask: What is the probability that X will show an outcome $\leq v$? This probability is usually denoted as $\mathbb{P}(X \leq v)$; it evidently depends on the value of v. This dependence can be captured in the form of a function $F : \mathbb{R} \to [0,1]$ by setting

$$F(v) := \mathbb{P}(X \leq v) \qquad \text{for all } v \in \mathbb{R}. \tag{1}$$

The function F is called the *cumulative probability distribution function* or more briefly the *distribution function* of X. If necessary the function can be denoted by F_X to emphasise that it belongs to the variable X. Such a distribution function is always an increasing function, increasing from 0 to 1 when v runs from $-\infty$ to $+\infty$. The function may show discontinuities, exhibiting themselves as jumps in its graph (see Figure 1 below).

If such a jump occurs, for example at $v=1$, then the size of the jump is equal to the probability that X will show the outcome 1. Generally one has

$$\mathbb{P}(X=v) = \text{size of jump at } v. \qquad (2)$$

When there is no jump at v then the size of the jump is understood to be zero. For example, in Figure 1 the jump at $v=5$ is zero. Hence $\mathbb{P}(X=5)=0$. This should be read as follows: The probability that X will show the outcome 5 with *infinite precision* is equal to zero. The probability that X will show an outcome between for example 4.9 and 5.1 is equal to $F(5.1)-F(4.9)$. More precisely and more generally one has

$$\mathbb{P}(a<X\le b) = F(b)-F(a). \qquad (3)$$

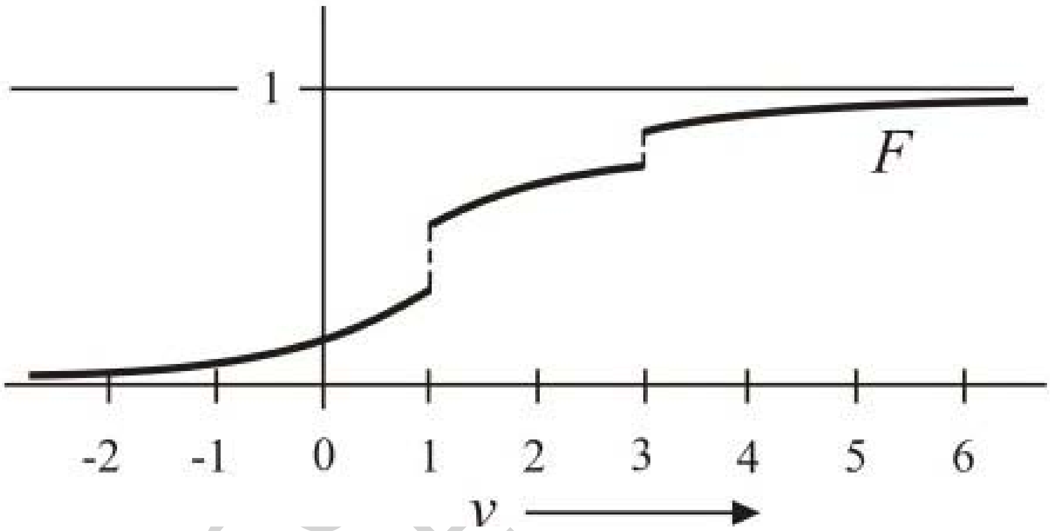

Figure 1: Distribution Function

Two real-valued random variables X and Y are said to be *identically distributed* if their distribution functions are identical. This is usually denoted by something like $X \simeq Y$ or $X \triangleq Y$. It should be noted here that $X \simeq Y$ does not imply $X = Y$. Here is an easy example in this:

Example 3: Throwing a die twice, let X be the number of pips in the first and Y the number of pips in the second throw. Of course then X and Y can very well show different outcomes, hence $X \neq Y$. However, X and Y evidently have identical distribution functions and therefore one does have $X \simeq Y$ here.

Being identically distributed is not compatible with addition or multiplication. That is to say, it can very well occur that $X \simeq V$ and $Y \simeq W$, but that neither the variables in the pair $X+Y$ and $V+W$ nor those in the pair XY and VW are identically distributed. This negative phenomenon is illustrated in the next example:

Example 4: Let X and Y be as in Example 3 and let $V = X$ and $W = X$. Then, of course, one has $X \simeq V$ and $Y \simeq W$. However, the variables $X+Y$ and $V+W$ are not identically distributed. Namely, $V+W$ is actually equal to $2X$ and can therefore assume even outcomes only. The variable $X+Y$, however, can very well show an odd outcome. Hence $X+Y$ and $V+W$ can impossibly be identically distributed. Similar arguments show that XY and VW are not identically distributed in this example.

From the mathematical point of view notations like $\mathbb{P}(X \le v)$, which were used to define the distribution function of a random variable X, are nonchalant. Actually, the event that X will show an outcome $\le v$ is defined by the subset

$$E = \left\{ \omega \in \Omega \;\middle|\; X(\omega) \le v \right\} \tag{4}$$

where $(\Omega, \mathcal{F}, \mathbb{P})$ is assumed to be the underlying probability space of X. One can only speak about the probability that E is going to occur, that is, about $\mathbb{P}(E)$, if E is a member of $\mathcal{F}$.

For this reason it is usually incorporated in the mathematical definition of a real-valued random variable X that subsets of the form $\{\omega \in \Omega \mid X(\omega) \le v\}$ be members of $\mathcal{F}$. In mathematical measure theory this is expressed by saying that X is $\mathcal{F}$*-measurable.*

3. Classification of Random Variables

In mathematics a set A, for example a set of real numbers, is said to be *countable* if it can be presented as a sequence of elements. That is to say, a set A is countable if it can be defined as

$$A := \{ a_1, a_2, a_3, \ldots \} \tag{5}$$

Here the sequence $a_1, a_2, a_3, \ldots$ may be finite or infinite. It is known that there exist sets that are not countable. For example, it is impossible to present the elements of the unit interval [0,1] as a sequence of elements. Hence the unit interval is not countable.

For an arbitrary random variable X the *range* is understood to be the set of all possible outcomes of X. The range of a variable X will be denoted as M_X. If the underlying probability space of X is $(\Omega, \mathcal{F}, \mathbb{P})$ then the range of X can be presented as

$$M_X = \left\{ X(\omega) \;\middle|\; \omega \in \Omega \right\}. \tag{6}$$

A random variable X is said to be *discrete* if M_X is a countable set. So in the case of a discrete variable X the possible outcomes of X can be listed as a finite or infinite sequence $x_1, x_2, x_3, \ldots$ Denoting the probability that X will show outcome x_i by p_i, one can make a so-called *probability table* for X:

x	x_1	x_2	x_3	$\cdots$
$\mathbb{P}(X=x)$	p_1	p_2	p_3	$\cdots$

Figure 2: Probability table for X

The table in Figure 2 can be visualised in the form of a needle diagram in which the length of the i^{th} needle is equal or proportional to p_i and their sum equal to one:

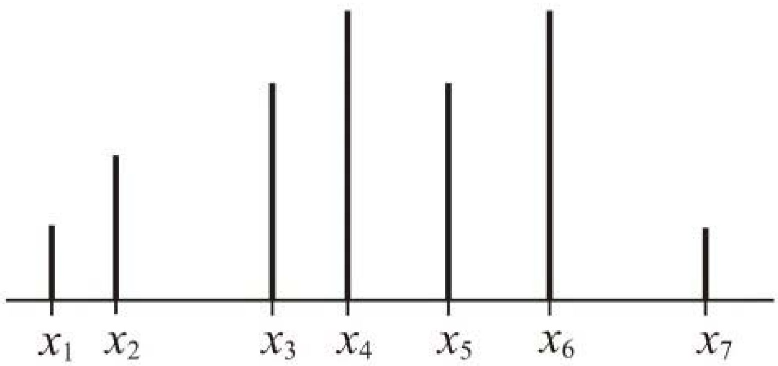

Figure 3: The distribution function of a discrete variable X is related to the p_i

The distribution function of a discrete variable X is related to the p_i as

$$F(v) = \sum_{x_i \le v} p_i = \text{sum of all } p_i \text{ for which } x_i \le v. \tag{7}$$

Distribution functions of discrete random variables as a rule show in their graphs a stair-shape, such as sketched in Figure 4 below:

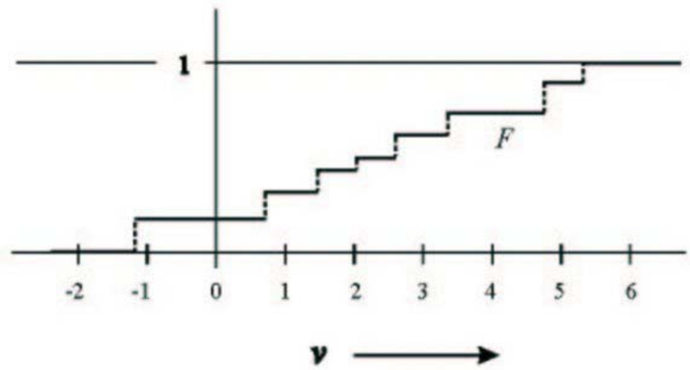

Figure 4: Distribution functions of discrete random variables

The jumps occur at places where $v = x_i$. Conversely, every random variable showing a stair-shaped distribution function is discrete. Thus discreteness of a random variable can be read off from its distribution function.

A random variable is called *continuous* if its distribution function is continuous, that is to say, without any jumps in its graph.

Figure 5 shows an example of such a distribution function. A special class of continuous random variables are those that are called ‘absolutely continuous’: A random variable X is said to be *absolutely continuous* if its distribution function F can be presented as a left-tail area function of some function $f:\mathbb{R}\to\mathbb{R}_+$. The latter function is then called the *probability density function* of X.

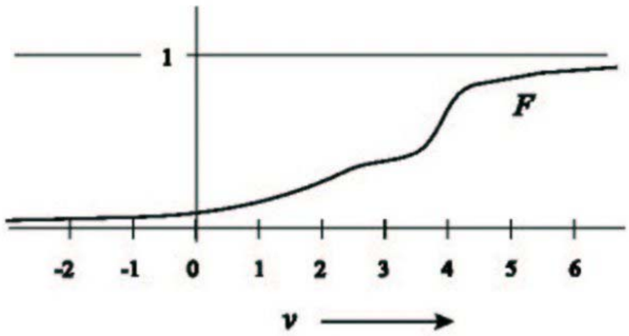

Figure 5: An example of such a distribution function

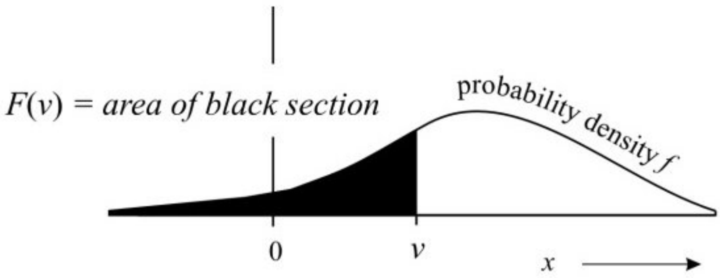

Figure 6: Probability density function

Mathematically this can be described in terms of an integral:

$$F(v) = \int_{-\infty}^{v} f(x)\, dx . \tag{8}$$

As a rule, aside from some difficult mathematical complications in cases where f is not continuous, one then has

$$F'(x) = f(x) \tag{9}$$

where $F'(x)$ is the derivative of F at x.

The classification of random variables into discrete and continuous variables is exclusive but not exhaustive. Namely, it can occur that a variable is neither discrete nor continuous. Figure 1 shows, as an example, a distribution function of which the underlying random variable is neither discrete nor continuous. In probability literature the distinction between 'continuous' and 'absolutely continuous' is often disregarded. It is indeed so, as already suggested by the terminology, that an absolutely continuous variable is always continuous. It can, however, very well occur that a random variable is continuous, but not absolutely continuous. In the past such variables were considered to be not of any practical value. Today, however, a prominent role is allotted to them in theoretical physics.

-
-
-

Bibliography

Ash, R.B. & Doléans-Dade, C.A.: Probability and measure theory. Academic Press, San Diego 1999. [An introduction to probability theory with an emphasis on its natural connection to mathematical measure theory. The concepts of Lebesgue and Stieltjes integrals, which were occasionally used in this article, are also explained.]

Bauer, H.: Probability Theory. Walter de Gruyter, Berlin 1996. [A profound introduction to probability theory with rigorous mathematical proofs. Very clear presentation. Intended for mathematics students. Also available in the German language, in which it was originally published.]

Bourbaki, N.: Éléments de Mathématiques, fascicule **35**, livre **6**, Intégration, chapitre **9**. Hermann, Paris 1969. [In this chapter of the celebrated series on mathematics a thorough introduction to infinite-dimensional normal distributions and Brownian motion can be found.]

Chandra, T.K. & Chatterjee, D.: A First Course in Probability. Alpha Science, Pangbourne, UK 2001. [A fine modern introduction to probability theory. The development is systematic and rigorous. Many illustrating examples.]

Feller, W.: An Introduction to Probability Theory and its Applications, vol. I, II. John Wiley & Sons, New York 1971. [Volume I is an excellent introduction to the subject. In volume II more advanced topics are treated, such as for example the theory of Brownian motion. In spite of the fact that the notations used are a bit old-fashioned, both volumes are very pleasant to read. Lots of interesting examples.]

Grimmett, C.R. & Stirzaker, D.R.: Probability and Random Processes. Oxford University Press, Oxford 1993. [A well-written modern introduction to the subject. Many examples. Good balance between rigour and readability.]

Pestman, W.R.: Mathematical Statistics. Walter de Gruyter, Berlin 1998. [An introduction to statistics and the probability distributions used in this branch of mathematics. Most theorems are proved rigorously. Lots of examples illustrating the theory. No much mathematical foreknowledge required.]

Ventsel, H.: Théorie des Probabilités. Edition Mir, Moscou 1977. [An excellent first introduction to probability theory. Very little mathematical foreknowledge required. A clear, pragmatic and pleasant style. Also available in the Russian language, in which it was originally written.]

Biographical Sketch

Wiebe R. Pestman has studied mathematics, physics and astronomy at the University of Groningen, where he also got his doctor's degree in mathematics. He then quit scientific life for a while. Eventually he returned to maths and became lecturer at the University of Nijmegen. At the moment Pestman is lecturer at the University of Utrecht. His mathematical interests are in functional analysis, probability and statistics, operator algebras and harmonic analysis. He authored several publications on various subjects in mathematics, including a textbook on mathematical statistics.