Random Variables, pt. I

ECON 3640–001

Marcio Santetti

Spring 2022

Motivation

Housekeeping

Notes based on `Blitzstein & Hwang (2019)`, ch. 3

- sections 3.1—3.3.

The road so far

Let us quickly **recap** what we have seen so far:

- **Descriptive Statistics**: *visual* techniques; *uni* and *bivariate* descriptive methods;
- **Probability Theory**:
  - Probabilities as measures of *uncertainty*; *frequentist* vs. *Bayesian* interpretations;
  - *Joint*, *marginal*, and *conditional* probability;
  - Law of Total Probability (LTP) and Bayes' Theorem.

Now, we are still within the "*Probability realm*," but we must learn new concepts to study the **behavior** of a **random variable**.

Random variables

Random variables

> Given an experiment in the sample space S, a **random variable** X is a function from S to the real number line $\mathbb{R}$

Translation: a random variable X is a **numerical** outcome of an experiment in the sample space S.

There are **two** types of RVs:

1. **Discrete**: the list of possible outcomes of X is *finite* or *countably infinite*.
2. **Continuous**: list of possible outcomes of X is *uncountably infinite* (such as an interval).

Discrete random variables

Discrete random variables

Let us start with an example:

Assume we flip a *fair* coin twice.

What is the **sample space** for this experiment's **outcomes**?

- $S = \{HH, TH, HT, TT\}$

How do we *frame* this situation in the context of random variables?

- Let X be the number of **Head** outcomes.
- How many values can it take on?

X, the number of **Heads**, can be 0, 1, or 2.

Discrete random variables

But where does the **random** part come from?

- The randomness comes from the experiment itself!
- An outcome $s \in S$ is chosen according to a **probability function** P.
- Before the experiment is performed, we **do not know** the result of the coin flipping, but we can calculate the **probability** associated with each possible value of this random variable!

Thus, what are the probabilities associated with each value for X?

- $P(X = 2) = 1/4$;
- $P(X = 1) = 2/4 = 1/2$;
- $P(X = 0) = 1/4$;

Probability distributions

Probability distributions

The **distribution** of a random variable specifies the probabilities of all events associated with it.

For a **discrete** RV, its distribution is called a **probability mass function (PMF)**.

> A probability mass function *P(X)* describes the plausible values of a discrete RV *X*.

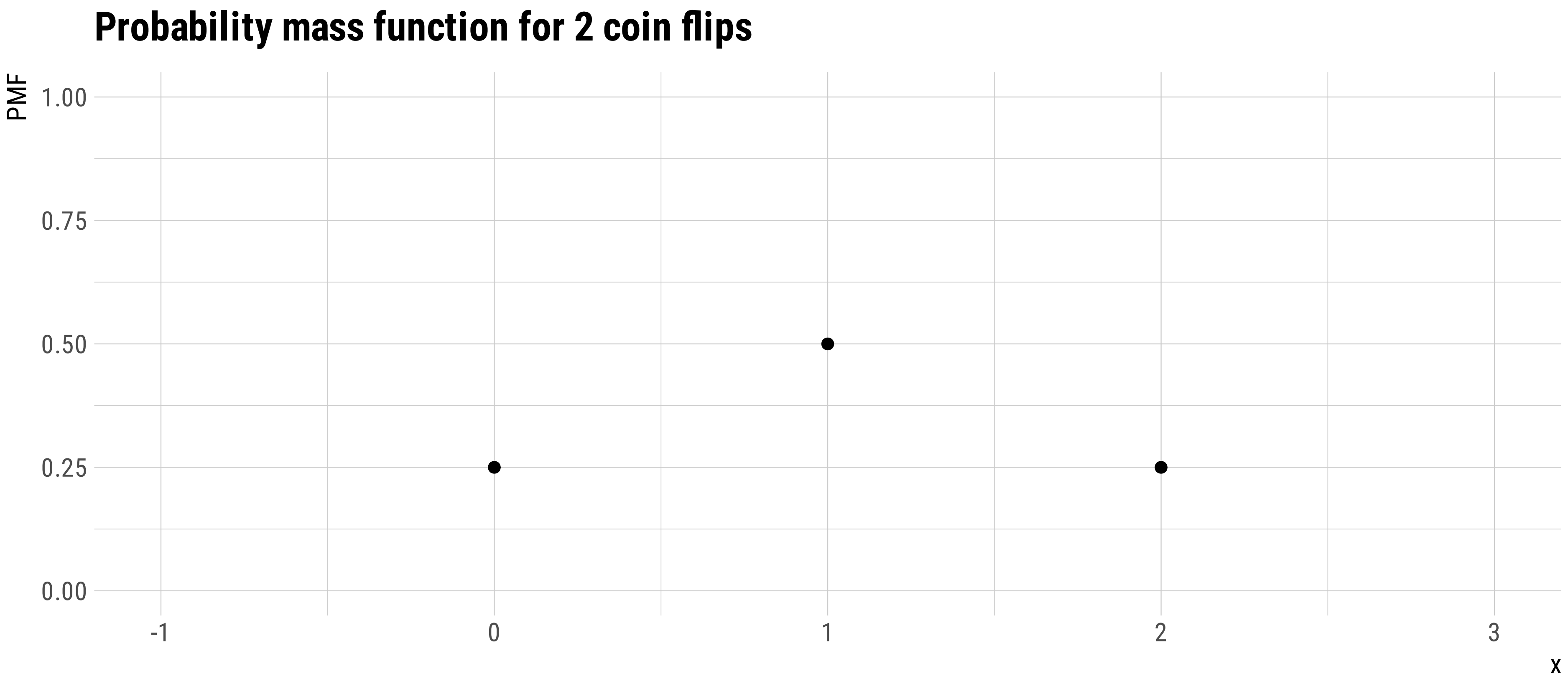
It must satisfy:

- $0 \leq P(X) \leq 1$;
- $\sum_{all\ x} P(X) = 1$

The statement $P(X = x)$ is read as the "probability with which the random variable *X* takes on the numerical value *x*."

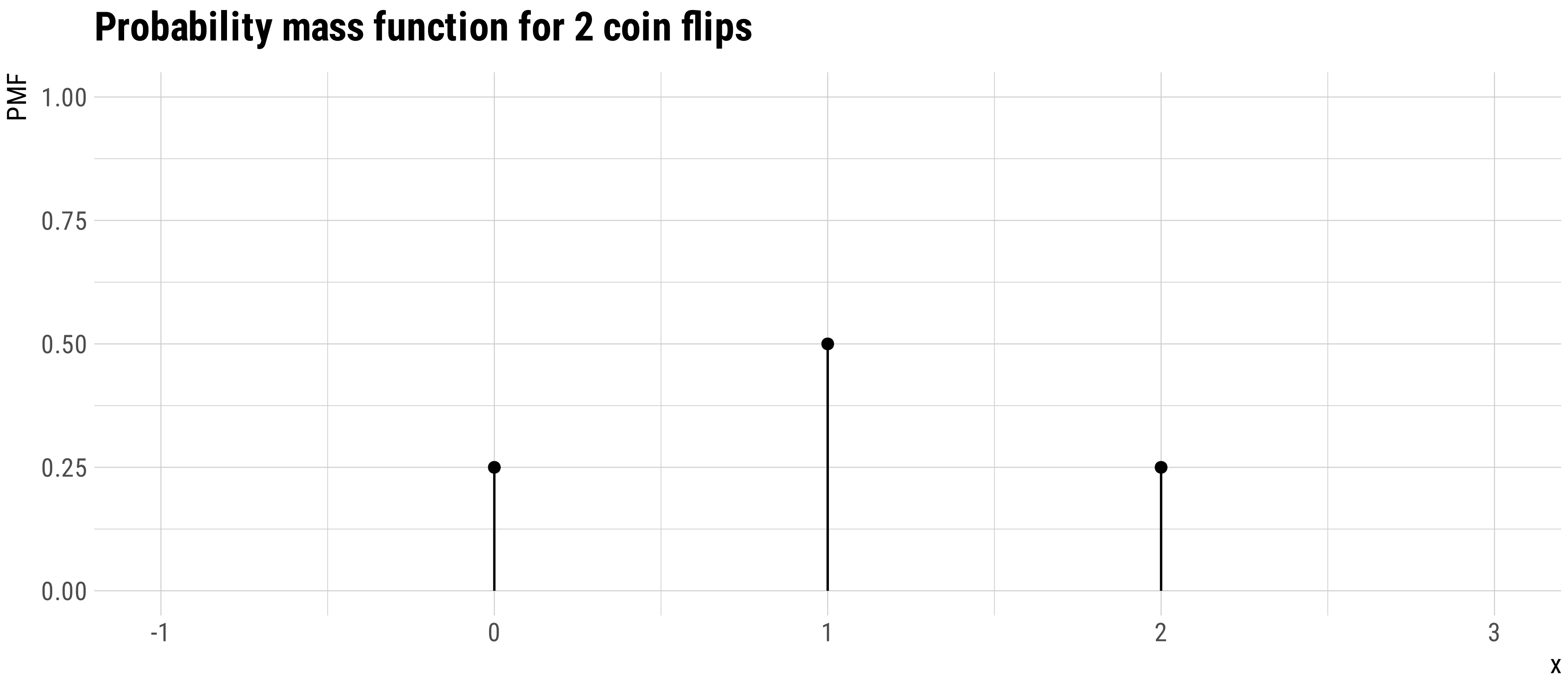
Probability distributions

Let us **visually** look at the previous results:



Probability distributions

Let us **visually** look at the previous results:



The Binomial distribution

The Binomial distribution

The **simplest** discrete probability distribution to start with is the **Binomial distribution**.

Suppose an RV that can take on only **two** possible values: 0 and 1.

There are **four** properties a binomial experiment must fulfill:

1. It consists of a *fixed* number of *trials* (n);
2. Each trial has two possible outcomes: *success* or *failure*;
3. The probability of success is denoted by p, while that of failure is $1 - p$;
4. Trials are *independent*. In other words, the outcome of one trial does not affect other outcomes.

If properties 2, 3, and 4 are satisfied, we have what is called a ***Bernoulli*** process. By adding the first property, a **binomial** experiment is defined.

The Binomial distribution

A **binomial random variable** is the number of *successes* in the experiment's n trials.

It can take on values 0, 1, 2,..., n.

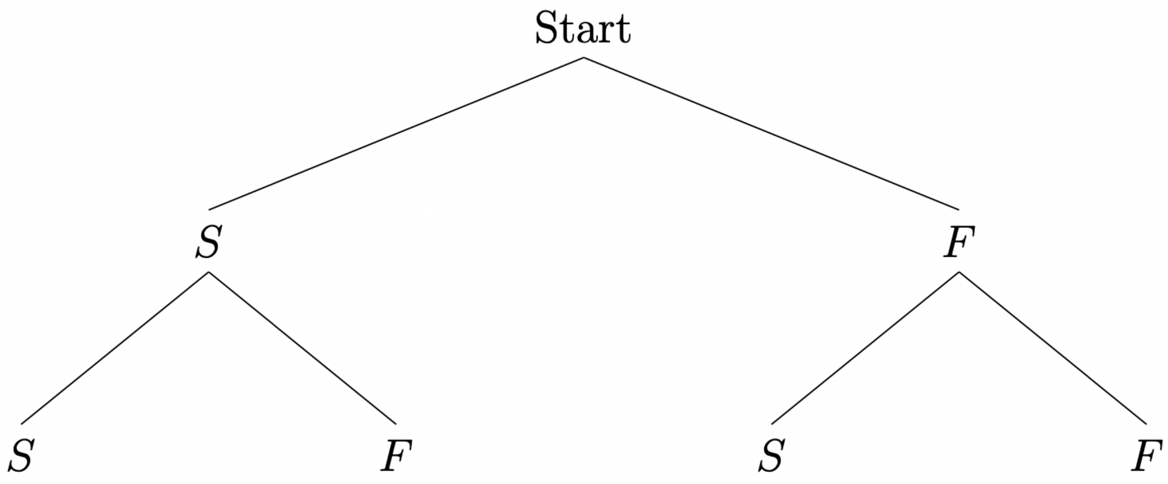
- That is, **discrete** values.

If an experiment has n trials, we can have X successes

- X is our binomial random variable!
- X occurs with probability p, and
- $n - X$ failures occur with probability $1 - p$.

The Binomial distribution

Using a *probability tree* to illustrate a binomial experiment (with $n = 2$ trials):



For **any** number of trials, to count the number of branch sequences that produce x successes and $n - x$ failures, we use:

$$C_x^n = \frac{n!}{x!(n-x)!}$$

The Binomial distribution

In addition to this *counting* relation, the *probability* for each sequence of branches that represent x successes and $n - x$ failures is represented by:

$$p^x(1-p)^{n-x}$$

Combining these **two** components, we have the **probability function** for a Binomial distribution:

$$P(X=x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}\ , \text{ for } x = 0, 1, 2, 3, \ldots, n$$

The Binomial distribution

An example:

A quiz consists of **10** multiple choice questions. Each question has 5 possible answers, only one being correct. A student plans to ***guess*** the answer to each question.

Calculate the probability of

- *(a)* no correct answers, and
- *(b)* of two correct answers.

Cumulative probabilities

Cumulative probabilities

In case we wish to find the probability that a random variable is **less than or equal to** a value, we are looking for a **cumulative probability function**.

- Such a probability is denoted by $P(X \leq x)$.

Using the previous problem, find the probability that the student **fails** the quiz.

- Failure occurs when the mark is below 50%.

The binomial distribution

We can also calculate descriptive measures of **central location** and **variability** for specific distributions.

In the case of a **binomial distribution**, its mean, variance, and standard deviation measures are:

- **Mean**: $\mu = np$
- **Variance**: $\sigma^2 = np(1-p)$
- **Standard deviation**: $\sigma = \sqrt{np(1-p)}$

As an example, assume that the entire class from the previous example is as prepared as the aforementioned student for the quiz. That is, *everyone* plans to *guess* all answers.

- What will be the average mark?
- What will be the standard deviation?

The Poisson distribution

The Poisson distribution

The *Poisson* distribution is **similar** to the Binomial.

A Poisson RV is the number of occurrences of **success** events in an *interval of time* or *specific region of space*.

Some examples of Poisson random variables:

- number of cars arriving at a specific service station in a 1-hour interval;
- number of flaws within a certain portion of a product's assembly line;
- number of accidents registered in one day on a particular stretch of highway.

The Poisson distribution

The Poisson distribution also has **four properties** a discrete random variable must fulfill to be considered a Poisson RV:

1. The number of successes (# successes) in any interval is independent of the number of successes in any other interval;
2. The probability of success, *P(success)*, in an interval is the same for all equally-sized intervals;
3. *P(success)* in an interval is proportional to the size of the interval;
4. The probability of more than one success in an interval approaches zero as the interval becomes smaller.

The Poisson distribution

The probability with which a Poisson random variable assumes the value of *x* successes in an interval is given by:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \text{ , for } x = 1, 2, 3, \ldots$$

- where λ is the average occurrence of *successes* for our Poisson random variable;
- e is the base of the natural logarithm (approximately 2.71828...)
- *x* is the number of successes we are interested in.

Finally, the **mean** and **variance** of a Poisson random variable because they are the same for the Poisson distribution.

- Therefore, $\lambda = \sigma^2$.

The Poisson distribution

A quick example:

Assume that the number of *typos* in a textbook is Poisson distributed, with an average of 1.5 typos per 100 pages.

- What is the probability of *no typos* if we randomly select 100 pages of this textbook?

Next time: Continuous random variables