

# MATH 2510-001: RANDOM VARIABLES

BLITZSTEIN AND HWANG (EDITED BY COLTON)  
SEPTEMBER 29, 2019

This was prepared for University of Colorado’s MATH 2510-001 (Introduction to Statistics) in Fall 2019. Please send questions, comments, complaints, and corrections to [colton.grainger@colorado.edu](mailto:colton.grainger@colorado.edu). Any mistakes are my own. These notes are scraped from section 3.1 on random variables in [BH19] and section 3.4 on the binomial distribution in [DCB15].

## CONTENTS

1. Random variables	1
2. Distributions and probability mass functions	4
3. Binomial distribution	5

1.

## Random variables

We introduce random variables, an incredibly useful concept that simplifies notation and expands our ability to quantify uncertainty and summarize the results of experiments.

To see why our current notation can quickly become unwieldy, consider the gambler’s ruin problem. In this problem, we may be very interested in how much wealth either gambler  $A$  or gambler  $B$  has at any particular time.

So we could make up notation like letting  $A_{jk}$  be the event that gambler  $A$  has exactly  $j$  dollars after  $k$  rounds, and similarly defining an event  $B_{jk}$  for gambler  $B$ , for all  $j$  and  $k$ .

This is already too complicated. And what’s to say that we’re not interested in other quantities, such as the difference of their wealths after  $k$  rounds, or the duration of the game (the number of rounds until one player is bankrupt)?

Expressing the event “the duration of the game is  $r$  rounds” in terms of the  $A_{jk}$  and  $B_{jk}$  would involve a long, awkward string of unions and intersections.

And then, what if we want to express gambler  $A$ ’s wealth as the equivalent amount in euros rather than dollars? We can multiply a *number* in dollars by a currency exchange rate, but we can’t multiply an *event* by an exchange rate.

Instead of having convoluted notation that obscures how the quantities of interest are related, wouldn’t it be nice if we could say something like the following?

Let  $X_k$  be the wealth of gambler  $A$  after  $k$  rounds. Then  $Y_k = N - X_k$  is the wealth of gambler  $B$  after  $k$  rounds (where  $N$  is the fixed total wealth); then

$$X_k - Y_k = 2X_k - N \quad \text{is the difference in wealths after } k \text{ rounds,}$$

and

$$c_k X_k \quad \text{is the wealth of gambler } A \text{ after } k \text{ rounds}$$

where  $c_k$  is the euros per dollar exchange rate after  $k$  rounds; and the duration is

$$R = \min \{n \text{ such that } X_n = 0 \text{ or } Y_n = 0\}.$$

The notion of a random variable will allow us to do exactly this! It needs to be introduced carefully though, to make it both conceptually and technically correct.

Sometimes a definition of “random variable” is given that is a barely paraphrased version of “a random variable is a variable that takes on random values”, but such a feeble attempt at a definition fails to say where the randomness come from.

Nor does it help us to derive properties of random variables: we’re familiar with working with algebraic equations like  $x^2 + y^2 = 1$ , but what are the *valid mathematical operations* if  $x$  and  $y$  are *random* variables?

To make the notion of a random variable precise, we define it as a *function* mapping the sample space to the real line.

**Definition 1.1** (Random variable). Given an experiment with sample space  $\Omega$ , a *random variable* (r.v.) is a function from the sample space  $\Omega$  to the real numbers  $\mathbb{R}$ . It is common, but not required, to denote random variables by capital letters.

Thus, a random variable assigns a numerical value  $X(E)$  to each possible outcome  $E$  of the experiment. The randomness comes from the fact that we have a random experiment (with probabilities determined by the probability measure  $\mathbb{P}$ ); the mapping itself is deterministic, as illustrated in 1.

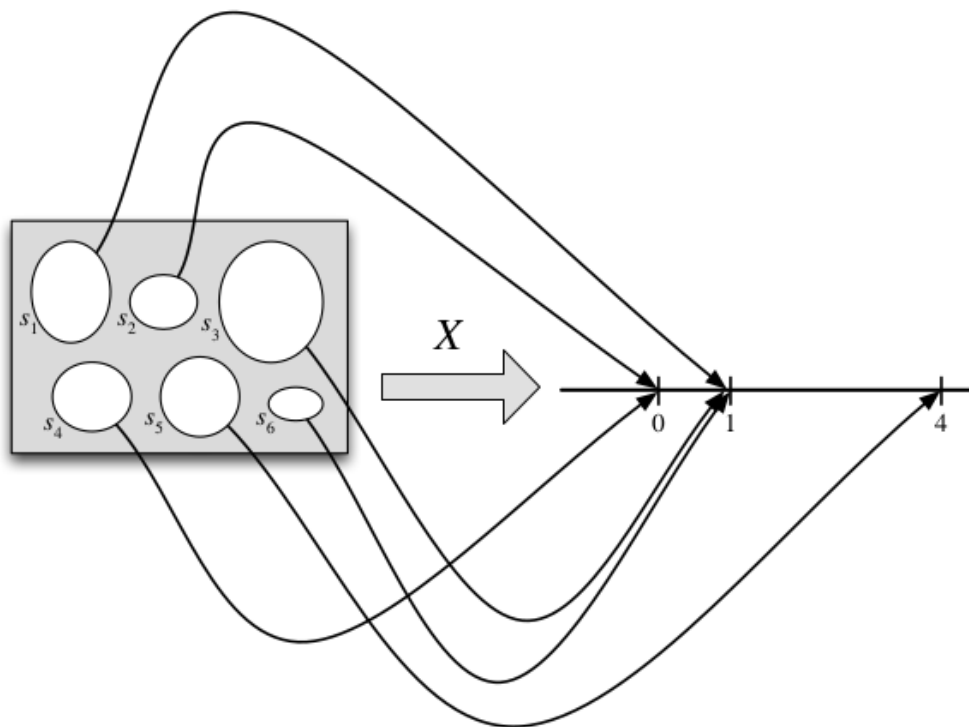


FIGURE 1. A random variable maps the sample space into the real line. The r.v.  $X$  depicted here is defined on a sample space with 6 elements, and has possible values 0, 1, and 4. The randomness comes from choosing a random pebble according to the probability function  $\mathbb{P}$  for the sample space.

A simpler way to “see” the mapping is simply to write the value of  $X(s)$  on each simple event  $s$  in  $\Omega$ .

As we’ve mentioned earlier, the source of the randomness in a random variable is the experiment itself, in which a sample outcome  $E \in \Omega$  is chosen according to a probability function  $\mathbb{P}$ . Before we perform the experiment, the outcome  $E$  has not yet been realized, so we don’t know the value of  $X$ ,

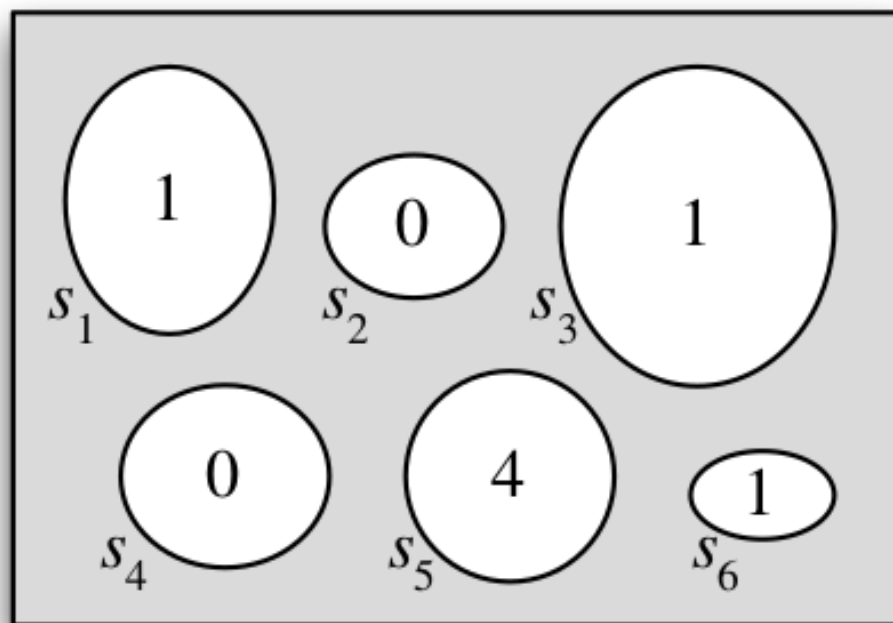


FIGURE 2. The random variable we just defined, with labels instead of arrows.

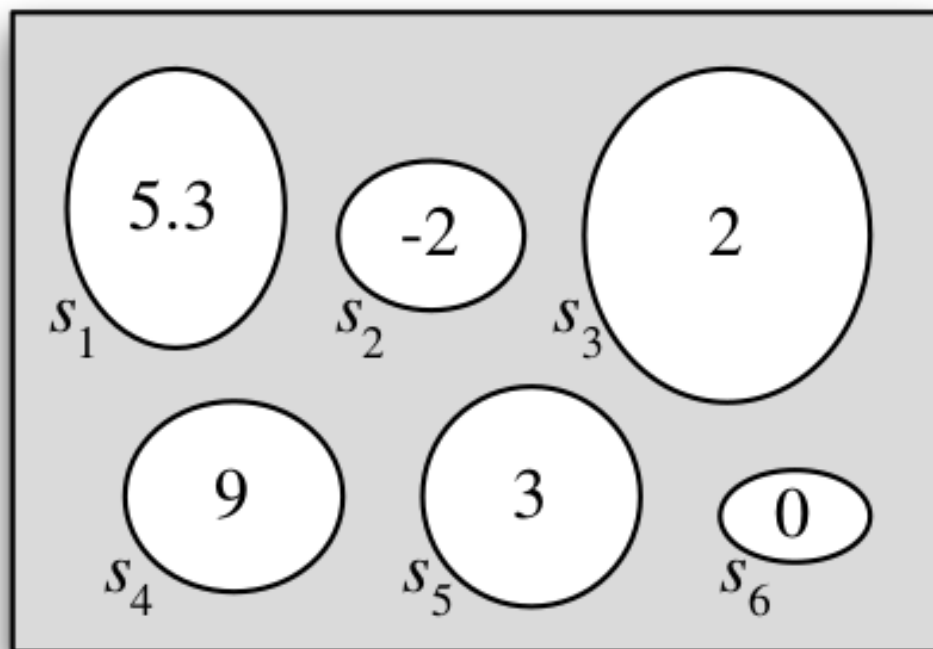


FIGURE 3. A different random variable, defined on the same sample space.

though we could calculate the probability that  $X$  will take on a given value or range of values. After we perform the experiment and the outcome  $s$  has been realized, the random variable crystallizes into the numerical value  $X(E)$ .

Random variables provide numerical summaries of the experiment in question. This is very handy because the sample space of an experiment is often incredibly complicated or high-dimensional, and the outcomes  $E \in \Omega$  may be non-numeric. For example, the experiment may be to collect a random sample of people in a certain city and ask them various questions, which may have numeric (e.g., age or height) or non-numeric (e.g., political party or favorite movie) answers. The fact that r.v.s take on numerical values is a very convenient simplification compared to having to work with the full complexity of  $\Omega$  at all times.

2.

## Distributions and probability mass functions

**Definition 2.1** (Discrete random variable). A random variable  $X$  is said to be discrete if there is a finite list of values  $a_1, a_2, \dots, a_n$  or an infinite list of values  $a_1, a_2, \dots$  such that

$$\mathbb{P}(X = a_j \text{ for some } j) = 1.$$

If  $X$  is a discrete r.v., then the finite or countably infinite set of values such that  $\mathbb{P}(X = x) > 0$  is called the *support* of  $X$ .

Most commonly in applications, the support of a discrete r.v. is a set of integers. In contrast, a continuous r.v. can take on any real value in an interval (possibly even the entire real line); such r.v.s will be defined more precisely on Wednesday.

Given a random variable, we would like to be able to describe its behavior using the language of probability. For example, we might want to answer questions about the probability that the r.v. will fall into a given range: if  $L$  is the lifetime earnings of a randomly chosen U.S. college graduate, what is the probability that  $L$  exceeds a million dollars? If  $M$  is the number of major earthquakes in California in the next five years, what is the probability that  $M$  equals 0?

The *distribution* of a random variable provides the answers to these questions; it specifies the probabilities of all events associated with the r.v., such as the probability of it equaling 3 and the probability of it being at least 110. We will see that there are several equivalent ways to express the distribution of an r.v. For a discrete r.v., the most natural way to do so is with a *probability mass function*, which we now define.

**Definition 2.2** (Probability mass function). The *probability mass function* (PMF) of a discrete r.v.  $X$  is the function  $p_X$  given by

$$p_X(n) = P(X = n) \quad (\text{for } n \text{ usually an integer; go back to the definition of a discrete r.v.}).$$

Note that this is positive if  $n$  is in the *support* of  $X$  and 0 otherwise.

*Note* (Warning!). In writing  $\mathbb{P}(X = n)$  we are using  $X = n$  to denote an *event*, consisting of all outcomes  $E$  to which  $X$  assigns the number  $n$ . This event is also written as  $\{X = n\}$ ; formally,  $\{X = n\}$  is defined as

$$\{E \in \Omega : X(E) = n\}$$

but writing  $\{X = n\}$  is shorter and more intuitive. Thinking of the sample space of two fair coin flips, if  $X$  is the number of **heads** in two fair coin tosses, then  $\{X = 1\}$  consists of the sample outcomes  $HT$  and  $TH$ , which are the two outcomes to which  $X$  assigns the number 1. Since  $\{HT, TH\}$  is a subset of the sample space, it is an event. So it makes sense to talk about  $\mathbb{P}(X = 1)$ , or more generally,  $P(X = n)$ . If  $\{X = n\}$  were anything other than an event, it would make no sense

to calculate its probability! It *doesn't make a lick of sense* to write  $P(X)$ , we can only take the probability of an event, not of an r.v. ◀

**Theorem 2.3** (Valid probability mass functions). *Let  $X$  be a discrete r.v. with support  $x_1, x_2, \dots$  (assume these values are distinct and, for notational simplicity, that the support is countably infinite; the analogous results hold if the support is finite). The PMF  $p_X$  of  $X$  must satisfy the following two criteria:*

**Non-negative:**  $p_X(x) > 0$  if  $x = x_j$  for some  $j$ , and  $p_X(x) = 0$  otherwise;

**Sums to 1:**  $\sum_{j=1}^{\infty} p_X(x_j) = p_X(x_1) + p_X(x_2) + \dots = 1$ .

The PMF is one way of expressing the distribution of a discrete r.v. This is because once we know the PMF of  $X$ , we can calculate the probability that  $X$  will fall into a given subset of the real numbers by summing over the appropriate values of  $x$ , as the next example shows.

**Example 2.4** (Sum of two fair die rolls). Let  $T$  be the sum of two fair die rolls.

- (1) Calculate the PMF of  $T$ .
- (2) Now suppose we're interested in the probability that  $T$  is in the interval  $[1, 4]$ . There are only three values in the interval  $[1, 4]$  that  $T$  can take on, namely, 2, 3, and 4. We know the probability of each of these values from the PMF of  $T$ .
- (3) Compute  $\mathbb{P}(1 \leq T \leq 4) = \mathbb{P}(T = 2) + \mathbb{P}(T = 3) + \mathbb{P}(T = 4)$ .
- (4) It's  $6/36$ , or  $1/6$ .

In general, given a discrete r.v.  $X$  and a set  $B$  of real numbers, if we know the PMF of  $X$  we can find  $\mathbb{P}(\{X \in B\})$ , the probability that  $X$  is in  $B$ , by summing up the heights of the vertical bars at points in  $B$  in the plot of the PMF of  $X$ .

*Knowing the PMF of a discrete r.v. determines its distribution.*

3.

## Binomial distribution

Some distributions are so ubiquitous in probability and statistics that they have their own names. We will introduce these named distributions throughout the book, starting with a very simple but useful case: an r.v. that can take on only two possible values, 0 and 1.

**Definition 3.1** (Bernoulli distribution). A r.v.  $X$  is said to have the *Bernoulli distribution* with parameter  $p$  if  $\mathbb{P}(X = 1) = p$  and  $\mathbb{P}(X = 0) = 1 - p$ , where  $0 < p < 1$ . We write this as  $X \sim \text{Bern}(p)$ . The symbol  $\sim$  is read “is distributed as”.

*Note* (Families of distributions). This number  $p$  in  $\text{Bern}(p)$  is called the parameter of the distribution; it determines which specific Bernoulli distribution we have. Thus there is not just one Bernoulli distribution, but rather a family of Bernoulli distributions, indexed by  $p$ . For example, if  $X \sim \text{Bern}(1/3)$ , it would be correct but incomplete to say “ $X$  is Bernoulli”; to fully specify the distribution of  $X$ , we should both say its name (Bernoulli) and its parameter value ( $1/3$ ), which is the point of the notation  $X \sim \text{Bern}(1/3)$ . ◀

**Definition 3.2** (Bernoulli trial). An experiment that can result in either a success or a failure (but not both) is called a *Bernoulli trial*. A Bernoulli random variable can be thought of as the *indicator of success* in a Bernoulli trial: it equals 1 if success occurs and 0 if failure occurs in the trial.

Because of this story, the parameter  $p$  is often called the success probability of the  $\text{Bern}(p)$  distribution. Once we start thinking about Bernoulli trials, it's hard not to start thinking about what happens when we have more than one trial.

**Definition 3.3** (Binomial distribution). Suppose that  $n$  independent Bernoulli trials are performed, each with the same success probability  $p$ . Let  $X$  be the number of successes. The distribution of  $X$  is called the *Binomial distribution* with parameters  $n$  and  $p$ . We write  $X \sim \text{Bin}(n, p)$  to mean that  $X$  has the Binomial distribution with parameters  $n$  and  $p$ , where  $n$  is a positive integer and  $0 < p < 1$ .

Notice that we define the Binomial distribution not by its PMF, but by a story about the type of experiment that could give rise to a random variable with a Binomial distribution. The most famous distributions in statistics all have stories which explain why they are so often used as models for data, or as the building blocks for more complicated distributions.

Thinking about the named distributions first and foremost in terms of their stories has many benefits. It facilitates pattern recognition, allowing us to see when two problems are essentially identical in structure; it often leads to cleaner solutions that avoid PMF calculations altogether; and it helps us understand how the named distributions are connected to one another. Here it is clear that  $\text{Bern}(p)$  is the same distribution as  $\text{Bin}(1, p)$ : the Bernoulli is a special case of the Binomial.

Using the story definition of the Binomial, let's find its PMF.

**Theorem 3.4** (Derivation of the Binomial PMF). *The binomial distribution describes the probability of having exactly  $k$  successes in  $n$  independent Bernoulli trials with probability of a success  $p$ . If  $X \sim \text{Bin}(n, p)$ , then the PMF of  $X$  is*

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

for  $k = 0, 1, \dots, n$  and  $\mathbb{P}(X = k) = 0$  otherwise when  $k > n$ .

*Proof.* Let an unfair coin have probability  $p$  of landing **heads**. Clearly the probability of **tails** is  $1 - p$ . Now consider  $\mathbb{P}(\text{heads})$  under the general case of  $k$  **heads** and  $n - k$  **tails** across the  $n$  Bernoulli trials. In any such scenario, we apply the Multiplication Rule for independent events:

$$p^k (1 - p)^{n-k}.$$

This is our general probability for  $\mathbb{P}(\text{heads})$ , which is just the probability of a single event. Secondly, we introduce a general formula for the number of ways to choose  $k$  successes in  $n$  trials, i.e. arrange  $k$  successes and  $n - k$  failures:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

The quantity  $\binom{n}{k}$  is read  *$n$  choose  $k$* . The exclamation point notation (e.g.  $k!$ ) denotes a factorial expression.

$$0! = 1$$

$$1! = 1$$

$$2! = 2 \times 1 = 2$$

$$3! = 3 \times 2 \times 1 = 6$$

$$4! = 4 \times 3 \times 2 \times 1 = 24$$

$$n! = n \times (n - 1) \times \cdots \times 3 \times 2 \times 1$$

For example, using the formula, we can compute the number of ways to choose  $k = 3$  successes in  $n = 4$  trials:

$$\binom{4}{3} = \frac{4!}{3!(4-3)!} = \frac{4!}{3!1!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{(3 \cdot 2 \cdot 1)(1)} = 4$$

Substituting  $n$  choose  $k$  for the number of arrangements of  $k$  **heads** in  $n$  trials, and also substituting  $p^k (1 - p)^{n-k}$  for the probability of the single event **heads**, yields the general binomial formula.  $\square$

What does the binomial PMF look like?

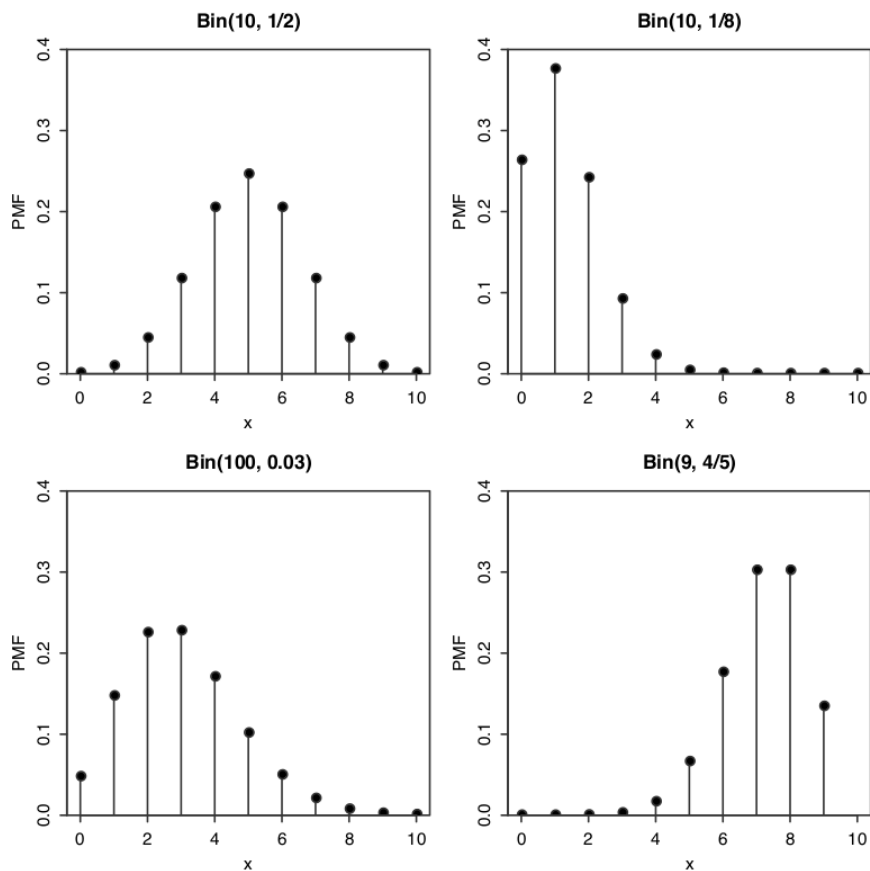


FIGURE 4. Some Binomial PMFs. In the lower left, we plot the  $\text{Bin}(100, 0.03)$  PMF between 0 and 10 only, as the probability of more than 10 successes is pretty darn close to 0.

*Note* (Is  $X$  distributed binomially?). Is your random variable binomially distributed? With parameters  $p, k, n$ ? You have four conditions to check.

- (1) The trials are independent.
- (2) The number of trials,  $n$ , is fixed.
- (3) Each trial outcome can be classified as a success or failure.
- (4) The probability of a success,  $p$ , is the same for each trial.

◀

**Corollary 3.5** (The binomial distribution begins to approximate the normal distribution). *Let  $X \sim \text{Bin}(n, p)$  with  $p = 1/2$  and  $n$  even. Then the distribution of  $X$  is symmetric about  $n/2$ , in the sense that  $\mathbb{P}(X = n/2 + j) = \mathbb{P}(X = n/2 - j)$  for all nonnegative integers  $j$ .*

## References

- [BH19] Joseph Blitzstein and Jessica Hwang. *Introduction to Probability*. CRC Press, 2nd edition, 2019. 1
- [DCB15] David Diez, Mine Cetinkaya-Rundel, and Christopher Barr. *OpenIntro Statistics*. Leanpub, July 2015. 1

## Index of Terms and Keywords

$n$  independent, 6

Bernoulli distribution, 5

Bernoulli trial, 5

Binomial distribution, 6

distribution, 4

event, 1, 4

function, 2

indicator of success, 5

must, 5

number, 1

probability mass function, 4

random, 2

random variable, 2

support, 4

valid mathematical operations, 2