Random Variables

<div style="text-align: right">**4**</div>

Students of economics are confronted with random variables very early in their programs. They are confronted with this term not only in statistics and econometrics but practically in all economic subdisciplines, in particular in microeconomics and finance. The meaning of a random variable, however, remains somewhat vague. It is usually considered sufficient if students understand it to be data whose actual value is not guaranteed. However, we will not remain on the surface but provide more fundamental insights of random variables. The reader will learn that random variables are *functions* with specific properties.

A Standard Example of Random Variables Assume that an experiment is carried out where the respective daily yields of both the S&P 500 index $x_1, \ldots, x_n$ and the Apple stock $y_1, \ldots, y_n$ are determined on all trading days of a year.[1] A plot of the daily yields presented in pairs may help to support the assumption that there is a linear correlation between the yield of the Apple stock and the S&P 500. A model of the form

$$y_i = \alpha + \beta x_i + \varepsilon_i \tag{4.1}$$

is used to estimate the regression line with α and β being the relevant parameters. It is commonly assumed that the interfering (noise) terms ε_i are independent of each other and have identical probability distributions. Typically, the interfering terms have an expected value of $E[\varepsilon_i] = 0$ and a variance $Var[\varepsilon_i] = \sigma^2$. If the noise is normally distributed, one usually writes $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.

While this depiction may not be a problem for most economic applications, it is far too simple for readers interested in a closer look at probability.

---

[1]The (log) daily yield of a financial asset is easily determined with the help of $\ln\left(\frac{\text{current rate}}{\text{previous rate}}\right)$.

Referring to the above comments on the regression function one will notice that the interference terms follow a particular distribution but nothing is being said about the underlying state space Ω. The state space was not even mentioned. Is it infinite? Does it include the real numbers or is it a larger set such as the space of the continuous functions $C[0, \infty)$? What is the relation between the states $\omega \in \Omega$ and the realizations observed? One would probably call this relationship "causal," because the state generated the realization which has occurred. While all these questions remain typically unanswered, at best the realizations with their probabilities are stated.

In Eq. (4.1) there exists a random variable ε_i, but it remains absolutely unclear what the connection between "a random event" $\omega \in \Omega$ and the "random-driven variable" y_i looks like. We are going to clarify this causal relationship in the following section.

4.1 Random Variables as Functions

We can certainly state that ε_i is influenced by "randomness" and can take different values. In order to express this relation formally, in a first step chance draws an arbitrary element ω from the state space Ω. Second, this state ω then exerts a causal influence. The resulting quantity $\varepsilon(\omega) = \varepsilon_i$ should always be a real number. This allows us to use a random variable ε as a function

$$\varepsilon : \Omega \rightarrow \mathbb{R}\,. \tag{4.2}$$

We will now illustrate the view of random variables with several examples.

Example 4.1 (St. Petersburg Paradox)   The St. Petersburg paradox is often discussed in decision theory. The formalism we have presented so far is particularly useful to describe this game.

Consider an experiment performed only once. The game master tosses a coin until "heads" appears. The payment to the participant is given in Table 4.1 and depends on the number of tosses required to obtain "heads" for the first time. Although the expected value of the payment is infinite,[2] hardly anyone is willing to sacrifice more than $10 to participate in the game.

A binomial model is used successfully in order to describe this game formally. Heads are represented by u and tails by d. An elementary event is a sequence of

---

[2]With a fair coin the two events "heads" and "tails" are equally probable. Then the expected payment of the game is

$$\lim_{n \to \infty} \sum_{i=1}^{n} 2^i 2^{-i} = \lim_{n \to \infty} n = \infty$$

Table 4.1 St. Petersburg
paradox payment

First "heads"	Payment $\varepsilon(\omega)$
After 1 toss	$2^1 = 2$
After 2 tosses	$2^2 = 4$
After 3 tosses	$2^3 = 8$
$\vdots$	$\vdots$
Never	0

tosses, i.e., an element[3]

$$\omega \in \{u, d\}^{\mathbb{N}}. \tag{4.3}$$

If one wants to determine the number of tosses necessary for the game to end, it is the natural number associated with the first u in state ω. Since it is at least conceivable that no tails will ever appear, one must differentiate two cases by defining the following function:

$$g(\omega) := \begin{cases} k, & \exists k \in \mathbb{N} \quad d = \omega_1 = \ldots = \omega_{k-1}, \quad u = \omega_k, \\ 0, & \text{else.} \end{cases} \tag{4.4}$$

The payment in dollars is calculated as follows:

$$\text{payment } \varepsilon(\omega) = 2^{g(\omega)}. \tag{4.5}$$

Example 4.2 (Dice Roll) We roll a dice and note that the payment is double the score. In this case the random variable corresponds to the payment in dollars and can be described as

$$\varepsilon(\omega) = 2 \cdot \omega. \tag{4.6}$$

ω varies from $1, \ldots, 6$. Let us now discuss a more difficult example.

Example 4.3 (Continuous-Time Stock Prices) The state space consisting of the set of all continuous functions $\Omega = C[0, \infty)$ is required for the construction of the Brownian motion. This state space is the natural candidate for considering stock prices that vary continuously in time. Every elementary event $\omega \in \Omega$ is a function of real numbers. Hence, we can also write $\omega(t) : \mathbb{R} \to \mathbb{R}$ with t being time.

   If we want to construct a random variable for the event space Ω, we must determine the real number which is generated by an event ω. The value of the

---

[3]Instead of $\{u, d\}^{\mathbb{N}}$ it is possible to write also $\{u, d\}^{\infty}$.

random variable $\omega(t)$ ("effect") is the realization of the event ω ("cause") at a predetermined time t. The random variable is denoted by

$$\varepsilon(\omega) := \omega(t). \tag{4.7}$$

Obviously, it is the value of one of an infinite number of functions $\omega \in \Omega$ at time t.

Example 4.4 Instead of focusing on a single point in time we are interested in the average of all values of the function $\omega(\cdot)$ in the interval $[0, t]$. In other words we are not restricting ourselves to the value of the elementary event at time t but are interested in the average of a finite time interval. This random variable would be defined in the form

$$\varepsilon(\omega) := \frac{1}{t} \int_0^t \omega(s)\, ds\,. \tag{4.8}$$

4.2   Random Variables as Measurable Functions

Not every function is a random variable. There are two classes of functions, those that are random variables and those that are not. In order for functions to be called random variables, they must have a certain property which will be discussed on page 65. As a prerequisite to that discussion one has to understand why we look at random variables at all.

In dealing with random variables we are primarily interested in their probabilities. However, assigning probabilities to realizations of random variables is not always an easy task. In the following two examples we show initially the case where the assignment of probabilities does not create any difficulties and subsequently where it will.

Example 4.5 (Dice, Ideal and Manipulated) In the case of a dice roll one can assign to each realization a corresponding probability, regardless whether the dice is ideal or manipulated. In doing so the inverse function ε^{-1} would have to be considered and we can determine the probability of the outcome. Formally this would be $a \in \mathbb{R}$ for a specific realization, so

$$\mu(\varepsilon^{-1}(a)) : \mathbb{R} \;\rightarrow\; [0, 1]. \tag{4.9}$$

To illustrate the above let us assume that the payment after a roll is double the score. Our mapping would be the same as in Example 4.2 on page 61

$$\varepsilon(\omega) = 2 \cdot \omega\,. \tag{4.10}$$

Table 4.2 Probabilities of two dice

Dice\score	1–5	6
Ideal	$\frac{1}{6}$	$\frac{1}{6}$
Manipulated	$\frac{2}{15}$	$\frac{5}{15}$

With respect to this random variable we can specify the probabilities directly: since the inverse function $\varepsilon^{-1}(a) = \frac{a}{2}$ exists, the corresponding probability can be calculated easily. For each $a = 1, \ldots, 6$ the probability amounts exactly to $\frac{1}{6}$.

With a manipulated dice, for example a score of six would be rolled with a higher probability than with an ideal dice, this manipulation would be reflected in the function $\varepsilon(\omega)$. The inverse function $\varepsilon^{-1}(6)$ would not return $\frac{1}{6}$ but a higher value; correspondingly the other scores must have lower probabilities.

To grasp this, imagine two different dice, an ideal dice and one being manipulated. With the manipulated dice the occurrence of a score of six is twice as likely as with the ideal dice. Since we can clearly assign a probability to each realization of both dice according to the following table, the dice are clearly distinguishable from each other (Table 4.2).

Given the payment rule (4.10) it is easy to conclude which dice is rolled: if the ideal dice is rolled over and over again, the payment of \$12 (equivalent to a score of 6) will occur as often as a payment of \$6 (equivalent to a score of 3); if however the manipulated dice is rolled, \$12 are paid out much more often than \$6.

As shown in the following example matters are not always as simple as illustrated above.

Example 4.6 Let the state space cover the set of all real numbers, $\Omega = \mathbb{R}$. For any real number drawn by chance, the payment shall again be twice the real number as postulated in Eq. (4.10), i.e., $\varepsilon(\omega) = 2\omega$. All we need to do now is to specify how we will measure the probability of an event in $\mathbb{R}$. For this purpose we use the Stieltjes measure μ introduced above,[4] leaving the actual function g unspecified for the moment.

Constructing the inverse function as in (4.9) and measuring the probability, we obtain an extremely unsatisfactory result. If the state $\frac{a}{2}$ occurs the payment of $a \in \mathbb{R}$ will result. The probability that this will be the case can be determined directly: it is simply zero because $\mu([\frac{a}{2}, \frac{a}{2}]) = g(\frac{a}{2}) - g(\frac{a}{2}) = 0$. This result is entirely independent of the function g chosen. One must realize that a different procedure is required.

With the dice roll example the probabilities of the payments are always positive and allow us to determine whether the ideal or the manipulated dice was rolled. The probability of a score of 6 points and a payout of \$12 is significantly higher when rolling the manipulated dice.

---

[4]See page 52.

With the real number example, however, we cannot achieve a similar result because the probability of a payout is always zero, regardless of whether one uses the function g_1 (analog to the ideal dice) or g_2 (analog to the manipulated dice).

The solution to the problem is not to focus on a particular realization, but on an interval of realizations. We no longer ask which state results *exactly in the value a*, rather we ask which states will deliver realizations *between the values b and a* with $b < a$. This leads to a meaningful result. We have to ask when the state ω returns a value from the interval $[b, a]$. Hence

$$
\begin{aligned}
\mu(\varepsilon^{-1}([b, a])) &= \mu \{\omega \,:\, \varepsilon(\omega) \in [b, a]\} \\
&= \mu \{\omega \,:\, 2 \cdot \omega \in [b, a]\} \\
&= \mu \left\{\omega \,:\, \omega \in \left[\frac{b}{2}, \frac{a}{2}\right]\right\} \\
&= \mu \{[b, a]\} \\
&= g\left(\frac{a}{2}\right) - g\left(\frac{b}{2}\right).
\end{aligned}
\tag{4.11}
$$

Obviously the particular function g has a direct influence on the probability that the realization of the random variable lies in the interval $[b, a]$.

However, our proposal also has a weakness. The probability that a realization ω will fall in the interval $[b, a]$ depends on two variables b and a. This is a multidimensional function, and functions like these are always difficult to handle. It makes sense to standardize the first variable b, and $b \rightarrow -\infty$ has proven to be useful. It is a common practice to omit the equal sign in $-\infty < \varepsilon(\omega) \le a$. This finally gives us the definition used nowadays to characterize a random variable.

> Using a random variable will answer the question: what is the probability of an event leading to a realization *being less than a*?

For each random variable ε, we are considering the probability

$$
\mu \left(\{\omega \,:\, \varepsilon(\omega) < a\}\right) \ .
\tag{4.12}
$$

This function, depending on a, is called *distribution function* of ε. However, we still have to make sure that the set $M := \{\omega \,:\, \varepsilon(\omega) < a\}$ is measurable.

Definition 4.1 (Random Variables) A function ε is called a random variable if for each real number a the event

$$F_\varepsilon(a) := \{\omega \in \Omega \ : \ \varepsilon(\omega) < a\} \tag{4.13}$$

is measurable. Random variables are therefore also called measurable functions. $F_\varepsilon(a)$ is the distribution function of the random variable ε.

The definition of the distribution function allows us to establish something similar to "probabilities for certain realizations." The derivative F' exists if the distribution function is differentiable. This derivative can be interpreted as the "weight" of the distribution function in the neighborhood of a, because

$$F(a + h) - F(a) \approx F'(a) \cdot h \tag{4.14}$$

applies in linear approximation. The probability of a realization of the random variable in the interval $(a, a + h)$ can be approximated by the product $F'(a) \cdot h$. Remember the probability of exactly realizing a is zero. But if you depart from the point value to a linear estimation of a sufficiently small interval, you obtain—for differentiable distribution functions—a variable that is easy to interpret. F' is called density function.

Let us point out some facts in the context of random variables. From common analysis one knows: adding, subtracting, or multiplying continuous functions result in functions which remain continuous. It is useful to know whether this property holds also for measurable functions (i.e., random variables). The following proposition provides the answer.[5]

Proposition 4.1 (Properties of Random Variables) *If X and Y are random variables, then the sum $X + Y$, the product $X \cdot Y$, and the ratio $\frac{X}{Y}$ (with $Y \neq 0$) are also random variables.*

For the purpose of brevity we omit a proof.

For enhancing the understanding of random variables three additional examples will be presented.

Example 4.7 (Dice Roll) We refer to the dice roll example of page 41 and define the following payout function depending on the score,

$$f(\omega) := \begin{cases} 100, & \text{if } \omega = 1, 3, 5; \\ 200, & \text{if } \omega = 4; \\ 0, & \text{else.} \end{cases}$$

---

[5]We will see later: for Riemann-integrable functions analog relations apply. For example, if X and Y can be integrated, then this holds also for the sum of the functions.

Because of the relationship

$$\{\omega \; : \; f(\omega) < 200\} = \{1, 2, 3, 5, 6\} = \Omega \backslash \{4\},$$

the function f is not $\mathcal{F}_1$-measurable since this set does not belong to the σ-algebra $\mathcal{F}_1$. At time $t = 1$, the function f is therefore not a random variable. Based on the knowledge available at time $t = 1$ it is not possible to decide how high the payout associated with f will be. One learns only at time $t = 2$, whether event $\{4\}$ or event $\{2, 6\}$ has occurred.

Since the σ-algebra $\mathcal{F}_2$ includes all subsets of the possible number of points, the payout function is $\mathcal{F}_2$-measurable. Thus, the function f represents a random variable at $t = 2$. Now you can see if a 4 or other even number or any other number has been rolled.

Example 4.8 Using the same dice roll example we will now consider how a function must be constructed to be a random variable at time $t = 1$. Intuitively, the answer is clear: since one can only distinguish odd and even scores at this point in time, a function is only measurable if it returns identical payouts for all even and all odd scores respectively. Thus, a function of the form

$$f(\omega) = \begin{cases} o, & \text{at odd scores } \omega, \\ e, & \text{at even scores } \omega, \end{cases} \tag{4.15}$$

with $o \neq e$ will ensure that f is measurable at $t = 1$.[6]

This intuition-based statement will now be proven formally for $o > e$. To do so we have to show that the set $M := \{\omega \; : \; f(\omega) < a\}$ is measurable for any real number a. With a given a we can distinguish four conceivable cases.[7]

Case 1: $a < o$ and $a < e$ Since both the even and odd numbers provide function values above a, the set remains empty, $M = \emptyset$. This set can be measured according to the above stated prerequisite.

Case 2: $a < o$ and $a > e$.   With this choice of a, the set M captures the even scores. The odd scores do not belong to M. This set can also be measured.

Case 3: $a > o$ and $a < e$.   Only the odd scores are captured in M, while the even scores do not belong to M. This set is measurable.

Case 4: $a > o$ and $a > e$.   Both even and odd numbers return function values below a. Therefore, all even and all odd numbers are in the set M, which therefore includes all rolls. This set is also measurable.

---

[6]The reader is encouraged to explore other functions capable of guaranteeing measurability.

[7]Depending on the numbers u and g, one of the four cases identified may not even occur. For example, if $g = 0$ and $u = 1$, there is no a to satisfy case 3.

The set M is measurable in all conceivable cases. Therefore, it can be stated that the function f is measurable and thus represents a random variable.

Example 4.9 We consider the Borel-measurable sets on the real line and look for functions $f : \mathbb{R} \rightarrow \mathbb{R}$ that are random variables. According to our definition this implies that the set of

$$A := \{x \ : \ f(x) < a\} \tag{4.16}$$

must be measurable or, as one might say, belongs to the Borel-σ-algebra.

Restricting ourselves to continuous functions f implies: if a point x belongs to the set A, i.e., the function $f(x) < a$ is valid, then there exists also a (possibly small) interval of x in A. Given continuity it follows that $f(x \pm \delta) < a$ also applies. Hence, the set A is an open set and thus Borel-measurable.[8] We can summarize: all continuous functions are random variables; however, the existence of further Borel-measurable functions is not excluded.

4.3 Distribution Functions

Random variables are measurable functions and are also called distribution functions. Describing such functions in full for a specific case can be very time-consuming. In order to get at least a rough idea of a distribution function it is common to characterize it by its moments.

The most important moment is the expectation[9] which can be illustrated by returning to Example 4.2 from page 61. We are interested in the payout a participant could realize on average if this game was played very often (strictly speaking: infinitely often). The amount in question is calculated by weighting the random payouts with their probabilities and adding them over all conceivable states. Hence,

$$E[X] = \sum_{\omega=1}^{6} X(\omega) \cdot \frac{1}{6} = (2 + 4 + \ldots + 12) \cdot \frac{1}{6} = 7. \tag{4.17}$$

The average payment of this game therefore amounts to \$7. The distribution function in our dice example is very straightforward.

Unfortunately, determining expected values is not always easy. Calculating the expected value is far more difficult when dealing with a random variable X, for which $X : \Omega \rightarrow \mathbb{R}$ applies. Since the number of possible realizations is infinitely large and the probability of a specific realization is zero, an integral replaces the sum.

---

[8]See page 47.

[9]Further ratios are addressed in standard textbooks on statistics, for example Mood et al. (1974), chapter 4, or Hogg et al. (2013, p. 52 ff).

With the Brownian motion we are dealing with the state space $\Omega = C[0, \infty)$, i.e., the set of all continuous functions starting at zero. However, anyone wanting to determine the expected value quickly gets into considerable difficulties with the Riemann integral known from high school mathematics. In the following chapter we will show the reason for these difficulties and how they can be overcome.