

Section 3: Random variables – discrete

STA 35C – Statistical Data Science III

Instructor: Akira Horiguchi

Fall Quarter 2025 (Sep 24 – Dec 12)
MWF, 12:10 PM – 1:00 PM, Olson 158
University of California, Davis

Basic concepts

Based on Chapter 3 of textbook: <https://www.probabilitycourse.com/>

- 1 Basic concepts
- 2 Independent random variables
- 3 Special distributions

We usually focus on some *numerical aspects* of a random experiment.

- For example, in a soccer game we may be interested in the number of goals, shots, shots on goal, corners kicks, fouls, etc.
- On any given day at UCD, we may be interested in the number of Cheeto sightings.
- These are examples of *random variables*.

Definition 1: Random variable

A **random variable** $X: \Omega \rightarrow \mathbb{R}$ is a function from the sample space Ω to the real numbers.

- E.g., toss a coin three times. Sample space is

$$\Omega = \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}.$$

obs H : 0 1 1 2 1 2 2 3

We can define a random variable X whose value is the number of observed heads.

- Usually denote random variables by capital letters such as X , Y , and Z .
- The **range** of a random variable X is the set of possible values for X . For example:
 - ▶ I toss a coin 100 times. Let X be the number of heads I observe. **range = $\{0, 1, 2, \dots, 100\}$**
 - ▶ I toss a coin until the first heads appears. Let Y be the total number of coin tosses. **N**
 - ▶ The random variable T is defined as the time (~~in hours~~) from now until the next earthquake occurs in a certain city. **range = $(0, \infty)$**

Discrete random variables

A random variable X is *discrete* if its range is countable.

■ Recurring examples:

1. number of heads after two coin flips,
2. number of coin flips needed before a heads turns up.

■ Here probabilities can be assigned to each realizable value.

1. For $\{0, 1, 2\}$ (finite), we can assign probabilities $1/4$, $1/2$, and $1/4$.
2. For \mathbb{N} (countably infinite), we can assign probabilities $(1/2)^k$ to each $k \in \mathbb{N}$.

■ For a discrete r.v. X with range $\{x_1, x_2, x_3, \dots\}$, the function $f_X(\cdot)$ defined as

$$f_X(x_k) = P(X = x_k), \quad \text{for } k = 1, 2, 3, \dots,$$

is called the *probability mass function (PMF)* of X .

1. $f_X(0) = 1/4$, $f_X(1) = 1/2$, and $f_X(2) = 1/4$.
2. $f_X(k) = (1/2)^k$ for each $k \in \mathbb{N}$.

Here $f_X(a)$ is "*the probability that X equals a .*"

Discrete random variables: PMF

PMF of the number of heads after two flips of a fair coin.

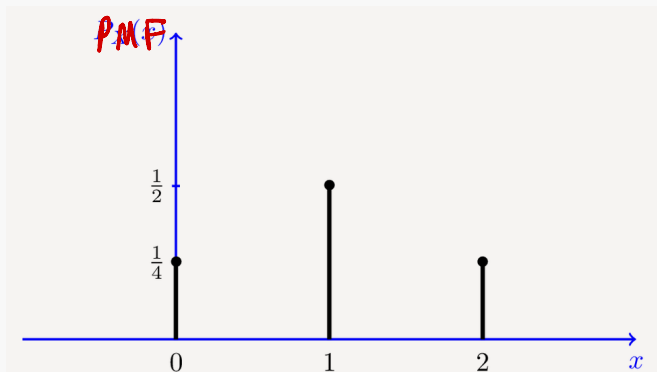


Fig.3.1 - PMF for random Variable X in Example 3.3.

The PMF of a discrete random variable is also called the r.v.'s *probability distribution*.

Discrete random variables: PMF

A PMF is a probability measure, so it satisfies Definition 1 from Section 2.

- In particular, it satisfies countable additivity.
- This lets us deduce the probability $P(X \in A)$ that a discrete r.v. X lies in an event A :

$$\underline{P(X \in A)} = P\left(\bigcup_{a \in A} [X = a]\right) = \sum_{a \in A} \frac{f_X(a)}{P(X=a)}, \quad (1)$$

$$[X \in A] \Leftrightarrow \bigcup_{a \in A} [X = a]$$

$a \in A \Leftrightarrow$ outcome a in event A

End of 10/3
lecture

Independent random variables

Independent random variables

When dealing with more than one random variable, often need to consider the *dependence/correlation* between them.

- Concept of *independent random variables* is similar to that of independent events.
- Two random variables are independent if knowing the value of one does not change the probabilities for the other.

Two independent random variables

Definition 2: Two independent random variables

Two discrete random variables X and Y are *independent* if

$$P(X = x, Y = y) = P(X = x) P(Y = y) \quad (2)$$

for all x, y .

If two random variables are independent, then we can write

$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B)$$

for all sets A, B . We can also write

$$P(Y = y|X = x) = P(Y = y)$$

for all x, y .

Example

Toss a fair coin four times.

- Let X be the number of heads observed in the first two coin flips.
- Let Y be the number of heads observed in the third and fourth coin flips.

Find $P((X < 2) \text{ and } (Y > 1))$.

≥ 2 independent random variables

Definition 3: ≥ 2 independent random variables

Discrete random variables $X_1, X_2, X_3, \dots, X_n$ are *independent* if

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_1 = x_1) P(X_2 = x_2) \dots P(X_n = x_n) \quad (3)$$

for all x_1, x_2, \dots, x_n .

Special distributions

A random variable X with values in a finite set M is *uniformly* distributed if each element in M has the same probability:

$$P(X = k) = \frac{1}{|M|} \quad \text{for all } k \in M$$

- Such distributions occur when all possible outcomes are equally likely.
- We write $X \sim U(M)$ or $X \sim \text{Unif}(M)$.
- Nine random draws in R: `sample(c(1,2,3,4,5,6), size=9, replace=T)`

A random variable X is *Bernoulli* distributed with parameter $p \in (0, 1)$, if $P(X = 1) = p$ and $P(X = 0) = 1 - p$.

- For when our random experiment has only two possible outcomes ("success" and "failure").
- Example: flip a coin with probability p of heads ("success"). Is it heads?
- We write $X \sim \text{Ber}_p$ or $X \sim \text{Bern}(p)$.
- Nine random draws in R: `rbinom(n=9, size=1, prob=1/3)`

Binomial distribution

A random variable X is **Binomial** distributed with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$ if

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{for all } k = 0, \dots, n.$$

- We think of n as the number of experiments and p the success probability. In the above equation, k is the number of successes.
- For measuring the probability of the number of successes of n independent Bernoulli experiments with parameter p .
- Example: flip a coin n times, each flip with probability p of heads ("success"). How many heads?
- We write $X \sim \text{Bin}_{n,p}$ or $X \sim \text{Bin}(n, p)$.
- A random draw in R: `rbinom(n=3, size=1, prob=0.25) |> sum()`