Chapter 3: Random Variables and Discrete Distributions

MATH/STAT 394: Probability I

Arman Jahangiri

University of Washington Department of Mathematics

Summer 2026

Learning Goals

By the end of this lecture, students should be able to:

- define a random variable;
- compute and interpret PMFs and CDFs;
- recognize and use common discrete distributions;
- work with functions of random variables;

Random Variables

Why Random Variables?

In many experiments, the sample space itself is complicated.

Examples:

- sequences of coin tosses;
- poker hands;
- paths of a random walk;
- survey responses.

Usually we are not interested in the full outcome itself.

Instead, we care about a numerical summary:

- number of Heads (i.e., 3 instead of $HHTHTTT$);
- total score;
- duration of a game;
- number of successes.

Definition of a Random Variable

Definition

Let $(\mathcal{S}, \mathcal{F}, P)$ be a probability space. A **random variable** (r.v.) is a (measurable) function

$$X : \mathcal{S} \to \mathbb{R}.$$

Measurability means that for every (Borel) set $B \in \mathcal{B}$,

$$X^{-1}(B) = \{\, s \in \mathcal{S} : X(s) \in B \,\} \in \mathcal{F}.$$

Important

The randomness comes from the outcome

$$s \in \mathcal{S}.$$

The function X itself is deterministic; it simply converts outcomes into numerical values.

Flip a fair coin twice.

$$\mathcal{S} = \{HH, HT, TH, TT\}.$$

Define:

$$X = \text{number of Heads.}$$

Then,

$$X(HH) = 2,$$
$$X(HT) = 1,$$
$$X(TH) = 1,$$
$$X(TT) = 0.$$

Thus,

$$X : \mathcal{S} \to \{0, 1, 2\}.$$

Indicator Random Variable

For an event A, define

$$I_A = \begin{cases} 1, & \text{if } A \text{ occurs}, \\ 0, & \text{otherwise}. \end{cases}$$

Indicator random variables are extremely useful.

Example.

$$I_{\{\text{first toss is Heads}\}}.$$

Discrete Random Variables

Definition

A random variable is **discrete** if it takes values in a finite or countably infinite set.

Examples:

- number of Heads in 10 tosses;
- number of emails received today;
- number of customers entering a store.

PMFs and Distributions

Probability Mass Functions (PMFs)

Definition

The probability mass function (PMF) of a discrete r.v. X is

$$\boxed{p_X(x) = \mathbb{P}(X = x).}$$

The PMF assigns probability mass to each possible value of the random variable.

PMF Example: Number of Heads

Suppose:

$$X = \text{number of Heads in two fair tosses.}$$

Then,

$$p_X(0) = \frac{1}{4}, \qquad p_X(1) = \frac{1}{2}, \qquad p_X(2) = \frac{1}{4}.$$

All other values have probability 0. We will see in the future that this is called a Binomial random variable with parameters $n = 2$ and $p = 0.5$.

Properties of a Valid PMF

A valid PMF must satisfy:

1. Nonnegativity:
$$p_X(x) \geq 0.$$
2. Total probability equals 1:
$$\sum_x p_X(x) = 1.$$

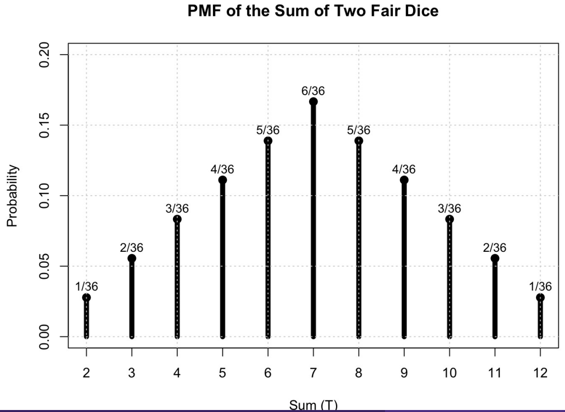
Interpretation

A PMF distributes total probability mass 1 across the support of the random variable.

Example: Sum of Two Dice

Roll two fair dice. Define T = sum of the two die rolls. The possible values of T, is $Range(T) = \{2, 3, \ldots, 12\}$. The PMF table is

t	2	3	4	5	6	7	8	9	10	11	12
$p_T(t) = \mathbb{P}(T = t)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



Named Distributions

Bernoulli Distribution

A Bernoulli random variable models the outcome of a binary experiment: success $\leftrightarrow 1, \qquad$ failure $\leftrightarrow 0.$

Bernoulli PMF

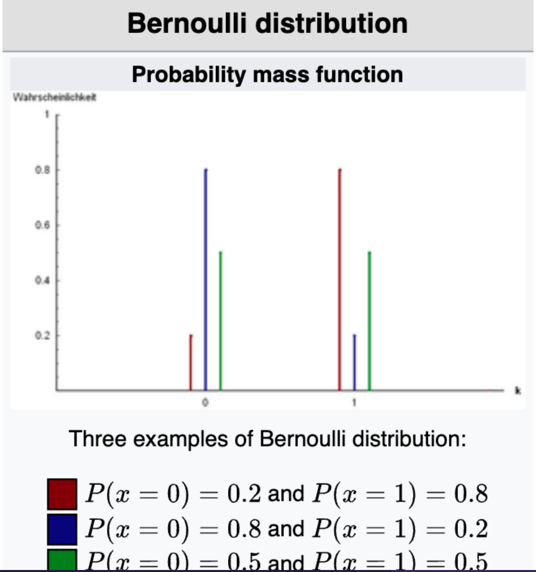
A random variable X has the Bernoulli distribution with parameter $0 \leq p \leq 1$, written $X \sim \text{Ber}(p)$, if its PMF is given by

$$p_X(x) = \begin{cases} p, & x = 1, \\ 1 - p, & x = 0, \\ 0, & \text{otherwise}, \end{cases}$$

Interpretation: $\mathbb{P}(\text{success}) = \mathbb{P}(X = 1) = p, \qquad \mathbb{P}(\text{failure}) = \mathbb{P}(X = 0) = 1 - p$. **PMF verification**: Since $0 \leq p \leq 1$,, we have $p_X(x) \geq 0$. Also,

$$\sum_x p_X(x) = p + (1 - p) = 1.$$

Hence $p_X(x)$ satisfies the properties of a PMF.



$$X \sim \mathrm{Bern}(p),$$

$$\mathbb{P}(X=1)=p,$$

$$\mathbb{P}(X=0)=1-p.$$

A Bernoulli random variable models a single success/failure experiment.

Binomial Distribution

Suppose an experiment consists of

- n independent Bernoulli trials;
- where each trial share the same success probability p.

If X is the number of successes, then

$$X \sim \mathrm{Bin}(n, p).$$

If

$$X \sim \text{Bin}(n, p),$$

then

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

for

$$k = 0, 1, \ldots, n.$$

Interpretation:

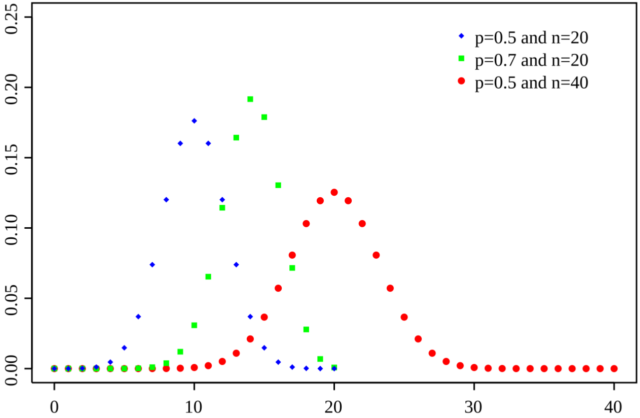
- $\binom{n}{k}$: choose which k trials are successes;
- p^k: probability of the k successes;
- $(1-p)^{n-k}$: probability of the $n-k$ failures.

PMF verification: Since $0 \leq p \leq 1$, we have $p_X(k) \geq 0$. Also,

$$\sum_{k=0}^{n} p_X(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1,$$

by the Binomial Theorem.

Binomial Distribution: PMF



$$X \sim \text{Bin}(n, p),$$

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

Binomial distribution counts the number of successes in n independent Bernoulli trials.

Example: Coin Tosses

Flip a fair coin 10 times.
Let:

$$X = \text{number of Heads.}$$

Then,

$$X \sim \text{Bin}(10, \tfrac{1}{2}).$$

For example:

$$\mathbb{P}(X = 3) = \binom{10}{3} \left(\frac{1}{2}\right)^{10}.$$

Geometric Distribution

Suppose independent Bernoulli trials are repeated until the first success occurs.
Let

$$X = \text{number of trials until the first success.}$$

If each trial share the same success probability p, then

$$X \sim \text{Geom}(p).$$

Possible values:

$$1, 2, 3, \ldots \quad \text{(i.e., all positive integers)}$$

Interpretation:

- $X = 1$ means success on the first trial;
- $X = 2$ means failure then success;
- $X = 3$ means two failures then success;
- etc.

Geometric PMF

If

$$X \sim \text{Geom}(p),$$

then

$$\boxed{\mathbb{P}(X = k) = (1-p)^{k-1}p,}$$

for

$$k = 1, 2, 3, \ldots$$

Interpretation:

- the first $k-1$ trials must fail;
- the k-th trial must succeed.

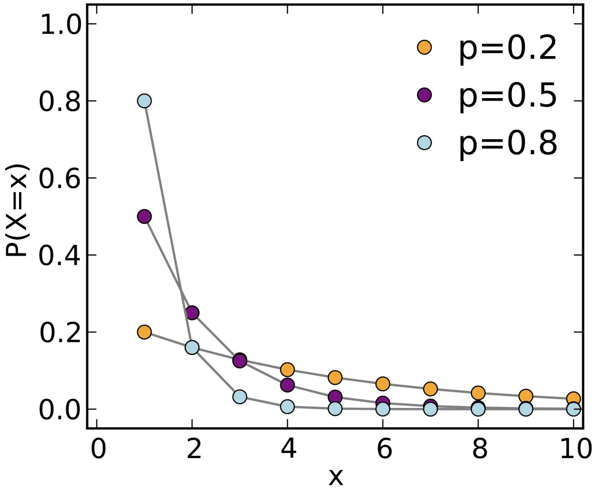
Hence,

$$\mathbb{P}(X = k) = \underbrace{(1-p)^{k-1}}_{\text{failures}} \underbrace{p}_{\text{success}}.$$

PMF verification: Since $0 \le p \le 1$, we have $p_X(x) \ge 0$. Also,

$$\sum_{k=1}^{\infty} p_X(k) = \sum_{k=1}^{\infty} (1-p)^{k-1} p = p \sum_{k=0}^{\infty} (1-p)^k = p \times \frac{1}{1-(1-p)} = p \times \frac{1}{p} = 1,$$

using the geometric sequence formula.



$$X \sim \text{Geom}(p),$$

$$p_X(x) = \mathbb{P}(X = k) = (1-p)^{k-1}p,$$

$$k = 1, 2, 3, \ldots$$

The Geometric distribution models the waiting time until the first success.

Memoryless Property of $Geom(p)$

The Geometric distribution is the **only discrete distribution** that satisfies the **memoryless property**. We will see the continuous counterpart (exponential distribution) in Chapter 5b.

$$\boxed{\mathbb{P}(X > m + n \mid X > m) = \mathbb{P}(X > n).}$$

Interpretation:

If the first m trials were all failures, the process probabilistically restarts from scratch.

Example:

$$\mathbb{P}(X > 10 \mid X > 7) = \mathbb{P}(X > 3).$$

The past failures do not affect future waiting times.

Negative Binomial Distribution

Suppose independent Bernoulli trials are repeated until the rth success occurs. Let

$$X = \text{number of trials needed to obtain } r \text{ successes.}$$

If each trial has success probability p, then

$$X \sim \text{NegBin}(r, p).$$

Possible values:

$$r, r+1, r+2, \ldots$$

Interpretation:

- the experiment stops once the rth success appears;
- Geometric distribution is the special case

$$r = 1.$$

Negative Binomial PMF

If
$$X \sim \text{NegBin}(r, p),$$
then
$$\boxed{p_X(k) = \mathbb{P}(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r},}$$
for
$$k = r, r+1, r+2, \ldots$$

Interpretation:

- among the first $k-1$ trials, there must be exactly $r-1$ successes;
- the k-th trial must be a success.

Thus,
$$\mathbb{P}(X = k) = \underbrace{\binom{k-1}{r-1}}_{\text{arrange first } r-1 \text{ successes}} \underbrace{p^r}_{r \text{ successes}} \underbrace{(1-p)^{k-r}}_{k-r \text{ failures}}.$$

PMF verification: Since $0 \le p \le 1$, we have $p_X(x) \ge 0$.

PMF verification (conitnue)

$$\sum_{k=r}^{\infty} p_X(k) = \sum_{k=r}^{\infty} \binom{k-1}{r-1} p^r (1-p)^{k-r}.$$

Let

$$j = k - r.$$

Then $k = j + r$, so

$$\sum_{k=r}^{\infty} p_X(k) = p^r \sum_{j=0}^{\infty} \binom{j+r-1}{r-1} (1-p)^j.$$

Using the negative binomial series identity,

$$\sum_{j=0}^{\infty} \binom{j+r-1}{r-1} x^j = \frac{1}{(1-x)^r}, \qquad |x| < 1,$$

with $x = 1 - p$, we get

$$\sum_{k=r}^{\infty} p_X(k) = p^r \frac{1}{(1-(1-p))^r} = p^r \frac{1}{p^r} = 1.$$

Understanding the Formula

To have

$$X = k,$$

the following must happen:

1. In the first $k-1$ trials, $r-1$ successes must occur.
2. The final trial must be a success.

Number of ways:

$$\binom{k-1}{r-1}.$$

Probability of each arrangement:

$$p^{r-1}(1-p)^{k-r} \cdot p = p^r(1-p)^{k-r}.$$

Multiplying gives:

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

Note that both the formula and the definition implies that the Negative Binomial distribution is a generalization of the Geometric distribution (when $r = 1$, Negative Binomial reduces to Geometric).

Example: Third Heads

Flip a fair coin repeatedly.
Let

$$X = \text{number of tosses needed to obtain 3 Heads.}$$

Then,

$$X \sim \text{NegBin}(3, \tfrac{1}{2}).$$

For example,

$$\mathbb{P}(X = 5) = \binom{4}{2} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2.$$

Interpretation:

- among the first 4 tosses, exactly 2 must be Heads;
- the 5th toss must also be Heads.

Geometric vs. Negative Binomial

Distribution	Stops After	Support
Geometric	first success	$1, 2, 3, \ldots$
Negative Binomial	rth success	$r, r+1, \ldots$

Key Relationship

The Geometric distribution is a special case of the Negative Binomial distribution:

$$\text{Geom}(p) = \text{NegBin}(1, p).$$

Hypergeometric Distribution

Hypergeometric Story

An urn contains:

- w white balls,
- b black balls.

Sample n balls without replacement.
If X is the number of white balls selected, then

$$X \sim \text{HGeom}(w, b, n).$$

If

$$X \sim \text{HGeom}(w, b, n),$$

then

$$\boxed{\mathbb{P}(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}.}$$

The possible values are

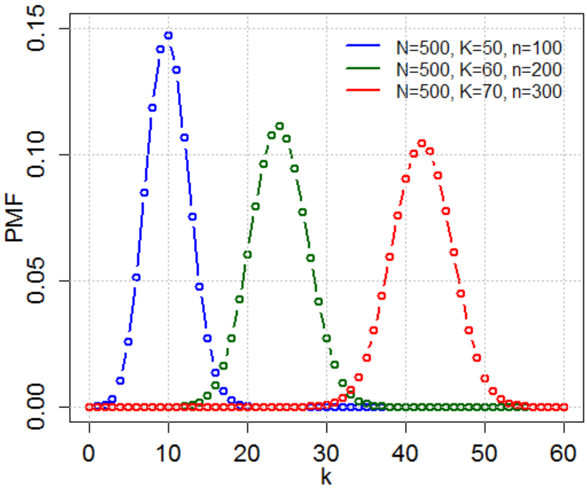
$$\max(0, n-b) \le k \le \min(w, n).$$

Interpretation:

- choose k white balls from the w white balls;
- choose $n-k$ black balls from the b black balls;
- divide by the totral number of samples of size n.

Thus,

$$\mathbb{P}(X = k) = \frac{\underbrace{\binom{w}{k}}_{\text{choose white balls}}\underbrace{\binom{b}{n-k}}_{\text{choose black balls}}}{\underbrace{\binom{w+b}{n}}_{\text{all samples of size } n}}.$$



$$X \sim \mathrm{HGeom}(w, b, n),$$

$$\mathbb{P}(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}.$$

The Hypergeometric distribution arises from sampling without replacement.

Discrete Uniform Distribution

Discrete Uniform

Choose one value uniformly from a finite set C.
Then,

$$X \sim \text{DUnif}(C).$$

All values are equally likely.

Example:

$$X \sim \text{DUnif}(\{1, 2, \ldots, 6\}).$$

Poisson Distribution

A random variable X has the Poisson distribution with parameter $\lambda > 0$ if

$$\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots.$$

Notation:

$$X \sim \text{Pois}(\lambda).$$

Poisson random variables often model counts of rare events.

PMF verification: Since

$$\lambda > 0,$$

we have

$$\mathbb{P}(X = k) \geq 0.$$

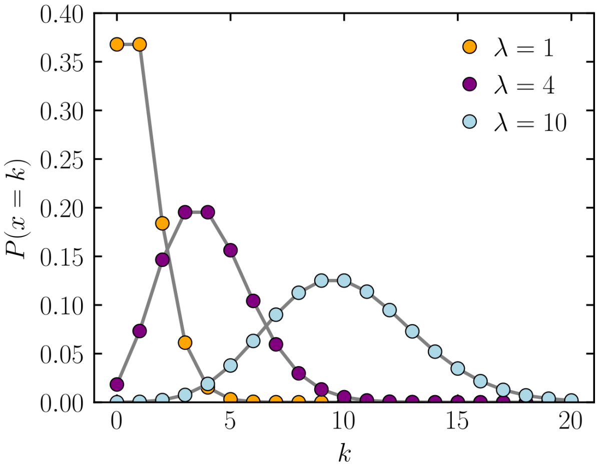
Also,

$$\sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}.$$

Using the exponential series,

$$\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{\lambda},$$

Hence $\mathbb{P}(X = k)$ satisfies the properties of a PMF.



$$X \sim Pois(\lambda),$$

$$\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

The Poisson distribution models the number of events occurring in a fixed interval of time or space.

Typical examples include:

- phone calls arriving,
- radioactive decays,
- website visits,
- accidents in a day.

CDFs

Definition

The cumulative distribution function (CDF) of X is

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

- The CDF accumulates all the probability mass that exists from $-\infty$ to (inclusive) x.
- In other words, $F_X(x) = \mathbb{P}(X \in (-\infty, x]) = \mathbb{P}(-\infty < X \leq x)$ are three different ways of introducing the same thing, the CDF of X at point x.
- Note that $X \leq x$ (or $X \in (-\infty, x]$, or $-\infty < X \leq x$), actually refer to an EVENT:

$$\{X \leq x\} = \{s \in \mathcal{S} \;:\; X(s) \leq x\} = X^{-1}((-\infty, x]))$$

- In general, $\{X \in A\} = \{s \in \mathcal{S} \;:\; X(s) \in A\} = X^{-1}(A)$

PMF vs. CDF

For discrete random variables:

- PMF:

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

  gives probability at one point;
- CDF:

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

  gives cumulative probability.

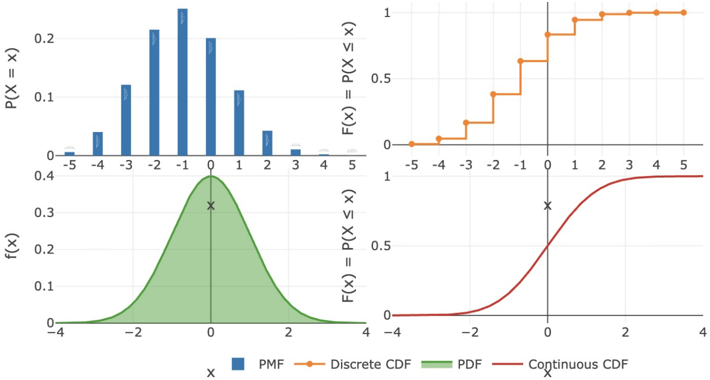
The CDF is increasing and ranges from 0 to 1.

Properties of a CDF

Key properties

Every CDF satisfies:

- $0 \leq F(x) \leq 1$,
- $F(x)$ is increasing,
- $\lim_{x \to -\infty} F(x) = 0$,
- $\lim_{x \to \infty} F(x) = 1$.



Interactive visualization of PMF/PDF and CDF accumulation.

Functions of Random Variables

If X is a random variable and

$$g : \mathbb{R} \to \mathbb{R},$$

then

$$Y = g(X)$$

is also a random variable.

Examples:

$$Y = X^2, \qquad Y = \sqrt{X}, \qquad Y = |X|, \qquad Y = 2X + 1.$$

Question: How to find $F_Y(y)$? We will see later in Transformations of r.v.s.

Several Random Variables on One Space

Different random variables may be defined on the same experiment.

Example:

- X = first die $\in \{1, 2, 3, 4, 5, 6\}$;
- Y = second die $\in \{1, 2, 3, 4, 5, 6\}$;
- $X + Y$ = total $\in \{1, 2, \ldots, 11, 12\}$;
- $\max(X, Y)$ = maximum $\in \{1, 2, 3, 4, 5, 6\}$;
- $|X - Y|$ = difference $\in \{0, 1, 2, 3, 4, 5\}$;

Each gives a different numerical summary of the same experiment.

Summary

- A random variable is a function from the sample space to $\mathbb{R}$.
- Discrete random variables are described by PMFs.
- Important distributions:

  Bernoulli, Binomial, Hypergeometric, Discrete Uniform.

- CDFs accumulate probability:
$$F_X(x) = \mathbb{P}(X \leq x).$$
- Functions of random variables are themselves random variables.
- Random variables help convert complicated experiments into manageable numerical objects.

Next lecture

Expectation