3: Random Variables

Fall 2023

Matthew Blackwell

Gov 2002 (Harvard)

Where are we? Where are we going?

- Up to now: probability of abstract events, but data is numeric!
- Connection between probability and data: **random variables**.
- Long-term goal: inferring the data generating process of this variable.
    - What is the true Biden approval rate in the US?
- Today: given a probability distribution, what data is likely?
    - If we knew the true Biden approval, what samples are likely?

Roadmap

1. Random variables
2. Famous distributions
3. Cumulative distribution functions
4. Functions of random variables
5. Independent random variables

1/ Random variables

What are random variables?

Definition

A **random variable (r.v.)** is a function that maps from the sample space of an experiment to the real line or $X : \Omega \to \mathbb{R}$.

- Numeric representation of uncertain events $\rightsquigarrow$ we can use math!
- The r.v. is X and the numerical value for some outcome ω is $X(\omega)$.
- Randomness comes from the randomness of the experiment.

Example: sampling senators

- For any experiment, there can be many random variables.
- Randomly sample 2 senators $\rightsquigarrow$ 4 outcomes: $\Omega = \{DD, RD, DR, RR\}$.
  - $X =$ number of Democrats in the two draws.
  - $X(DD) = 2$, $X(RD) = X(DR) = 1$, $X(RR) = 0$
  - Another r.v. $Y =$ number of Republicans in the two draws, $Y = 2 - X$
  - $Z = 1$ if draw is two Democrats (DD), 0 otherwise.
- Usually abstract away from the underlying sample space fairly quickly.

Types of r.v.s

- Two main types of r.v.s: discrete and continuous. Focus on discrete now.

Definition

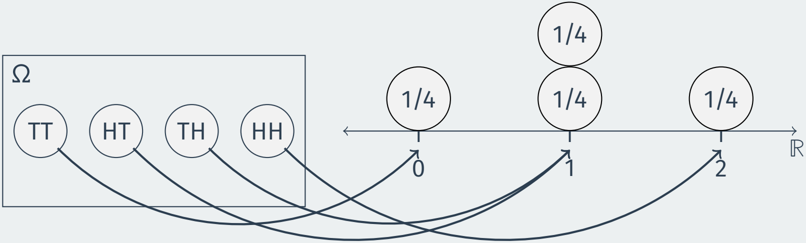
A r.v. X is **discrete** the values it takes with positive probability is finite ($X \in \{x_1, \dots, x_k\}$) or countably infinite ($X \in \{x_1, x_2, \dots\}$).

- The **support** of X is the values x such that $\mathbb{P}(X = x) > 0$.

The random in random variable

- How are r.v.s **random**?
  - Uncertainty over $\Omega \rightsquigarrow$ uncertainty over value of X.
  - We'll use probability to formalize this uncertainty.
- The **distribution** of a r.v. describes its behavior in terms of probability.
  - Specifies probabilities of all possible events of the r.v.
  - $X =$ number of times a randomly chosen citizen contributed to a campaign in 2020.
  - What's the $\mathbb{P}(X > 5)$? $\mathbb{P}(X = 0)$?
- Often there are many ways to express a distribution.

Inducing probabilities



- Let X be the number of heads in two coin flips.

ω	$\mathbb{P}(\{\omega\})$	$X(\omega)$
TT	1/4	0
HT	1/4	1
TH	1/4	1
HH	1/4	2

x	$\mathbb{P}(X = x)$
0	1/4
1	1/2
2	1/4

Expressing a distribution

- **Probability mass function (p.m.f.)**: $p_X(x) = \mathbb{P}(X = x)$
    - **Careful**: $\mathbb{P}(X = x)$ makes sense b/c $\{X = x\}$ is an event.
    - $\mathbb{P}(X)$ doesn't make any sense since X is just a mapping.
- Some properties of valid p.m.f. of a discrete r.v. X with support $x_1, x_2, \ldots$:
    - Nonnegative: $p_X(x) > 0$ if $x \in x_1, x_2, \ldots$ and $p_X(x) = 0$ otherwise.
    - Sums to 1: $\sum_{j=1}^{\infty} p_X(x_j) = 1$.
- Probability of a set of values $S \subset \{x_1, x_2, \ldots\}$:

$$\mathbb{P}(X \in S) = \sum_{x \in S} p_X(x)$$

Example - random assignment to treatment

- You want to run a randomized control trial on 3 people.
- Use the following procedure:
  - Flip independent fair coins for each unit
  - Heads assigned to Control (C), tails to Treatment (T)
- Let X be the number of treated units:

$$X = \begin{cases} 0 & \text{if } (C, C, C) \\ 1 & \text{if } (T, C, C) \text{ or } (C, T, C) \text{ or } (C, C, T) \\ 2 & \text{if } (T, T, C) \text{ or } (C, T, T) \text{ or } (T, C, T) \\ 3 & \text{if } (T, T, T) \end{cases}$$

- Use independence and fair coins:

$$\mathbb{P}(C, T, C) = \mathbb{P}(C)\mathbb{P}(T)\mathbb{P}(C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

Calculating the p.m.f.

$$p_X(0) = \mathbb{P}(X = 0) = \mathbb{P}(C, C, C) = \frac{1}{8}$$

$$p_X(1) = \mathbb{P}(X = 1) = \mathbb{P}(T, C, C) + \mathbb{P}(C, T, C) + \mathbb{P}(C, C, T) = \frac{3}{8}$$

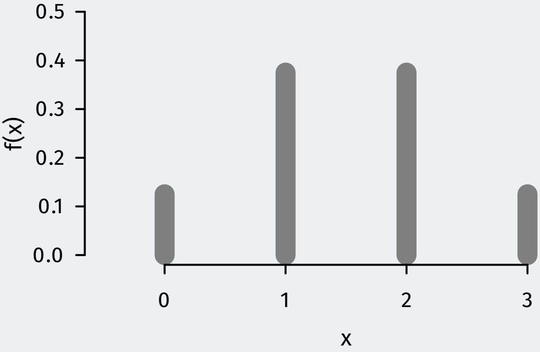
$$p_X(2) = \mathbb{P}(X = 2) = \mathbb{P}(T, T, C) + \mathbb{P}(C, T, T) + \mathbb{P}(T, C, T) = \frac{3}{8}$$

$$p_X(3) = \mathbb{P}(X = 3) = \mathbb{P}(T, T, T) = \frac{1}{8}$$

- What's $\mathbb{P}(X = 4)$? 0!

Plotting the p.m.f.

- We could plot this p.m.f. using R:



- **Question**: Does this seem like a good way to assign treatment? What is one major problem with it?

2/ Famous distributions

Bernoulli distribution

Definition

An r.v. X has a **Bernoulli distribution** with parameter p if $\mathbb{P}(X = 1) = p$ and $P(X = 0) = 1 - p$ and this is written as $X \sim \text{Bern}(p)$.

- Story: indicator of success in some trial with either success or failure.
- Actually a **family** of distributions indexed by p.
- Any event A has an associated Bernoulli r.v.: **indicator variable**:

$$\mathbb{I}(A) \sim \text{Bern}(p) \text{ with } p = \mathbb{P}(A)$$

Binomial distribution

Definition

Let X be the number of successes in n independent Bernoulli trials all with success probability p. Then X follows the **binomial distribution** with parameters n and p, which is written $X \sim \text{Bin}(n, p)$.

- Definition is based on a **story**: helps pattern match to our data.
- Also helps draw immediate connections:
  - $\text{Bin}(1, p) \sim$ Bern(p).
  - If $X \sim \text{Bin}(n, p)$, then $n - X \sim \text{Bin}(n, 1 - p)$.

Binomial p.m.f.

Binomial p.m.f.

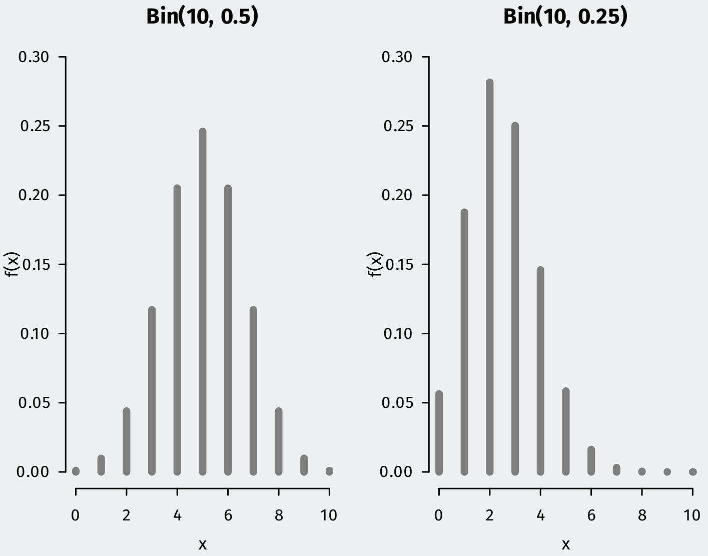
If $X \sim \text{Bin}(n, p)$, then the p.m.f. of X is

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k},$$

for all $k = 0, 1, \dots, n$.

- $p^k(1-p)^{n-k}$ is the probability of a **specific** sequence of 1's and 0's with k 1's.
- Binomial coefficient $\binom{n}{k}$ is how many of these combinations there are.

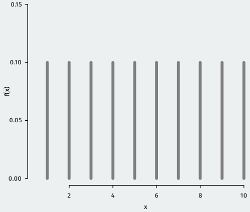
Some binomials



Discrete uniform distribution

Definition

Let C be a finite, nonempty set of numbers. If X is the number chosen randomly with all values equally likely, we say it follows the **discrete uniform** distribution.



- p.m.f. for a discrete uniform r.v.:

$$p_X(x) = \begin{cases} 1/|C| & \text{for } x \in C \\ 0 & \text{otherwise} \end{cases}$$

3/ Cumulative distribution functions

Cumulative distribution functions

Definition

The **cumulative distribution function (c.d.f.)** is a function $F_X(x)$ that returns the probability is that a variable is less than a particular value:

$$F_X(x) \equiv \mathbb{P}(X \leq x).$$

- Useful for all r.v.s since p.m.f. are unique to discrete r.v.s
- For discrete r.v.: $F_X(x) = \sum_{x_j \leq x} p_X(x_j)$

Example of discrete c.d.f

- Remember example where X is the number of treated units:

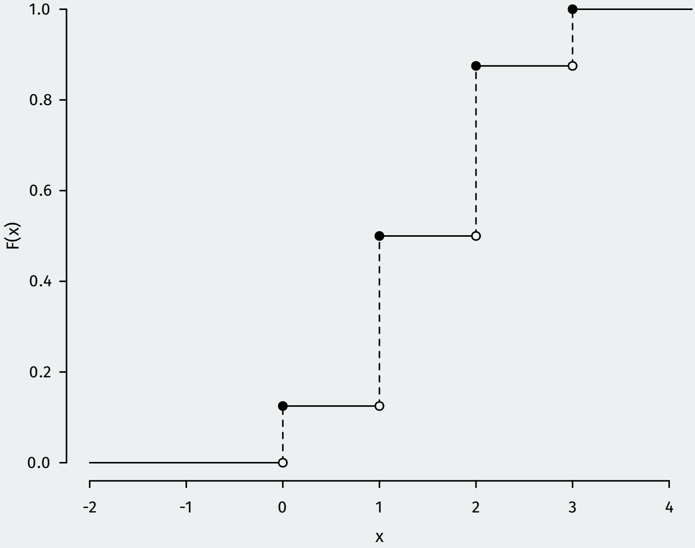
x	$\mathbb{P}(X = x)$
0	1/8
1	3/8
2	3/8
3	1/8

- Let's calculate the c.d.f., $F_X(x) = \mathbb{P}(X \leq x)$ for this:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1/8 & 0 \leq x < 1 \\ 1/2 & 1 \leq x < 2 \\ 7/8 & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

- What is $F_X(1.4)$ here? 0.5

Graph of discrete c.d.f.



Properties of the c.d.f.

- Finding the probability of any region:
  - $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$.
  - $\mathbb{P}(X > a) = 1 - F_X(a)$
- Properties of F_X:

1. **Increasing**: if $x_1 \leq x_2$ then $F_X(x_1) \leq F_X(x_2)$.
   - Proof: the event $X < x_1$ includes the event $X < x_2$ so $\mathbb{P}(X < x_2)$ can't be smaller than $\mathbb{P}(X < x_1)$.
2. **Converges to 0 and 1**: $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$.
3. **Right continuous**: no jumps when we approach a point from the right:

$$F(a) = \lim_{x \to a^+} F(x)$$

4/ Functions of random variables

Transforming a random variable

- $Y =$ numbers of citizens who vote in an election in a population of 1,000.
- We could model the distribution of Y as $\text{Bin}(1000, p)$.
    - Allows us to make statements like $\mathbb{P}(Y \geq 500)$.
- What about the proportion turnout $X = Y/1000$?
    - Can we make statements about $\mathbb{P}(X \geq 0.5)$?

Functions of random variables

- Any function of a random variable is a also a random variable.
- $Y = g(X)$ where $g() : \mathbb{R} \to \mathbb{R}$ is the function that maps from the sample space to $\omega : g(X(\omega))$
    - Let $x_1, \dots, x_k$ be the support of X and $y_j = g(x_j)$ be the support of Y
- If all x_j values map to a single y_j value ("one-to-one"), then we have:

$$\mathbb{P}(Y = g(x_j)) = \mathbb{P}(g(X) = x_j) = \mathbb{P}(X = x_j)$$

- If there are redundancies, we have to add those probabilities together:

$$\mathbb{P}(Y = y_j) = \mathbb{P}(g(X) = y_j) = \sum_{x_i : g(x_i) = y_j} \mathbb{P}(X = x_i)$$

Sum vs mean vs any

- $X \sim \text{Bin}(n, p)$: number of successes.
- $Y = X/n$: proportion of successes (one-to-one)
- $Z = \mathbb{I}(X > 0)$: any successes (not one-to-one)

x	$\mathbb{P}(X = x)$
0	1/8
1	3/8
2	3/8
3	1/8

y	$\mathbb{P}(Y = y)$
0	1/8
1/3	3/8
2/3	3/8
1	1/8

z	$\mathbb{P}(Z = z)$
0	1/8
1	3/8 + 3/8 + 1/8 = 7/8

Careful with r.v.s

- Easy to confuse r.v.s, their distribution, events, and values the r.v.s take.
- A few common examples:
  - If X and Y have the same distribution $\not\Rightarrow \mathbb{P}(X = Y) = 1$
  - Scaling an r.v. doesn't scale the p.m.f., so $Y = 2X$ does not have $p_Y(y) \neq 2p_X(x)$

5/ Independent random variables

Independence of r.v.s

- Two r.v.s are **independent** if

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y)$$

- For many r.v.s:

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1) \times \cdots \times \mathbb{P}(X_n \leq x_n)$$

  - Remember: $X_1, \ldots, X_n$ independent $\implies$ pairwise independent, but not vice versa.

- For discrete r.v.s (not continuous), we can write this as:

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

i.i.d. and the Bern/Bin connection

- **Independent and identically distributed (i.i.d.)** $X_1, \dots, X_n$
  - Identically distributed: all have the same p.m.f./c.d.f.
  - Extremely common data assumption
- Story of the binomial: if $X \sim \text{Bin}(n, p)$, we can write it as $X = X_1 + \cdots + X_n$ where X_i are i.i.d. Bern(p).
- **Theorem:** If $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(m, p)$ with X and Y independent, then $X + Y \sim \text{Bin}(n + m, p)$.

Connections to data

- Statistical modeling in a nutshell:

  1. Assume the data, $X_1, X_2, \ldots$, are i.i.d. with p.m.f. $p_X(x; \theta)$ within a family of distributions (Bernoulli, binomial, etc) with parameter θ.
  2. Use a function of the observed data to **estimate** the value of the θ: $\hat{\theta}(X_1, X_2, \ldots)$

- Example:

  - Sample n respondents from population with replacement.
  - $X_1, X_2, \ldots, X_n$: independent Bernoulli r.v.s indicating Biden approval.
  - p is the Biden approval rate in the population.
  - $\overline{X} = (1/n) \sum_i X_i$ is our estimate of p. Properties?