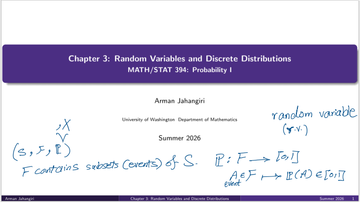
Lecture 5 (July 1)

Wednesday, July 1, 2026 11:57 AM

chapter3_random_vari...



Chapter 3: Random Variables and Discrete Distributions

MATH/STAT 394: Probability I

Arman Jahangiri

University of Washington Department of Mathematics

Summer 2026

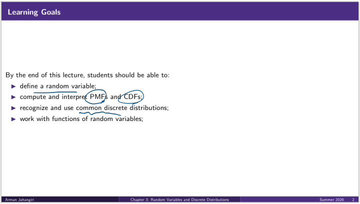
random variable (r.v.)

$$\overset{,X}{(S, \mathcal{F}, \mathbb{P})}$$

$\mathcal{F}$ contains subsets (events) of S.

$\mathbb{P} : \mathcal{F} \longrightarrow [0,1]$

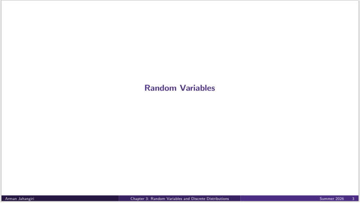
$\underset{\text{event}}{A} \in \mathcal{F} \longmapsto \mathbb{P}(A) \in [0,1]$



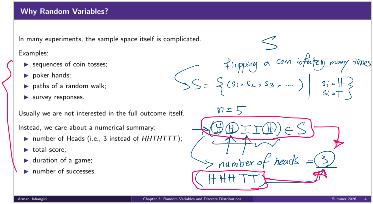
Learning Goals

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.

S

flipping a coin infinitely many times

$$S = \left\{ (s_1, s_2, s_3, \ldots\ldots) \;\middle|\; \begin{matrix} s_i = H \\ s_i = T \end{matrix} \right\}$$

$n = 5$

$(\textcircled{H}\ \textcircled{H}\ T\ T\ \textcircled{H}) \in S$

number of heads $= \textcircled{3}$

$(H\ H\ H\ T\ T)$

Definition of a Random Variable

$\mathbb{P} : \mathcal{F} \to [0,1]$

Definition

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

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

$X(s)$ $s \in S$

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

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

$X(s) \in \mathbb{R}$

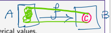
$B \subseteq \mathbb{R}$ $X^{-1}(B) \subseteq S$ (i.e. $X^{-1}(B) \in \mathcal{F}$)

Important

The randomness comes from the outcome

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

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



$f: A \to B$ if $C \subseteq B$ then $f^{-1}(C) = \{a \in A \mid f(a) \in C\} \subseteq A$

Coin Toss Example

Flip a fair coin twice.

$$\mathcal{S} = \{\underbrace{HH}_{s_1}, \underbrace{HT}_{s_2}, \underbrace{TH}_{s_3}, \underbrace{TT}_{s_4}\}.$$

Define:

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

Then,

$$X(HH) = 2,$$
$$\boxed{\begin{aligned} X(HT) &= 1, \\ X(TH) &= 1, \end{aligned}}$$
$$X(TT) = 0.$$

Thus,

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

In general, if we flip a fair coin n times. Then, if X = number of heads

$0 \leq X \leq n$

Indicator Random Variables

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.

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

I_A $A = \{\text{First toss is Heads}\}$

$I_A = 1 \iff \text{first toss} = H$

Discrete Random Variables

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

Definition

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

if Range (X) is countable (or $\operatorname{Im}(X)$)

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)

$\mathbb{P}: \mathcal{F} \to [0,1]$

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

Definition

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

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

small p

$\underset{X}{P}(x) : \text{Image}(X) \longrightarrow [0, 1]$

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

dummy variable

$P_X(z) = \mathbb{P}(X = z)$

your random variable.

$f(z) = z^2 \equiv g(y) = y^2$

PMF Example: Number of Heads

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

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 $p = 0.5$.

$P_X(0) = \mathbb{P}(X=0) = \mathbb{P}(\text{both Tails}) = \mathbb{P}(\text{first toss} = \text{Tails and second toss} = \text{Tails})$

$= \mathbb{P}(\text{first toss} = \text{Tails}) \times \mathbb{P}(\text{second toss} = \text{Tails}) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

$$P_X(x) = \begin{cases} 1/4 & x = 0 \\ 1/2 & x = 1 \\ 1/4 & x = 2 \\ 0 & \text{otherwise} \end{cases}$$

Note that in this example, we computed $\mathbb{P}(X=0) = 1/4$.

We previously said that $\mathbb{P}: \mathcal{F} \to [0,1]$ meaning that $\text{Domain}(\mathbb{P}) = \mathcal{F}$.

So the function $\mathbb{P}(\cdot)$ can assign numbers only to events (elements of $\mathcal{F}$).

But, is $X=0$ an event?

Answer: Actually, YES!

By $X=0$, we mean $\{X=0\} = X^{-1}(\{0\}) = \{TT\} \in \mathcal{F}$.

by definition

also check the definition of a r.v.

For any $B \subseteq \mathbb{R}$, by $\{X \in B\}$ we mean $X^{-1}(B)$.

$\{X \in B\} = X^{-1}(B) \in \mathcal{F}$

$\Rightarrow \mathbb{P}(X \in B) = \mathbb{P}(\{X \in B\}) = \mathbb{P}(X^{-1}(B))$

this makes sense b/c $X^{-1}(B) \in \mathcal{F}$ ($X^{-1}(B)$ is an event, so we can talk about $\mathbb{P}(X^{-1}(B))$)

Properties of a Valid PMF

A valid PMF must satisfy:

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

Interpretation

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

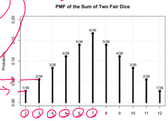
Example: Sum of Two Dice

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

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

Tabular representation of PMF

← values that your r.v. can take



graphical representation of PMF

Named Distributions

Bernoulli Distribution

a generalization of coin toss when the prob. of heads is $p \in [0,1]$

A Bernoulli random variable models the outcome of a binary experiment: success $\overset{X=}{\leftrightarrow}$ 1, failure $\overset{X=}{\leftrightarrow}$ 0.

Bernoulli PMF

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

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

$X=1 \Leftrightarrow$ success

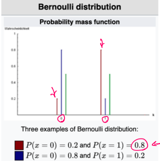
$X=0 \Leftarrow$ failure

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

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

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

Bernoulli Distribution: PMF



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

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

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

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

Binomial Distribution

trial ①
- - - -
trial ⓝ

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

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

follows

each trial results in one of the 2 outcomes (success / failure)

X = # of successes

$0 \le X \le n$

Binomial PMF

If

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

number of trials

prob. of success in trail ⓘ

then

$$\boxed{p_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k},} \ge 0$$

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 $0 \le p \le 1$, we have $p_X(k) \ge 0$. Also,

at home →

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

by the Binomial Theorem.

claim) if $n=1 \Rightarrow \text{Bin}(1,p) \equiv \text{Ber}(p)$

proof

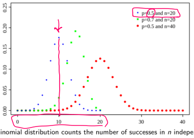
$$\boxed{P_X(k) = \binom{1}{k} p^k (1-p)^{1-k}}$$

← if $k=0$ $\binom{1}{0} p^0 (1-p)^{1-0} = 1-p$

if $k=1$ $\binom{1}{1} p^1 (1-p)^{1-1} = p$

$$\text{if } X \sim \text{Ber}(p) \Rightarrow P_X(x) = \begin{cases} p & x=1 \\ 1-p & x=0 \\ 0 & \text{o.w.} \end{cases}$$

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,

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

$n=10$
$p=\frac{1}{2}$

For example:

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

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

$n=10$
$p=1/2$

$$= \binom{10}{k}\boxed{\left(\tfrac{1}{2}\right)^{k}\left(\tfrac{1}{2}\right)^{10-k}}$$

$$= \binom{10}{k}\underbrace{\left(\tfrac{1}{2}\right)^{k-k}}_{1}\cdot\left(\tfrac{1}{2}\right)^{10}$$

Geometric Distribution

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

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

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

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

the first (k-1) trials
result

for

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

$$\underbrace{\underline{1-p}\ \underline{1-p}\ \underline{1-p} \ldots \underline{1-p}}_{k-1}\ \underbrace{p}_{k\text{th}}$$

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

Interpretation:

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

Hence,

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

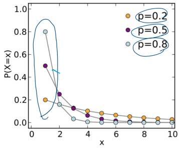
PMF verification: Since $0 \le p \le 1$, we have $p_X(x) \ge 0$. Also,

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

$\frac{1}{1-q}$ $q = 1-p$

using the geometric sequence formula.

Geometric Distribution: PMF



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

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

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

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.

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

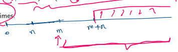
Interpretation:

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

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:

$$X \in \{ r, r+1, r+2, \dots \}$$

Interpretation:

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

$$r = 1.$$

Negative Binomial PMF

claim) if $r=1 \Rightarrow \text{NegBin}(1,p) \equiv \text{Geom}(p)$

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

$$k = r, r+1, r+2, \dots$$

prof if $r=1$

$$p_X(k) = \underbrace{\binom{k-1}{1-1}}_{1} \cdot p^1 (1-p)^{k-1}$$

$$= p(1-p)^{k-1}$$

$$\Downarrow$$

Geom(p)!

Interpretation:

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

Thus,

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

$X=k$

$\equiv$ at k^{th} trial, you have your r^{th} success

$r-1$ sucesses [1 2 . . . . . $k-1$] $\frac{P}{k}$

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

Lecture ⑥

July 5

PMF verification (conitnue)

Let

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

$X \geq r$

$$j = k - r.$$

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,

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

$r=1$ $\quad |x|<1$

$$\sum_{j=0}^{\infty} x^j = \frac{1}{1-x}$$

derivative

$$\frac{-1}{(1-x)^2} \times (-1)$$

$$\parallel$$

$$\frac{1}{(1-x)^2}$$

with $x = 1-p$, we get

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

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,

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

at home

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

$X =$ $\qquad$ range(X) = Image(X)

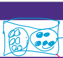
Distribution	Stops After	Support
Geometric	first success	$1, 2, 3, \ldots$
Negative Binomial	rth success	$r, r+1, \ldots$

Key Relationship

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

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

Hypergeometric Distribution



→ b black balls

w white balls

select n balls from the basket

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

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

$b \leftarrow N - w$

$$N = w + b$$

$$p(X=k) = \frac{\binom{w}{k} \times \binom{b}{n-k}}{\binom{w+b}{n}} \leftarrow N$$

Hypergeometric PMF

If $\qquad w + b = N$

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

then $\qquad n \le N = w+b$

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

$\underset{X}{}$

The possible values are $\qquad \boxed{n - b \le w} \quad \boxed{k \le w}$

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

1) $k \le w$
2) $k \le n$
3) $k \ge 0$
4) $\boxed{k \ge n - b}$

$$\boxed{k + b \ge n}$$

Interpretation:

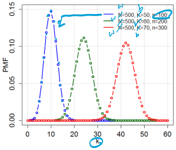
- choose k white balls from the w white balls;
- choose $n - k$ black balls from the b black balls;
- divide by the totral number of samples of size n.

Thus,

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

General fact:
$x \le y$ and $x \le z$
$\Rightarrow x \le \min(y, z) \le \max(y, z)$
$y \le x$ & $z \le x$
$\Rightarrow \max(y, z) \le x$

Hypergeometric Distribution: PMF



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

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

The Hypergeometric distribution arises from sampling without replacement.

Discrete Uniform Distribution

$|C| = 6$

Discrete Uniform

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

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

$C = \{c_1, c_2, \ldots, c_n\}$

All values are equally likely.

Example:

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

$\frac{1}{6} +$
$\frac{1}{6} +$
$\frac{1}{6} \ldots = ①$

$$\sum_{k=c_1}^{c_n} p_X(k) = \sum_{k=c_1}^{c_n} \frac{1}{|C|} = \frac{|C|}{|C|} = 1$$

Poisson Distribution

Poisson Distribution

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

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

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,

$$\times \quad e^{\lambda}$$

$$\sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}. = 1$$

calculus:

$$\sum_{n=0}^{\infty}\frac{x^n}{n!} = e^{x}$$

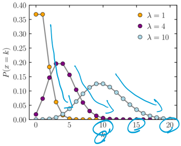
$$x = \lambda$$

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.

Poisson Distribution: 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

Cumulative Distribution Functions

r.v. X → PMF
→ CDF

Definition

The cumulative distribution function (CDF) of X is

PMF: $p_X(k) = \mathbb{P}(X = k)$

$$0 \leq F_X(x) = \mathbb{P}(X \leq x). \leq 1$$

Survival Modelling

$$S_X(x) = \mathbb{P}(X > x)$$

not in this course

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

$x \quad y$

$$x < y$$

$$\underbrace{\{X \leq x\}}_{A} \subseteq \underbrace{\{X \leq y\}}_{B}$$

$$\leq$$

$$P(A) \leq P(B)$$

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.

$F(x) = \boxed{\mathbb{P}(X=x)} + \mathbb{P}(X=x-1) + \cdots$

Properties of a CDF

$F(x) =$

$\{X \le x\} = \{X=x\} \cup \{X=x-1\} \cup \{X=x-2\} \cup \cdots \Rightarrow P(X\le x) = \sum_{k=-\infty}^{x} P(X=k)$

Key properties

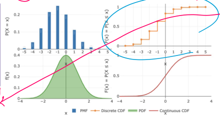
Every CDF satisfies:

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

$\boxed{F_X(x)} = \mathbb{P}(X \le x) = \boxed{\sum_{k\le x} P_X(k)}$

$\mathbb{P}$

$0 \quad 1 \quad 2 \quad 3 \quad 4 \to \mathbb{R}$



Interactive visualization of PMF/PDF and CDF accumulation.

Functions of Random Variables

Functions of Random Variables

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

X^2

$3X$

If X is a random variable and

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

then

$$Y = g(X)$$

is also a random variable.

Examples:

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

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

Binomial $(n,p) \Longrightarrow$ PMF: $P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ $\quad 0 \le k \le n$

CDF: $F_X(x) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$ $\quad x \le n$

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 $\in \{1,2,\ldots,11,12\}$;
- Z = $\max(X,Y)$ = 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:

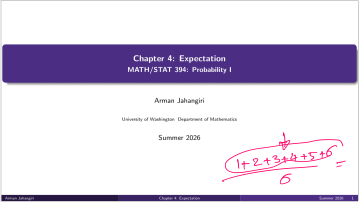
  $$F_X(x) = \mathbb{P}(X \le x).$$

- Functions of random variables are themselves random variables.
- Random variables help convert complicated experiments into manageable numerical objects.

Next lecture

Expectation

chapter4_expectation



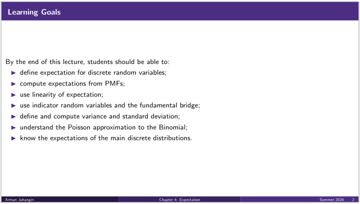
Chapter 4: Expectation
MATH/STAT 394: Probability I

Arman Jahangiri

University of Washington Department of Mathematics

Summer 2026

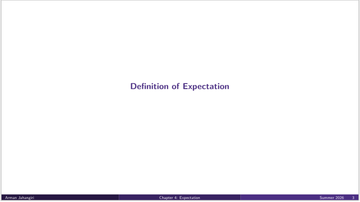
$$\frac{1+2+3+4+5+6}{6} = \frac{6\times 7}{6\times 2} = \boxed{3.5}$$



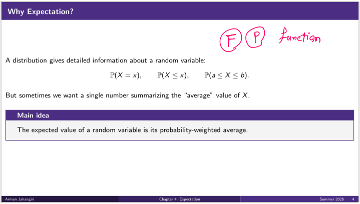
Learning Goals

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

- define expectation for discrete random variables;
- compute expectations from PMFs;
- use linearity of expectation;
- use indicator random variables and the fundamental bridge;
- define and compute variance and standard deviation;
- understand the Poisson approximation to the Binomial;
- know the expectations of the main discrete distributions.



Definition of Expectation



Why Expectation?

Ⓕ Ⓟ function

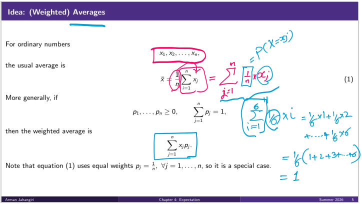
A distribution gives detailed information about a random variable:

$$\mathbb{P}(X = x), \qquad \mathbb{P}(X \le x), \qquad \mathbb{P}(a \le X \le b).$$

But sometimes we want a single number summarizing the "average" value of X.

Main idea

The expected value of a random variable is its probability-weighted average.



Idea: (Weighted) Averages

For ordinary numbers

$$\boxed{x_1, x_2, \dots, x_n,}$$

the usual average is

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_j = \sum_{j=1}^{n} \boxed{\frac{1}{n}} \times x_j \tag{1}$$

$= P(X=x_j)$

More generally, if

$$p_1, \dots, p_n \ge 0, \qquad \sum_{j=1}^{n} p_j = 1,$$

then the weighted average is

$$\boxed{\sum_{j=1}^{n} x_j p_j.}$$

$$\sum_{i=1}^{6} \frac{1}{6} \times i = \frac{1}{6}\times 1 + \frac{1}{6}\times 2 + \dots + \frac{1}{6}\times 6$$
$$= \frac{1}{6}(1+2+3+\dots+6)$$
$$= 1$$

Note that equation (1) uses equal weights $p_j = \frac{1}{n}$, $\forall j = 1, \dots, n$, so it is a special case.

Toss a die ⇒ X = die roll ⇒ $\mathbb{E}[X] = ?$

Definition: Expected Value

Definition (Expected value of a discrete r.v.)

Let X be a discrete random variable with possible values

$$x_1, x_2, \ldots .$$

The expected value of X is

$$\boxed{\mathbb{E}(X) = \sum_j x_j \mathbb{P}(X = x_j).}$$

$= \boxed{x_1} \times \underline{P(X=x_1)} + \boxed{x_2} \times \underline{P(X=x_2)} + \cdots$

Equivalently,

$$\boxed{\mathbb{E}(X) = \sum_x \boxed{x}\, p_X(x),}$$

$$\boxed{\sum_{x=0}^{\infty} P_X(x) = 1}$$

where

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

Interpretation

Expectation is the center of mass of the PMF (next slide).

$\underline{1} \cdot \underbrace{P(X=1)}_{1/6} + \underline{2} \times \underbrace{P(X=2)}_{1/6} + \underline{3} \times \underbrace{P(X=3)}_{''} + \underline{4} \times \underbrace{P(X=4)}_{''} + \underline{5} \times \underbrace{P(X=5)}_{''} + \underline{6} \times \underbrace{P(X=6)}_{''}$

$= 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = \underline{3.5}$

⊛ $E(X)$ is unique.

$$\begin{tikzcd} X \arrow[r] \arrow[dr] & E(X) \arrow[d, phantom, "\shortparallel"] \\ & E'(X) \end{tikzcd}$$

Expectation as a Center of Mass (Optional)

Suppose a discrete random variable X takes values

$$x_1, x_2, \ldots$$

with probabilities

$$p_1, p_2, \ldots$$

Think of the PMF as placing masses p_i at locations x_i on the number line. The **center of mass** (balance point) is a point c such that

$$\sum_i (x_i - c) p_i = 0.$$

This means:

- points to the right of c pull one way,
- points to the left of c pull the other way,
- and the system balances perfectly at c.

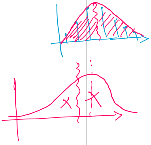
Expanding the equation:

$$\sum_i x_i p_i - c \sum_i p_i = 0.$$

Since

$$\sum_i p_i = 1,$$

we obtain



Example: One Fair Die

Let

$$X = \text{result of one fair die roll.}$$

Then

$$\mathbb{P}(X = k) = \frac{1}{6}, \qquad k = 1, 2, \ldots, 6.$$

Thus

$$\mathbb{E}(X) = \sum_{k=1}^{6} k \mathbb{P}(X = k) = \frac{1+2+3+4+5+6}{6}.$$

So

$$\boxed{\mathbb{E}(X) = 3.5.}$$

Important

The expectation does not have to be a possible value of X.

Example: Bernoulli

$X \in \{0, 1\}$

Let

$$\boxed{X \sim \text{Bern}(p).}$$

x	0	1
$P(X=x)$	$1-p$	p

Then

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

Therefore

$$\mathbb{E}(X) = 1 \cdot p + 0 \cdot (1 - p) = p.$$

So

$$\boxed{\mathbb{E}(X) = p.}$$

$\underbrace{0 \times (1-p) + 1 \times p}_{\sum_{x_i} x_i \times P(X=x_i)} = p$

Example: Discrete Uniform

Let

$$X \sim \text{DUnif}(\{\boxed{a}, a+1, \ldots, b\}).$$

$|C| < \infty$

$P(X=k) = \boxed{\frac{1}{b-a+1}}$ $a \le k \le b$

$\underbrace{C = \{a, a+1, a+2, \ldots, b\}}$

$|C| = b - a + 1$

All values are equally likely.

There are

$$b - a + 1$$

possible values.

Thus

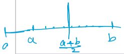
$$\boxed{\mathbb{E}(X)} = \boxed{\frac{\boxed{a} + \boxed{(a+1)} + \cdots + \boxed{b}}{\boxed{b - a + 1}}}.$$

$\sum_{k=a}^{b} \boxed{k} \times P(X=k)$

$= \sum_{k=a}^{b} k \cdot \boxed{\frac{1}{b-a+1}} = \underline{\frac{1}{b-a+1}} \sum_{k=a}^{b} k$

The average of equally spaced endpoints is the midpoint:

$$\boxed{\mathbb{E}(X) = \boxed{\frac{a+b}{2}}.}$$



Example: Geometric Distribution

Let

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

Range(X)

Recall:

$$\boxed{\mathbb{P}(X=k) = (1-p)^{k-1}p, \qquad k = 1, 2, \dots}$$

Then

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p.$$

Using the geometric-series identity,

$$\boxed{\sum_{k=1}^{\infty} k r^{k-1} = \frac{1}{(1-r)^2}}, \qquad |r| < 1,$$

with $r = 1 - p$, we obtain

$$\mathbb{E}(X) = \frac{1}{p}.$$

$|x| < 1$

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}$$

$$\Downarrow$$

$$\sum_{i=1}^{\infty} i x^{i-1} = \frac{1}{(1-x)^2}$$

Example: Negative Binomial Distribution

Let

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

where X counts the number of trials needed to obtain r successes. Write

$$X = X_1 + \cdots + X_r,$$

where each

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

Therefore,

$$\mathbb{E}(X) = r \cdot \frac{1}{p}.$$

Hence

$$\mathbb{E}(X) = \frac{r}{p}.$$

Example: Poisson Distribution

Let

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

Recall:

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

Then

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

Rearranging,

$$= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}.$$

Since

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

we obtain

$$\mathbb{E}(X) = \lambda.$$

Expectation Depends Only on the Distribution

Proposition

If X and Y have the same distribution, then

$$\mathbb{E}(X) = \mathbb{E}(Y),$$

provided the expectations exist.

Reason: the formula

$$\mathbb{E}(X) = \sum_x x\mathbb{P}(X = x)$$

depends only on the PMF of X.

Warning

The converse is false: two random variables can have the same expectation but different distributions.

Linearity of Expectation

Remark

$$X = X_1 + X_2 + \cdots + X_n$$

(Example) $X \sim \text{Binomial}(n=2, p)$

$$X \in \{0, 1, 2\}$$

$$X = X_1 + X_2$$

X: number of successes in 2 ~~experim~~ trials (i.e. trial 1 & trial 2)

X_1: number of successes in trial ① $\in \{0,1\}$

X_2: number of successes in trial ② $\in \{0,1\}$

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

$$X = X_1 + X_2 + \cdots + X_n$$

for any $i = 1, \dots, n$:

$$X_i = \begin{cases} 1 & \text{trial } ⓘ \text{ is successful} \\ 0 & \text{o.w} \end{cases}$$

$$X_i \sim \text{Ber}(p)$$

This proves that any Binomial (n,p) random variable can be written as sum of $\underline{\underline{n}}$ independent Ber(p) random variable.

$$\mathbb{E}[X+Y] = \mathbb{E}(X) + \mathbb{E}(Y)$$

special case of (*) in slide ⑯

$$X = X_1 + \cdots + X_n \implies \boxed{\mathbb{E}(X) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)}$$

$$\boxed{\mathbb{E}\Big(\underbrace{\sum X_i}_{X}\Big) = \sum_{i=1}^{n} \mathbb{E}(X_i)}$$

$$\boxed{\sum_{x_i=0}^{n} x_i \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i}}$$

$$\mathbb{E}(\text{Bin}(n,p)) = \sum_{i=1}^{n} \underbrace{\mathbb{E}(\text{Ber}(p))}_{p} = \boxed{n \times p}$$

Linearity of Expectation

$\sqrt{X+Y} \neq \sqrt{X} + \sqrt{Y}$

Theorem

For random variables X, Y and constant c,

$$\boxed{\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)}$$

Also,

$$\boxed{\mathbb{E}(cX) = c\mathbb{E}(X).}$$

$c \neq 0$

$\mathbb{E}(X) = \sum_i x_i P(X = x_i)$

$\mathbb{E}(cX) = \sum_i y_i P(Y = y_i)$

$= \sum_i c \cdot x_i \times \mathbb{P}(c \cdot X = c \cdot x_i)$

$= c \cdot \sum x_i P(X = x_i)$

$= \boxed{\mathbb{E}(X)}$

More generally,

Generalization $\Rightarrow$

$$\mathbb{E}\left(\sum_{j=1}^{n} c_j X_j\right) = \sum_{j=1}^{n} c_j \mathbb{E}(X_j).$$

Application of Linearity in finding the Expectation of Bin(n,p)

Let

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

Think of X as the number of successes in n Bernoulli trials.
Write

$$X = I_1 + \cdots + I_n,$$

where

$$I_j = \begin{cases} 1 & \text{success on trial } j, \\ 0 & \text{otherwise.} \end{cases}$$

Each

$$I_j \sim \text{Bern}(p),$$

so

$$\mathbb{E}(I_j) = p.$$

Therefore, by linearity of expectation,

linearity of expectation

$$\mathbb{E}(X) = \mathbb{E}[I_1 + \cdots + I_n] = \underset{p}{\mathbb{E}[I_1]} + \cdots + \underset{p}{\mathbb{E}[I_n]} = \underbrace{p + \cdots + p}_{n \text{ times}} = np.$$

Note that we did not even use the PMF of $Ber(n, p)$ to find its expectation using the main formula, which made it much easier computationally.

Example: Hypergeometric Expectation

Let

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

where X is the number of white balls in a sample of size n drawn without replacement.
Write

$$X = I_1 + \cdots + I_n,$$

0 or 1

where $I_j = 1$ if the jth draw is white.
For each j,

$$\mathbb{P}(I_j = 1) = \frac{w}{w+b}.$$

So

$$\mathbb{E}(I_j) = \frac{w}{w+b}.$$

By linearity,

$E(X) = E(I_1 + \cdots + I_n)$

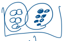
$= E(I_1) + \cdots + E(I_n)$

$= \frac{w}{w+b} + \cdots + \frac{w}{w+b}$

$= n \times \frac{w}{w+b}$

$$\boxed{\mathbb{E}(X) = n\frac{w}{w+b}.}$$

draw



w+b balls

$$\mathbb{E}(X) = \sum_{\min \le k \le \max} k \cdot \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$$

Key point

The draws are dependent, but linearity still works.

of expectation

Monotonicity of Expectation

Theorem

If

$$X \geq Y$$

with probability 1, then

$$\mathbb{E}(X) \geq \mathbb{E}(Y).$$

~~$X \geq Y$~~

~~$x_i \geq y_i$ for all i~~

~~$\sum x_i P(X = x_i)$~~

Sketch of proof:

$$X - Y \geq 0.$$

Therefore,

$$\mathbb{E}(X - Y) \geq 0.$$

By linearity,

$$\mathbb{E}(X) - \mathbb{E}(Y) \geq 0.$$

$Z \geq 0$

$z_i \geq 0$

$\mathbb{E}(Z) = \sum \underset{\geq 0}{z_i} \underbrace{P(Z = z_i)}_{\geq 0} \geq 0$

For a concise proof, we need more technical tools that are not covered in this course.

Indicator Random Variables and the Fundamental Bridge

Indicator Random Variables

For an event A, define

$$I_A = \begin{cases} 1, & \text{if } A \text{ occurs,} \\ 0, & \text{if } A \text{ does not occur.} \end{cases} \quad (A^c)$$

Then

$$\boxed{I_A \sim \text{Bern}(\mathbb{P}(A)).} \implies \mathbb{E}(I_A) = \underline{P(A)}$$

Interpretation

Indicators convert events into random variables.

Basic Indicator Properties

For events A and B:

$$I_{A^c} = 1 - I_A.$$

$$I_{A\cap B} = I_A I_B. \quad = 1$$

$$I_{A\cup B} = I_A + I_B - I_A I_B.$$

Reason

Indicators behave like yes/no algebra:

$$1 = \text{yes}, \qquad 0 = \text{no}.$$

The Fundamental Bridge

Fundamental Bridge

For any event $A \subseteq \mathcal{S}$,

$$\boxed{\mathbb{E}(I_A) = \mathbb{P}(A).}$$

Proof:

$$I_A \sim \text{Bern}(\mathbb{P}(A)).$$

Since the expectation of a Bernoulli random variable is its success probability,

$$\mathbb{E}(I_A) = \mathbb{P}(A).$$

Meaning

Probability questions can often be turned into expectation questions, and vice versa.

Example: Matching Problem

A deck has cards labeled

$$1, 2, \ldots, n.$$

interesting example (at home)

After shuffling, a match occurs if card j is in position j.
Let

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

Define

$$I_j = \begin{cases} 1, & \text{card } j \text{ is in position } j, \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$X = I_1 + \cdots + I_n.$$

4th

Since

$$\mathbb{P}(I_j = 1) = \frac{1}{n},$$

we get

$$\mathbb{E}(X) = n \cdot \frac{1}{n} = 1.$$

$$\boxed{\mathbb{E}(X) = 1.}$$

Example: Birthday Matches

Suppose there are n people.
Let

$$X = \text{number of pairs of people with the same birthday.}$$

Cont.

There are

$$\binom{n}{2}$$

pairs.
For each pair (i, j), define

$$I_{ij} = \begin{cases} 1, & \text{people } i, j \text{ have the same birthday,} \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$X = \sum_{i<j} I_{ij}.$$

Since

$$\mathbb{E}(I_{ij}) = \mathbb{P}(I_{ij} = 1) = \frac{1}{365},$$

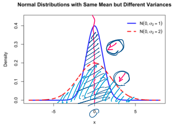
we have

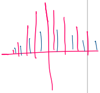
$$\mathbb{E}(X) = \binom{n}{2} \frac{1}{365}.$$

Variance

Why Variance?

Expectation measures the center of a distribution.
But two distributions can have the same expectation and very different spread.
We need a numerical measure of variability.
Variance measures the average squared distance from the mean.





Definition: Variance and Standard Deviation

$\mathbb{E}\left[X-\mathbb{E}(X)\right]^2$

Definition

The variance of X is

$$\boxed{\mathrm{Var}(X)=\mathbb{E}\left[(X-\mathbb{E}X)^2\right].}$$

Variance is a measure of dispersion (from the mean). It is a measure of how far a set of numbers are spread out from their mean, on average.

The standard deviation is

$$\boxed{\mathrm{SD}(X)=\sqrt{\mathrm{Var}(X)}.}$$

Variance has squared units; standard deviation returns to the original units.

Motivation behind the formula $\mathbb{E}[X-\mathbb{E}[X]]^2$?

Consider the function $\boxed{g(\mathbb{E}(X))=\mathrm{Var}(X)}$ $\boxed{g(c)=\mathbb{E}[(X-c)^2].}$

$\mathrm{Var}(X)=\mathbb{E}\left[\left[X-\overset{g}{\mathbb{E}(X)}\right]^2\right]$

The quantity $g(c)$ measures the average squared distance between X and the constant c.
We claim that $g(c)$ is minimized when

$$\rightarrow \boxed{c=\mathbb{E}[X].}$$

$c=\mathbb{E}(X)$

$\frac{d}{dc}g(c)=0$

$\Downarrow$

$c=\mathbb{E}(X)$

Expand the square:

$$g(c)=\mathbb{E}\left[(\overset{a}{X-\mathbb{E}[X]}+\overset{b}{\mathbb{E}[X]-c})^2\right]$$

$$\overset{a^2}{=}\mathbb{E}\left[(X-\mathbb{E}[X])^2\right]+\underbrace{2(\mathbb{E}[X]-c)\boxed{\mathbb{E}[X-\mathbb{E}[X]]}}_{2ab}$$

$$\overset{b^2}{+}(\mathbb{E}[X]-c)^2 \geq 0$$

$\mathrm{Var}(X)$

$g(c)=\mathrm{Var}(X)$ + nonnegative

$c=\mathbb{E}(X) \Leftrightarrow !!$

$\mathbb{E}(X)-\mathbb{E}(\mathbb{E}(X))$ $=0$

$\mathbb{E}(X)=$

$\boxed{\mathbb{E}(2)=2}$ ✓

Since $\mathbb{E}[X-\mathbb{E}[X]]=0$,, we obtain $g(c)=\mathrm{Var}(X)+(\mathbb{E}[X]-c)^2$.
Because

$$(\mathbb{E}[X]-c)^2\geq 0,$$

the minimum occurs when

$$c=\mathbb{E}[X].$$

Thus, the expectation is the *best constant predictor* of X in mean squared error.

Alternative Formula for Variance

Variance Formula

computationally easier ✓ $\boxed{\mathrm{Var}(X)=\mathbb{E}(X^2)-(\mathbb{E}X)^2.}$ $\mathrm{Var}(X)=\mathbb{E}\left[[X-\mathbb{E}(X)]^2\right]$

Proof : Let

$$\boxed{\mu=\mathbb{E}X.}$$

constant

$\boxed{\mathbb{E}(\mu)=\mu}$ $\sigma^2=\mathrm{Var}(X)$

Then $\mathbb{E}[X+3]$
$\mathbb{E}[X]+3$

$$\mathrm{Var}(X)=\mathbb{E}\left[(X-\mu)^2\right].$$

$\boxed{\mathbb{E}(aX+b)=a\mathbb{E}(X)+b}$ from before

Expanding:

linearity of $\mathbb{E}(\cdot)$

$$\mathbb{E}(X^2-2\mu X+\mu^2)=\mathbb{E}(X^2)-\underbrace{2\mu\mathbb{E}(X)}_{=-2\mu\times\mu\;=-2\mu^2}+\mu^2$$

Since $\mu=\mathbb{E}X$,

$$\mathrm{Var}(X)=\mathbb{E}(X^2)-\mu^2.$$

$=\mathbb{E}(X^2)-(\mathbb{E}X)^2$

$\mathbb{E}[X]$ $\mathbb{E}X$

Properties of Variance

Shifting does not change spread. Scaling by b scales squared distances by b^2.

$\mathbb{E}[X+a]=\mathbb{E}(X)+a$
$\mathbb{E}(bX)=b\mathbb{E}(X)$

//

$\mathbb{E}[aX+b]$
$=a\mathbb{E}(X)+b$

① $\boxed{\mathrm{Var}(X+a)=\mathrm{Var}(X)\text{ ✓ and } \overset{②}{\mathrm{Var}(bX)}=b^2\,\mathrm{Var}(X).}$ ✓

Warning

Unlike Expectation, Variance is not linear, i.e., in general,

$\mathbb{E}(X_1+\cdots+X_n)=\mathbb{E}(X_1)+\cdots+\mathbb{E}(X_n)$

$$\mathrm{Var}(X+Y)\neq\mathrm{Var}(X)+\mathrm{Var}(Y).$$

But if $X\perp Y$ (X & Y are independent), then $\mathrm{Var}[X+Y]=\mathrm{Var}[X]+\mathrm{Var}[Y]$

In fact,

$$Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y) \tag{2}$$

and for $n>1$ random variables,

proof of ①

let $Y=X+a$. We aim to show $\mathrm{Var}(\overset{X+a}{Y})=\mathrm{Var}(X)$.

$\mathrm{Var}(X)=\mathbb{E}\left[(X-\mathbb{E}[X])^2\right]$

$$\mathrm{Var}(Y)=\mathbb{E}\left[(\underset{X+a}{Y}-\mathbb{E}[Y])^2\right]=\mathbb{E}\left[((X+a)-(\mathbb{E}[X]+a))^2\right]=\mathbb{E}\left[(X-\mathbb{E}(X))^2\right]$$

$$=\mathrm{Var}(X)$$

$\mathbb{E}[Y]=\mathbb{E}[X+a]=\mathbb{E}[X]+\mathbb{E}[a]$

linearity of $\mathbb{E}(\cdot)$

$\mathbb{E}[a]=a$: expected value of a constant is itself.

Remember)
$\mathbb{E}[aX+b]=a\mathbb{E}[X]+b$ for any constants $a,b\in\mathbb{R}$.

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j) \quad (3)$$

(skip)

We will study this in further depth in Chapter 7 (Joint Distributions).

Variance of a Binomial Random Variable

Let

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

Write

$$I_j = \begin{cases} 1 & j^{th} \text{ trial} = \text{success} \\ 0 & \text{o.w.} \end{cases}$$

$$X = I_1 + \cdots + I_n,$$

where the I_j are independent Bernoulli(p) random variables. Each has variance

$$\text{Var}(I_j) = p(1-p).$$

Since the indicators are independent (*),

$X \sim Ber(p)$
$E(X) = p$, $Var(X) = p \cdot (1-p)$

$$\text{Var}(X) = \sum_{j=1}^{n} \overset{p(1-p)}{\text{Var}(I_j)}.$$

Therefore

$$\text{Var}(X) = np(1-p).$$

$$Var X = \mathbb{E}[[X - \mathbb{E}(X)]^2]$$
$$= \mathbb{E}(X^2) - [\underbrace{\mathbb{E}(X)}_{np}]^2$$
$$= \mathbb{E}(X^2) - n^2p^2 = \cdots = np(1-p)$$
$$\mathbb{E}(X^2) = \sum_{k=0}^{n} k^2 \cdot P(X=k)$$
$$= \boxed{\sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k}}$$

painful :)

(*): Note that we previously proved that the expected value of a sum of random variables is equal to the sum of their expected values. However, this property does *not* generally hold for variance. In general, the variance of a sum is *not* equal to the sum of the variances. Variances add only in the special case where the random variables have zero covariance, that is, when they are uncorrelated. In particular, independence implies zero covariance, so variances of independent random variables do add. We will study covariance, correlation, and independence in more detail later. For a direct derivation of the variance of the Binomial distribution using its PMF, click here.

A Useful Variance Formula (exercise)

$$Var(X) = \mathbb{E}[(X - EX)^2]$$
$$= E(X^2) - (EX)^2$$
$$= E(X(X-1)) + E(X) - (EX)^2$$

please do this at home

Show that for any random variable X,

$$\text{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2.$$

Exercise

E~~xer~~site

Discrete Uniform Variance

Let $D\,Uinf\{a, a+1, \dots, b\}$

$$X \sim \text{DUnif}\{1, 2, \dots, n\}.$$

equal likelihood:

$$a = p(X=1) = p(X=2) = \cdots = p(X=n)$$
$$1 = \underbrace{\sum_{i=1}^{n} p(X=i)}_{p(S)} = \sum_{i=1}^{n} a = n \cdot a$$
$$a = \frac{1}{n}$$

Then

$$\mathbb{E}(X) = \frac{a+b}{2} \xrightarrow[b=n]{a=1}$$

$$\mathbb{E}[X] = \frac{n+1}{2}.$$

Also,

$$Var[X] = E[X^2] - (\mathbb{E}(X))^2$$

$$\mathbb{E}[X^2] = \frac{1}{n}\sum_{k=1}^{n} k^2 = \frac{1}{\cancel{n}} \cdot \frac{\cancel{n}(n+1)(2n+1)}{6} = \boxed{\frac{(n+1)(2n+1)}{6}}.$$

$$\sum_{x=1}^{n} x^2 \cdot \underbrace{p(X=x)}_{1/n} = \frac{1}{n}\sum_{x=1}^{n} x^2$$

Therefore,

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

$$\text{Var}(X) = \frac{n^2-1}{12}.$$

① $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$

② $1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$

$$+\begin{cases} S = 1 + 2 + \cdots + n \\ S = n + (n-1) + \cdots + 1 \end{cases}$$
$$2S = (n+1) + (n+1) + \cdots + (n+1) = n(n+1) \Rightarrow \boxed{S = \frac{n(n+1)}{2}}$$

Poisson Variance

Let

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

$$\sum_{k=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^k}{k!} = 1$$

Then

PMF

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

$$\mathbb{E}(X) = \lambda$$

We use

$$\mathbb{E}[X(X-1)] = \sum_{k=0}^{\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!}. \quad k! = k \cdot (k-1) \cdot (k-2)!$$

$j = k-2$

For $k \geq 2$,

$$e^{-\lambda} k(k-1)\frac{\lambda^k}{k!} = \boxed{\frac{e^{-\lambda}\lambda^k}{(k-2)!}}. \Rightarrow \sum_{k=2}^{\infty} \frac{e^{-\lambda}\lambda^{k-2}\lambda^2}{(k-2)!} = e^{-\lambda} \cdot \lambda^2 \cdot \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!}$$
$$= e^{-\lambda} \cdot \lambda^2 \cdot \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$$

calculus:

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x \quad \forall x \in \mathbb{R}$$

Thus,

$$\mathbb{E}[X(X-1)] = \lambda^2 \cancel{e^{-\lambda}} \underbrace{\sum_{j=0}^{\infty}\frac{\lambda^j}{j!}}_{= \cancel{e^{\lambda}}} = \lambda^2.$$

Poisson Variance Continued

Since

$$\mathbb{E}[X] = \lambda,$$

we get

$$\text{Var}(X) = \underbrace{\mathbb{E}[X(X-1)]}_{\lambda^2} + \mathbb{E}[X] - (\mathbb{E}[X])^2.$$

Therefore,

$$\text{Var}(X) = \cancel{\lambda^2} + \lambda - \cancel{\lambda^2}.$$

$$\text{Var}(X) = \lambda.$$

So for a Poisson random variable,

$$\mathbb{E}[X] = \text{Var}(X) = \lambda.$$

dist	$\mathbb{E}(\cdot)$	$Var(\cdot)$
Ber(p)	p	$p(1-p)$
Bin(n,p)	$n \cdot p$	$np(1-p)$
poisson(λ)	λ	λ
Geom(p)	$1/p$	$\frac{1-p}{p^2}$

$$Var[aX + b] = a^2 Var[X]$$

b: no effect on variance

$$Var(bX) = b^2 Var(X)$$

proof of ②

let $Y = bX$. We want to show that $Var(Y) = b^2 Var(X)$

$$Var(Y) = \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \mathbb{E}[(bX - b\mathbb{E}[X])^2]$$

$Y = bX$, $\mathbb{E}[Y] = \mathbb{E}[bX] = b\mathbb{E}[X]$

$$= \mathbb{E}[b^2(X - \mathbb{E}[X])^2]$$
$$= b^2 \cdot \mathbb{E}[(X - \mathbb{E}[X])^2]$$
$$= b^2 \cdot Var(X) \quad ▨$$

by the linearity of Expectation

$$Var(900X + 300) = 900^2 Var(X)$$

If $X_1, \dots, X_n$ are independent then

$$Var[X_1 + \cdots + X_n] = Var[X_1] + \cdots + Var[X_n]$$

(and more generally, $Var[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i^2 Var[X_i]$)

Remember that this was true for $\mathbb{E}[\cdot]$ even with no independence assumption ($\mathbb{E}[\sum a_i X_i] = \sum a_i \mathbb{E}[X_i]$)

Geometric Variance

Let

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

where

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

Thus,

$$\boxed{\mathbb{P}(X = k) = (1-p)^{k-1}p,} \qquad k = 1, 2, 3, \dots$$

Let $q = 1 - p$. Then

$$\mathbb{E}[X] = \frac{1}{p}.$$

$$Var[X] = \underline{E(X^2)} - \underbrace{(EX)^2}_{(1/p)^2}$$

A standard power series identity gives

$$\sum_{k=1}^{\infty} k^2 q^{k-1} = \frac{1+q}{(1-q)^3}.$$

$$\boxed{\sum_{i=} (\;) x^i = \frac{1}{(1-x)^r}}$$

Therefore,

$$E(X^2) = \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} p = p \sum_{k=1}^{\infty} k^2 q^{k-1}$$

$$\mathbb{E}[X^2] = p \sum_{k=1}^{\infty} k^2 q^{k-1} = p \cdot \frac{1+q}{(1-q)^3}.$$

Geometric Variance Continued

Since

$$1 - q = p,$$

we have

$$\mathbb{E}[X^2] = \cancel{p} \frac{1+q}{p^{\cancel{3}}} = \frac{1+q}{p^2}.$$

Thus,

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

So

$$\text{Var}(X) = \frac{\cancel{1}+q}{p^2} - \frac{\cancel{1}}{p^2} = \frac{q}{p^2}.$$

Since $q = 1 - p$,

$$\boxed{\text{Var}(X) = \frac{1-p}{p^2}.} \neq \underbrace{E(X)}_{1/p}$$

Negative Binomial Variance

(please do this at home)

Let

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

where

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

A Negative Binomial random variable can be written as a sum of r independent Geometric random variables:

$$X = G_1 + G_2 + \cdots + G_r,$$

where each

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

Each G_i is the waiting time for one additional success.

Negative Binomial Variance Continued

Since

$$X = G_1 + \cdots + G_r$$

and the G_i are independent,

$$\text{Var}(X) = \text{Var}(G_1) + \cdots + \text{Var}(G_r).$$

From the Geometric variance formula,

$$\text{Var}(G_i) = \frac{1-p}{p^2}.$$

Therefore,

$$\text{Var}(X) = r \cdot \frac{1-p}{p^2}.$$

$$\text{Var}(X) = \frac{r(1-p)}{p^2}.$$

Hypergeometric Variance

Let

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

where an urn has w white balls and b black balls.
Let

$$N = w + b.$$

We draw n balls without replacement, and

$$X = \text{number of white balls selected.}$$

Write

$$\boxed{X = I_1 + I_2 + \cdots + I_n,}$$

where

$$I_j = \begin{cases} 1, & \text{if draw } j \text{ is white,} \\ 0, & \text{otherwise.} \end{cases}$$

Hypergeometric Variance Continued

np

$np(1-p)$

Each indicator satisfies

$$\mathbb{E}[I_j] = \frac{w}{N}.$$

So

$$\mathbb{E}[X] = \boxed{n\frac{w}{N}} = n \cdot \frac{w}{w+b}$$

Also,

$$\mathrm{Var}(I_j) = \frac{w}{N}\left(1 - \frac{w}{N}\right).$$

Because sampling is without replacement, the indicators are not independent.
For $i \neq j$,

$$\mathrm{Cov}(I_i, I_j) = -\frac{w}{N}\frac{b}{N}\frac{1}{N-1}.$$

Hypergeometric Variance Formula

Using

$$\mathrm{Var}(X) = \sum_{j=1}^{n} \mathrm{Var}(I_j) + 2\sum_{i<j} \mathrm{Cov}(I_i, I_j),$$

we get

$$\mathrm{Var}(X) = n\frac{w}{N}\frac{b}{N} - n(n-1)\frac{w}{N}\frac{b}{N}\frac{1}{N-1}.$$

Factor:

$$\mathrm{Var}(X) = n\frac{w}{N}\frac{b}{N}\left(1 - \frac{n-1}{N-1}\right).$$

Therefore,

$$\mathrm{Var}(X) = n\frac{w}{N}\frac{b}{N}\frac{N-n}{N-1}.$$

Nonnegativity of Variance

- Sanity Check

$\cancel{-3}$

Proposition

For any r.v. X,

$$\mathrm{Var}(X) \geq 0.$$

Reason:

$$X \approx \mathbb{E}(X) \overset{=0}{\Longleftarrow} \boxed{(X - \mathbb{E}X)}^2 \geq 0. \qquad \mathbb{E}(\geq 0)$$

Thus its expectation is nonnegative.
Later, we will learn the Jensen's Inequality, which implies this Proposition as a special case.
Also,

$$\mathrm{Var}(X) = 0$$

if and only if X is constant with probability 1.

if $Z \geq 0$
$\Rightarrow \mathbb{E}[Z] \geq 0$

Poisson Approximation to Binomial

Poisson Approximation to Binomial

When

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

AND

- n is large;
- p is small;
- $\lambda = np$ is moderate,

Then,

$$\boxed{X \approx \mathrm{Pois}(\lambda).}$$

That is,

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \approx e^{-\lambda}\frac{\lambda^k}{k!}.$$

Why the Poisson Approximation Works

Assume $X \sim \text{Bin}(n,p)$ with $p = \lambda/n$, where $\lambda > 0$ is fixed. For each fixed $k = 0,1,2,\ldots$, we have

$$\mathbb{P}_{Binom(n,p)}(X=k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{n!}{k!(n-k)!}\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}$$

$$= \underbrace{\left[(1)\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right]}_{\longrightarrow 1}\underbrace{\frac{\lambda^k}{k!}}_{\text{stays fixed}}\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\longrightarrow e^{-\lambda}}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\longrightarrow 1}$$

$$\longrightarrow 1\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}\cdot 1 = e^{-\lambda}\frac{\lambda^k}{k!} = \mathbb{P}_{X\sim Poisson(\lambda)}(X=k)$$

Thus,

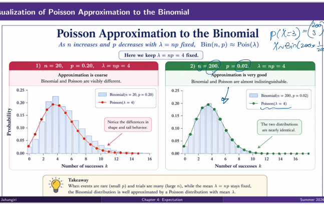
$$\mathbb{P}_{X\sim Bin(n,p)}(X=k) \xrightarrow{n\to\infty} e^{-\lambda}\frac{\lambda^k}{k!}.$$

This phenomenon is formally known as convergence in distribution (or weak convergence) in Probability theory. We will see more with Central Limit Theorem (CLT) and Law of Large Numbers (LLN) in Chapter 10B (Limit Theorems).

Visualization of Poisson Approximation to the Binomial

$P(X=3)=\binom{200}{3}\left(\frac{1}{200}\right)^{3}$

$X\sim Bin(200, \frac{1}{200})$



Poisson Approximation to the Binomial

As n increases and p decreases with $\lambda = np$ fixed, $\text{Bin}(n,p) \approx \text{Pois}(\lambda)$

Here we keep $\lambda = np = 4$ fixed.

Takeaway

When events are rare (small p) and trials are many (large n), while the mean $\lambda = np$ stays fixed, the Binomial distribution is well approximated by a Poisson distribution with mean λ.

Example: Website Visitors

Suppose 1,000,000 people independently decide whether to visit a website.
Each visits with probability

$$p = 2\times 10^{-6}.$$

Let

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

Then

$$X \sim \text{Bin}(1{,}000{,}000, 2\times 10^{-6}).$$

Here

$$\lambda = np = 2.$$

So

$$X \approx \text{Pois}(2).$$

For example,

$$\mathbb{P}(X=0) \approx e^{-2}.$$

Note that it is computationally hard to compute directly from the Binomial PMF:

$$\mathbb{P}(X=k) = \frac{1000000!}{k!(1000000-k)!}\left(\frac{2}{10^6}\right)^k\left(1-\frac{2}{10^6}\right)^{1000000-k} \tag{4}$$

Summary

- Expectation is a probability-weighted average:

$$\mathbb{E}(X) = \sum_x x\mathbb{P}(X=x).$$

- Linearity:

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y),$$

  with no independence required.
- Fundamental bridge:

$$\mathbb{E}(I_A) = \mathbb{P}(A).$$

- Variance measures spread:

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2.$$

- Poisson approximates Binomial when n is large, p is small, and $\lambda = np$ is moderate.

Next lecture

Continuous random variables.

A & B are indep

$$\boxed{P(A|B) = P(A)}$$

$$P(A|A) = P(A)$$

$$P(A|B) = \frac{P(A\cap B)}{P(B)}$$

$$\boxed{P(A\cap B) = P(A)\cdot P(B)}$$

$$\Downarrow$$

$$\underbrace{P(A\cap A)}_{P(A)} = \underbrace{P(A)\,P(A)}_{(P(A))^2}$$

$$\Downarrow$$

$$P(A)^2 - P(A) = 0$$

$$P(A)\left[P(A) - 1\right] = 0$$

$$P(A) = 0 \text{ or } 1$$

$$P(A|A) = \frac{P(A\cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1$$

$$- - - - = \frac{\binom{5}{4}}{\binom{5}{2}}$$