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

Section 1: Overview

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*.

Random variables

Definition 1: Random variable

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

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

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

  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.
  - I toss a coin until the first heads appears. Let Y be the total number of coin tosses.
  - The random variable T is defined as the time (in hours) from now until the next earthquake occurs in a certain city.

Discrete random variables

Countable vs uncountable sets. A random variable X is *discrete* if its range is countable.

- 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. Examples:
  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 random variable X with range $\{x_1, x_2, x_3, \cdots\}$, the function $f_X(\cdot)$ defined as

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

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

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

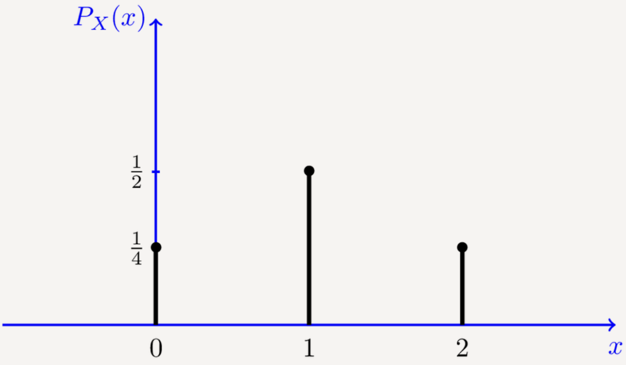

Fig.3.1 - PMF for random Variable X in Example 3.3.

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

Discrete random variables: PMF

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

- In particular, it satisfies countable additivity.
- So the probability $P(X \in A)$ that a discrete random variable X lies in an event A can be calculated by

$$P(X \in A) = \sum_{a \in A} f_X(a), \tag{1}$$

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

$$P(X = x, Y = y) = P(X = x)\,P(Y = y) \tag{2}$$

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

Definition 3: ≥ 2 independent random variables

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

$$P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = P(X_1 = x_1)\, P(X_2 = x_2) \cdots P(X_n = x_n) \tag{3}$$

for all $x_1, x_2, \cdots, x_n$.

Special distributions

Uniform distribution

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 $X \sim Unif(M)$.
- Nine random draws in R: `sample(c(1,2,3,4,5,6), size=9, replace=T)`

Bernoulli distribution

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 $X \sim Ber_p$ or $X \sim Bern(p)$.
- 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

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

- 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 $X \sim Bin_{n,p}$ or $X \sim Bin(n, p)$.
- A random draw in R: `rbinom(n=3, size=1, prob=0.25) |> sum()`