Random Variables

Idea: Often in an random experiment, we are not so much interested in the outcomes themselves, but rather in numerical values that are assigned to these outcomes.

Random variables provide numerical summaries of the experiment. (high dimensional, complex)

eg. exam score, tempreture,

credit score.

Ex Experiment: Toss a coin twice

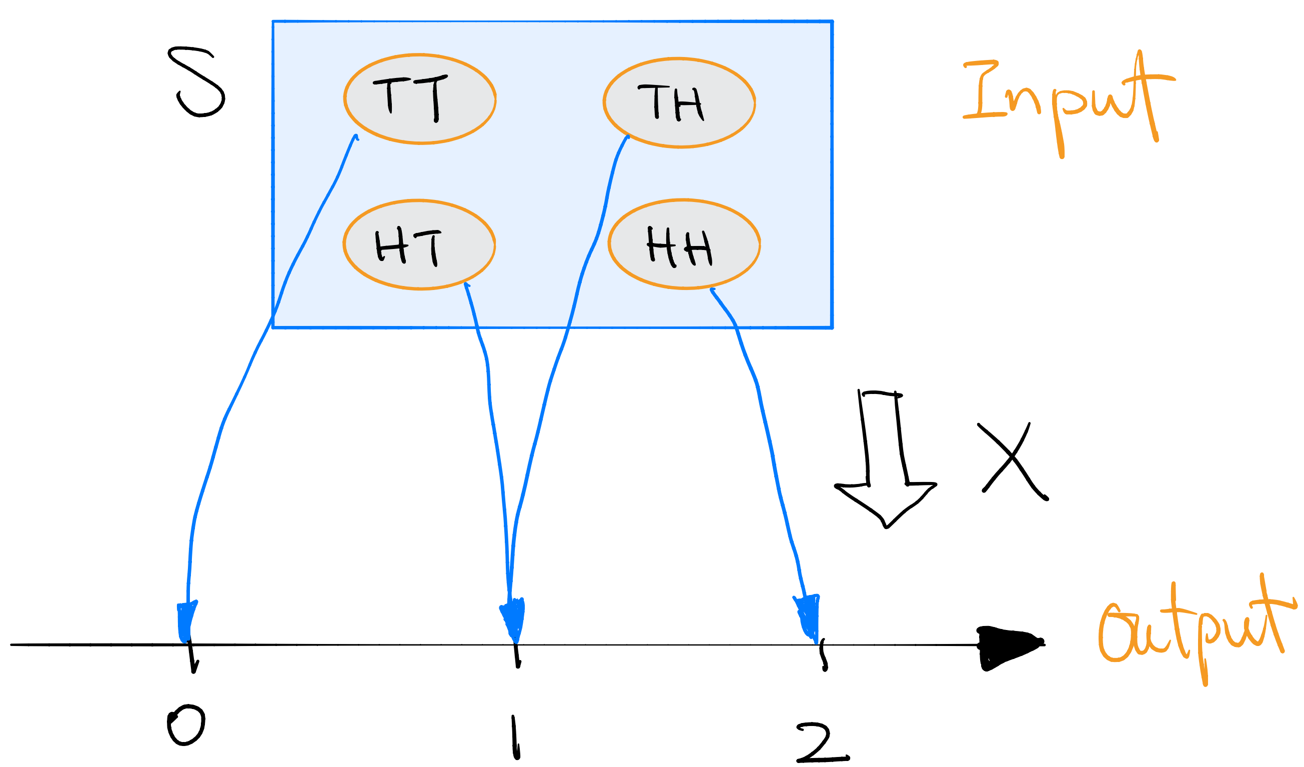
Sample space: $S = \{TT, TH, HT, HH\}$

Let X be the number of heads in the outcome

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

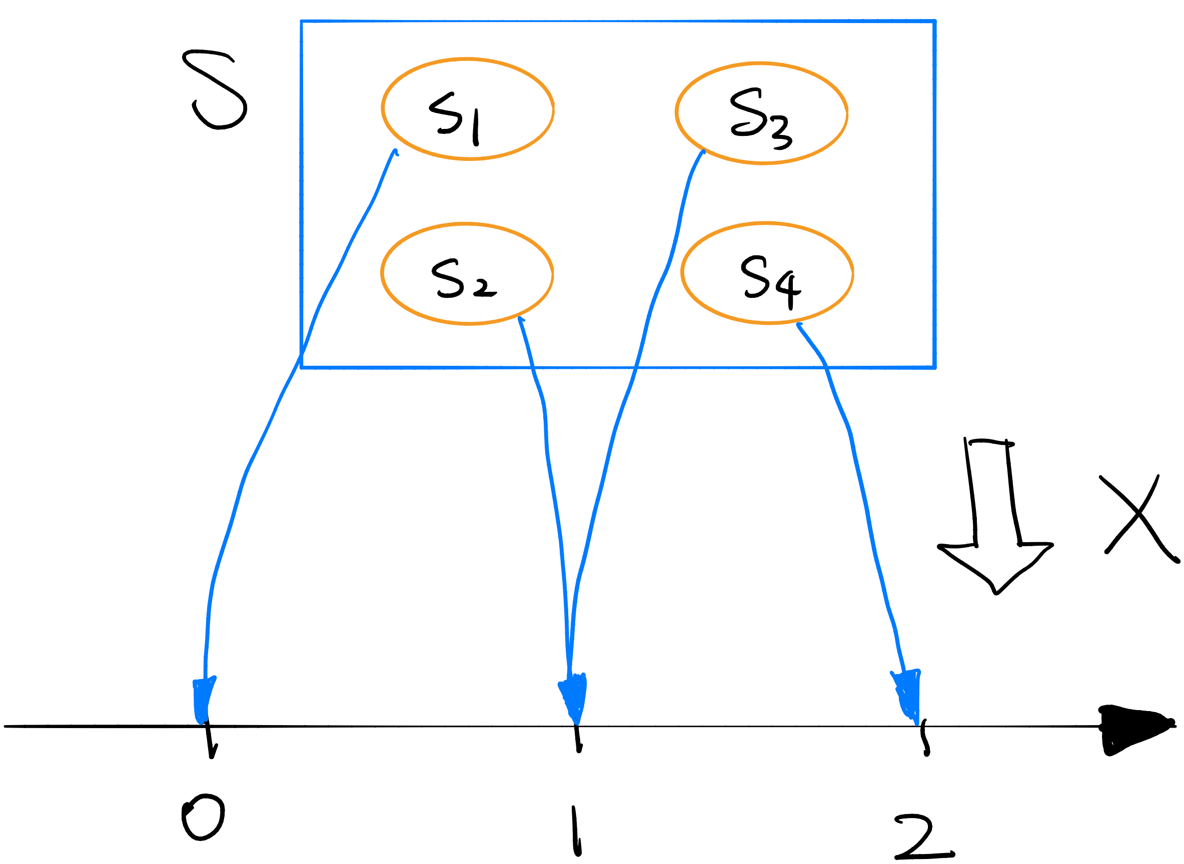
$X(TT) = 0, \; X(TH) = 1, \; X(HT) = 1, \; X(HH) = 2$

Then X can be viewed as a function of outcomes (mapping)



Definition: (Random variable)

Given an experiment with sample space S, a random variable (r.v.) is a function from sample space S to the real number $\mathbb{R}$



Note: mapping itself is deterministic randomness comes from the random experiment.

EX Let Y be the number of tails

$$
\begin{array}{lll}
X(HH)=2, & X(HT)=X(TH)=1, & X(TT)=0 \\
+ & + & + \\
Y(HH)=0, & Y(HT)=Y(TH)=1, & Y(TT)=2 \\
\| & \| & \| \\
2 & 2 & 2
\end{array}
$$

We have $Y(s)=2-X(s)$, for all $s\in S$

Hence $Y=2-X$

EX Let I be 1 if the first toss is Heads and 0 otherwise.

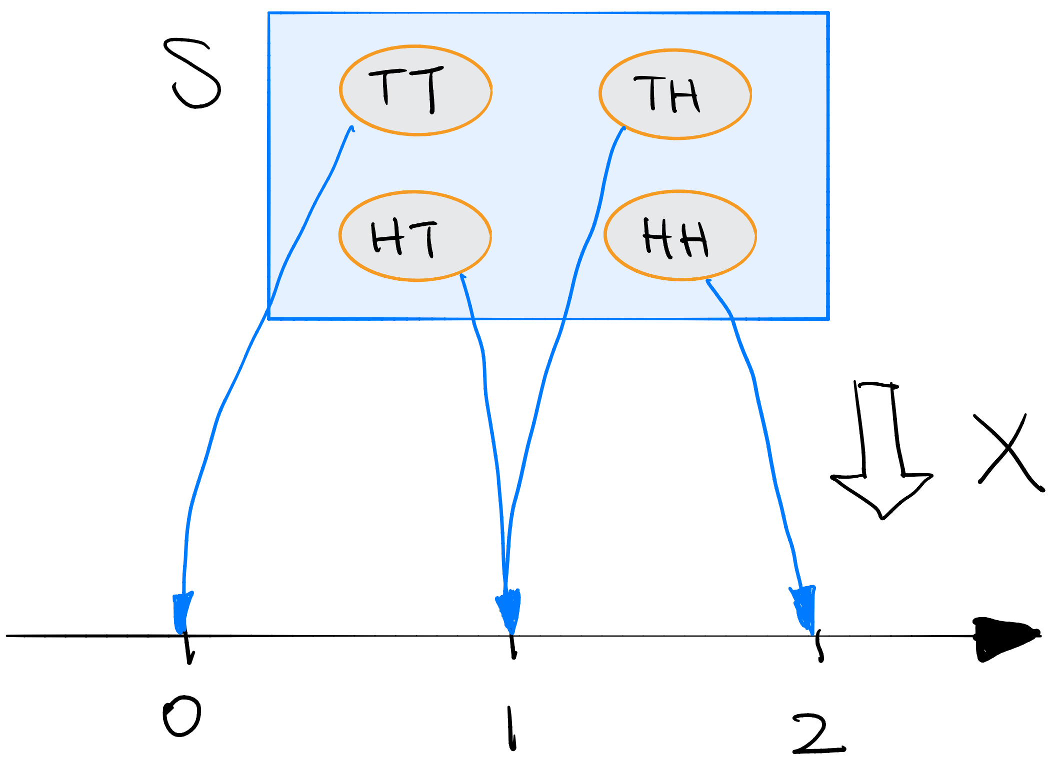
$$I(HH)=1, \quad I(TH)=0,$$

$$I(HT)=1 \quad I(TT)=0$$

Random variables provide numerical summaries of the experiment. (high dimensional, complex)

Use r.v. to represent event

Tossing a coin twice



$X = 0 \iff \{TT\}$

$$X = 1 \iff \{TH, HT\}$$

$$X = 2 \iff \{HH\}$$

Let

X : r.v.

x : realized value.

For a r.v. X, we use $X = x$ to represent an event.

$$X = x \iff \{s \in S . X(s) = x\}$$

This represents all s to which X assigns number x.

e.g. When $X = 1$

$$X = 1 \iff \{s \in S,\ X(s) = 1\}$$
$$\|$$
$$\{HT, TH\}.$$

$$X = 0 \iff \{s \in S,\ X(s) = 0\}$$
$$\|$$
$$\{TT\}$$

$$X = 2 \iff \{s \in S,\ X(s) = 2\}$$
$$\|$$
$$\{HH\}$$

Discrete random variable

Two types of random variables

discrete vs. continuous.

Discrete RV: possible values are countable

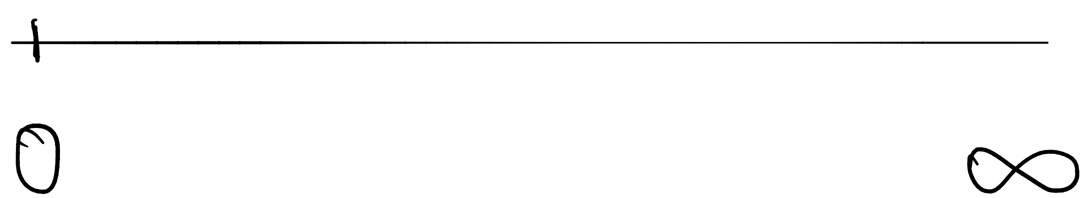
$X = 1, 2, 3,$ (finite) $X = 1, 2, 3, \ldots$ (infinite)

e.g. # of voters, shoe size, # of customers

Definition: A r.v. X is said to be discrete if there is a finite list of values $a_1, a_2, \ldots, a_n$ **or** an infinite list of values $a_1, a_2, \ldots$

Continuous RV: possible values correspond to points on some interval.

e.g. tempreture, Length, weight, height.



Probability Mass Function

Describe the behavior of a r.v. using probability.

$P(\text{Claim} > 1m)$ $\quad$ $P(\text{survival time} > 2\,yr)$

Ex

Tossing a coin twice

X : # of Heads in two tosses.

$X=0 \iff \{TT\}$ $\quad$ $P(X=0) = P(\{TT\}) = \frac{1}{4}$

$X=1 \iff \{HT, TH\}$ $\quad$ $P(X=1) = P(\{HT, TH\}) = \frac{1}{2}$

$X=2 \iff \{HH\}$ $\quad$ $P(X=2) = P(\{HH\}) = \frac{1}{4}$.

$P(X=0)$, $P(X=1)$, $P(X=2)$ specify the probability of all events associated with X

Definition (Probability Mass Function)

The probability mass function (PMF) of a discrete r.v. X is the function p_X given by

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

EX Two coin tosses. X : # of Heads.

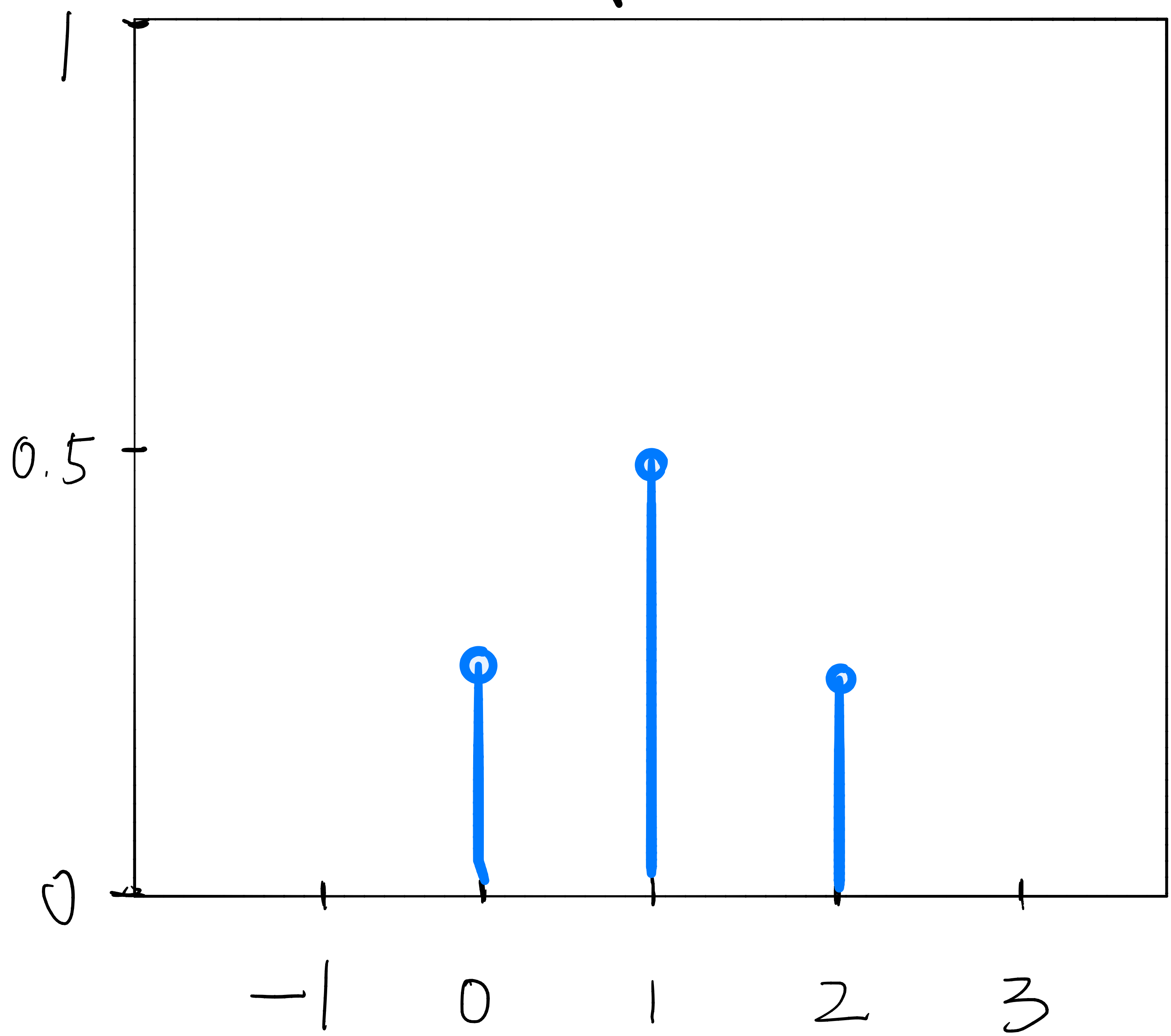
$$p_X(0) = \mathbb{P}(X = 0) = \frac{1}{4}$$

$$p_X(1) = P(X = 1) = \frac{1}{2}$$

$$p_X(2) = \mathbb{P}(X = 2) = \frac{1}{4}.$$

$p_X(0)$ for all other values of x.

PMF of RV X.



Ex Y : # of Tails . $Y = 2 - X$

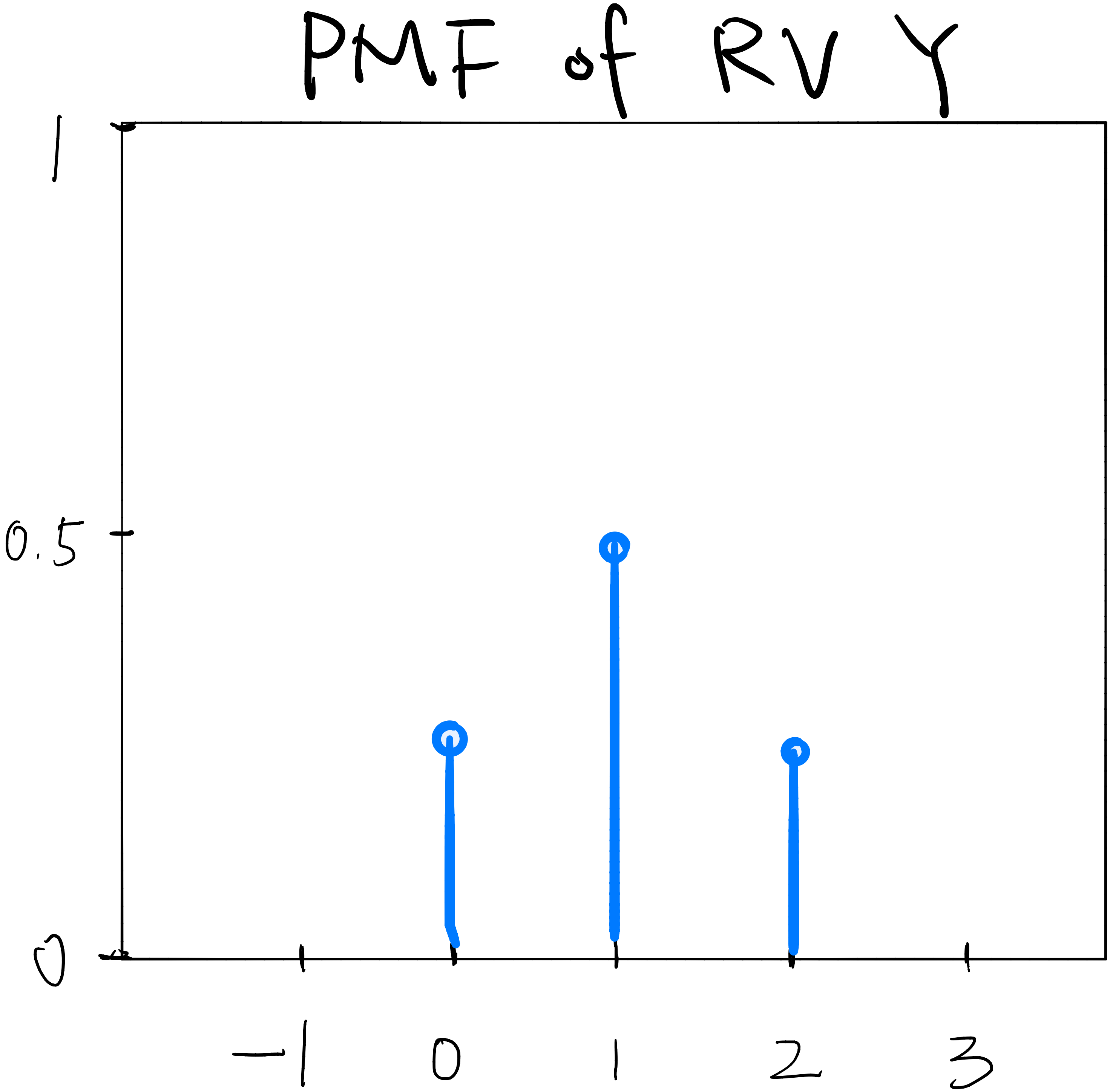
$$p_Y(y) = P(Y = y) = P(2 - X = y)$$
$$= \mathbb{P}(X = 2 - y)$$
$$= p_X(2 - y)$$

$$p_Y(0) = p_X(2-0) = p_X(2) = \frac{1}{4}$$

$$p_Y(1) = p_X(2-1) = p_X(1) = \frac{1}{2}$$

$$p_Y(2) = p_X(2-2) = p_X(0) = \frac{1}{4}$$

Note : X and Y have same PMF even though X and Y are not the same r.v. (X and Y are two different functions


PMF of RV Y
1
0.5
0
-1
0
1
2
3

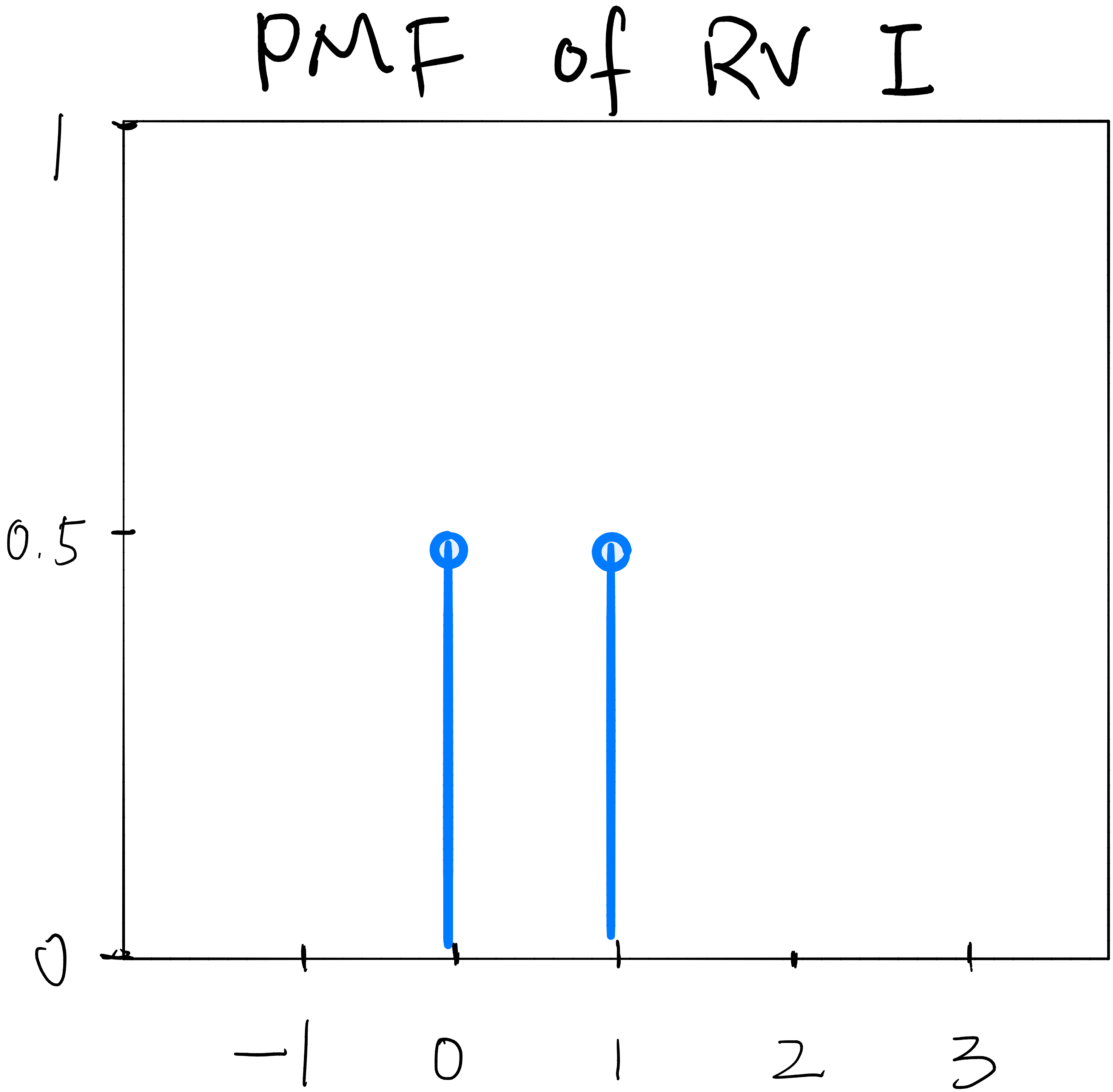
from $\{HH, HT, TH, TT\}$ to the real line.

EX $I = \begin{cases} 1 & \text{if the first toss is Heads.} \\ 0 & \text{otherwise.} \end{cases}$

PMF:

$$p_I(0) = \mathbb{P}(I=0) = \mathbb{P}(\{TH, TT\}) = \frac{1}{2}$$

$$p_I(1) = \mathbb{P}(I=1) = \mathbb{P}(\{HT, HH\}) = \frac{1}{2}.$$


PMF of RV I
1
0.5
0
-1
0
1
2
3

Theorem (Valid PMF)

Let X be a discrete R.V.

The PMF of X must satisfy:

1. $p_X(x) \geq 0$ for all value of x.

2. $\sum_x p_X(x) = 1$ (sum over all possible values of x)

Proof: (1) is true since probability is nonnegative.

$x_1 \quad x_2 \ldots x_K$

For (2), since $\{X = x_j\}$ are disjoint,

$$\sum_{j=1}^{K} P(X = x_j) = P\left(\bigcup_{j=1}^{K} \{X = x_j\}\right)$$

$$= P(X = x_1 \text{ or } X = x_2 \text{ or } X = x_3 \cdots \text{or}) = 1$$

Ex Whether valid or not.

x	0	1	2	3
$p_X(x)$	.2	.3	.3	.2

✓

x	-2	-1	0
$p_X(x)$	.25	.50	.20

✗

x	4	9	20
$p_X(x)$	$-.3$	1.0	.3

✗

x	2	3	5	6
$p_X(x)$	.15	.20	.40	.35

✗

EX

x	-2	-1	0	1	2
$p_X(x)$	.10	.15	.40	.30	.05

✓

(1) $P(X \le 0)$

$= P(X=-2 \text{ or } X=-1) = P(X=-2) + P(X=-1)$

$= 0.1 + 0.15 = 0.25$

(2) $P(X > -1)$

$= P(X=0 \text{ or } X=1 \text{ or } X=2) = P(X=0) + P(X=1) + P(X=2)$

$= .4 + .3 + .05$

$= .75$

(3) $P(-1 \le X \le 1)$

$= P(X=-1 \text{ or } X=0, \text{ or } X=1) = P(X=-1) + P(X=0)$

$+ P(X=1) = .15 + .4 + .3$

$= .85$

(4) $P(X < 2)$

(5) $\mathbb{P}(-1 < X < 2)$

(6) $\mathbb{P}(X < 1)$

EX Suppose that of those people that are infected by Ebola.

10%	infected by Blood
35%	" by urine
40%	" " vomit
15%	" " feces

Assume someone can only be infected by one source.

If someone infected by

blood	5 d.	incubation period.
urine	15 d.	" "
vomit	10 d.	" "
feces	10 d.	" "

(1) Find the PMF for incubation period.

X - incubation period $\quad p_X(x) = P(X = x)$

(2) If someone infected has no symptom up to and including 5 days, what is the

$$P(10 \le X < 15 \mid X \ge 5)$$

probability that have symptoms between 10 and 15 days (excluding 15 days)

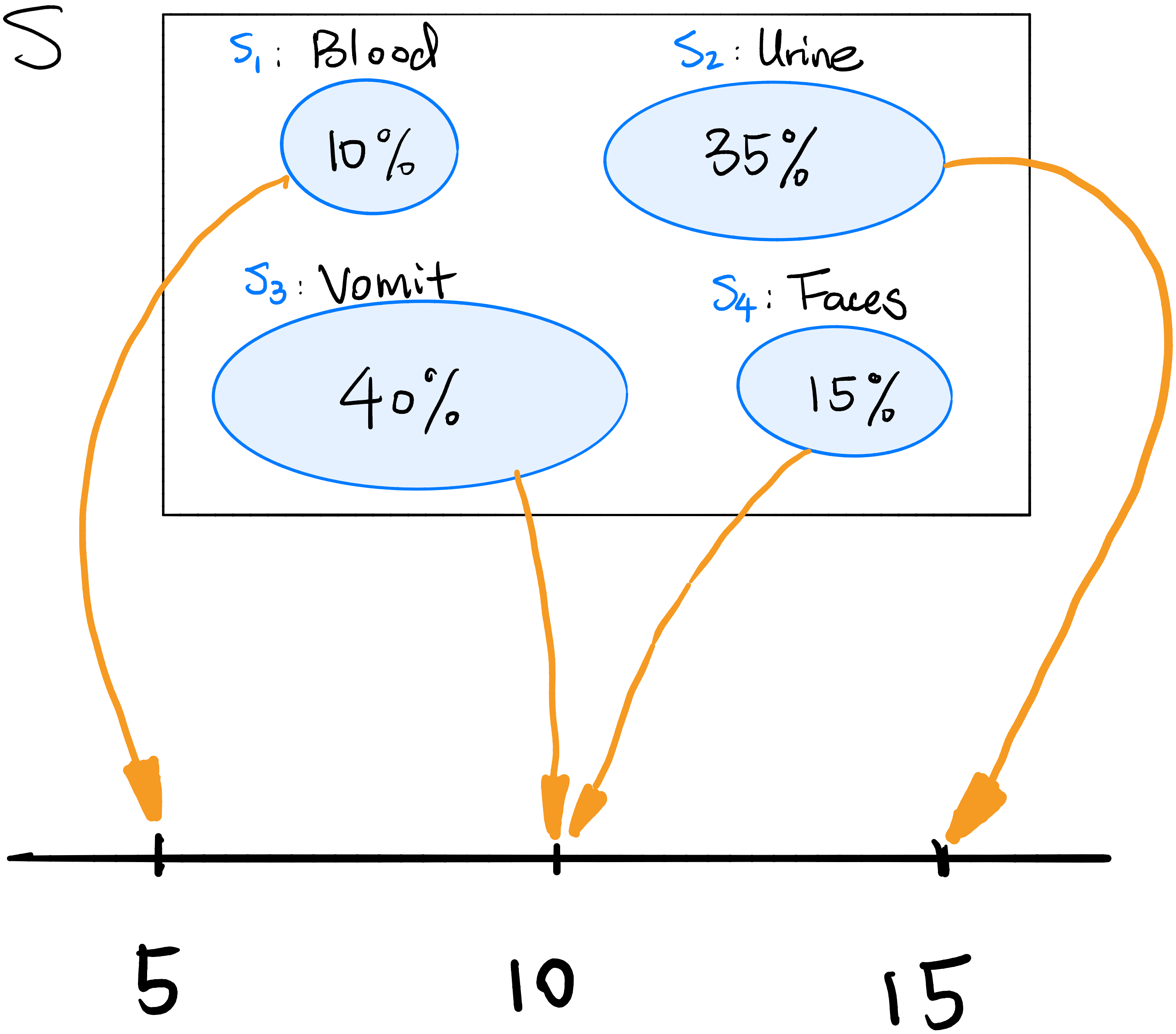
Solution:

(1) $S = \{ \underset{S_1}{\text{Blood}}, \underset{S_2}{\text{Urine}}, \underset{S_3}{\text{Vomit}}, \underset{S_4}{\text{Faces}} \}$

The associated probabilities

$P(S_1) = 0.1$ $\quad$ $P(S_2) = 0.35$

$P(S_3) = 0.4$ $\quad$ $P(S_4) = 0.15$



Incubation period is a random variable defined on this sample space, denoted by X

$$X(s_1) = 5 \qquad X(s_2) = 15$$

$$X(s_3) = 10 \qquad X(s_4) = 10$$

$$p_X(5) = P(X=5) = P(X(s_1)=5)$$

$$= P(s_1) = 0.1$$

$$p_X(10) = P(X=10) = P(X(\{s_3 \text{ or } s_4\})=10)$$

$$= 0.4 + 0.15 = P(s_3 \text{ or } s_4)$$

$$= 0.55 \qquad = P(s_3) + P(s_4)$$

$$p_X(15) = P(X=15) = P(X(s_2)=15)$$

$$= P(s_2) = 0.35$$

x	5	10	15
$p_X(x)$	0.1	0.35	0.55

(2) $$P(\overbrace{10 \le X < 15}^{B} \mid \overbrace{X > 5}^{A})$$

$$= \frac{P(\overbrace{\{10 \le X < 15\} \cap \{X > 5\}}^{B \cap A})}{P(\underbrace{X > 5}_{A})} \quad (*)$$

Note: If $\underbrace{10 \le X < 15}_{B}$ happens then

$\underbrace{X > 5}_{A}$ must happen. If $s \in B$ then $s \in A$.

This implies $B \subset A$, thus $B \cap A = B$

$$B \cap A = \{10 \le X < 15\} \cap \{X > 5\} = \{\underbrace{10 \le X < 15}_{B}\}$$

Hence

$$P(B \cap A) = P(B)$$

Hence

$$(*) \quad = \frac{P(B)}{P(A)}$$

$$= \frac{P(10 \le X < 15)}{P(X > 5)}$$

$$= \frac{P(X=10)}{P(X=10 \cup X=15)}$$

$$= P(X=10) + P(X=15)$$

$$= \frac{0.55}{0.55 + 0.35}$$

$$= \frac{55}{90}$$

For discrete R.V. knowing its PMF determines its distribution.

Bernoulli RV

Some certain RV's are commonly used

Definition: An RV X is said to have the Bernoulli distribution if X has only two possible values, 0 and 1, and with PMF

$$p_X(1) = P(X=1) = p$$

$$p_X(0) = P(X=0) = 1-p \quad , (0 \leq p \leq 1)$$

where p is a parameter. We write this as

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

"$\sim$" reads as "is distributed as".

EX (1) Toss a fair coin once

$$S = \{ \text{Head}, \text{Tail} \}$$

$$P(\text{Head}) = \frac{1}{2} = P(\text{Tail})$$

Define X : number of Head . 0 and 1

PMF : $P(X=0) = P(X=1) = 1/2$

$$X \sim \text{Bern}(p), \quad p = \frac{1}{2}$$

(2) If it is an biased coin

$$P(\text{Head}) = 2/3 \quad P(\text{Tail}) = 1/3$$

PMF : $P(X=1) = 2/3 \quad P(X=0) = 1/3$

$$X \sim \text{Bern}(p), \quad p = \frac{2}{3}$$

(3) Same biased coin, but now let

Y be the number of Tail,

$$Y = 1 - X$$

$$P(Y=0) = P(1-X=0) = P(X=1) = 2/3$$

$$P(Y=1) = P(1-X=1) = P(X=0) = 1/3$$

$$Y \sim \text{Bern}(p), \quad p = \frac{1}{3}$$

Bernoulli trial

An experiment that can result in either success or

failure (but not both) is called a Bernoulli trial . A Beroulli RV can be thought of as the indicator of success in a Bernoulli trial :

$$X = \begin{cases} 1 & \text{if sucess w prob } p \\ 0 & \text{if fail w prob } 1-p \end{cases}$$

$P(X=1) = p \quad P(X=0) = 1-p$

X is a Bernoulli RV which indicate Sucess / failure of this Bernoulli trial.

Ex Tossing a coin once.

Success - Head - 1

Failure - Tail - 0

Binomial RV

* Tossing a coin 4 times. $P(\text{Head}) = P(\text{Tail}) = \frac{1}{2}$

  $P(\#\text{ of Heads in 4 tosses} = 2) = ?$

* A family with 7 children.

  $P(\text{Boy}) = 0.51$

  $P(\text{Girl}) = 0.49$

  $P(\underbrace{\#\text{ of Boys}}_{=\,X\text{ : binomial RV}} = 5) = ?$

Definition (Binomial RV)

Suppose that

1. There are n identical Bernoulli trials.
2. The trials are independent.
3. Each trial has the same success probability p, and failure probability $1-p$.
4. Let X be the number of successes in n trials.

X is called the Binomial random variable with parameter n and p. We write

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

Note: $n \geq 1 \qquad 0 < p < 1$

Theorem (Binomial PMF)

If $X \sim \text{Bin}(n, p)$, then PMF of

X is $P(\,k$ successes in n trials$)$

$$p_X(k) = \overbrace{P(X=k)}^{\parallel} = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0, 1, 2, \dots, n$.

Let's check, whether $p_X(k)$ is indeed PMF

(1) $p_X(k) \geq 0$ for $k = 0, 1, \dots n$

(2) $\displaystyle\sum_{k=0}^{n} p_X(k) = 1$

$$= \sum_{k=0}^{n} \binom{n}{k} p^k \underbrace{(1-p)}_{q}^{n-k} = (p+q)^n$$

$$= 1^n = 1$$

e.g.

$$(p+q)^2 = \overset{\binom{2}{0}}{q^2} + \overset{\binom{2}{1}}{2pq} + \overset{\binom{2}{2}}{p^2} = \sum_{k=0}^{2} \binom{2}{k} p^k q^{2-k}$$

$$(p+q)^3 = \overset{\binom{3}{0}}{q^3} + \overset{\binom{3}{1}}{3pq^2} + \overset{\binom{3}{2}}{3p^2q} + \overset{\binom{3}{3}}{p^3} = \sum_{k=0}^{3} \binom{3}{k} p^k q^{3-k}$$

EX Tossing a coin twice

X : # of Heads

2 trials. $n = 2$, $p = P(\text{Head}) = P(\text{Tail}) = \frac{1}{2}$

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

PMF: $p_X(0) = P(X = 0)$

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

$$\binom{2}{0} = \frac{2!}{0!(2-0)!} = \frac{2!}{1 \cdot 2!} = 1$$

$$= \binom{2}{0}\left(\frac{1}{2}\right)^0\left(1-\frac{1}{2}\right)^{2-0}$$

$$= 1 \cdot 1 \cdot \frac{1}{4}$$

$$= \frac{1}{4}$$

$$p_X(1) = P(X=1)$$

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

$$= \binom{2}{1}\left(\frac{1}{2}\right)^1\left(1-\frac{1}{2}\right)^{2-1}$$

$$= 2 \cdot \frac{1}{2} \cdot \frac{1}{2}$$

$$= \frac{1}{2}$$

$$p_X(2) = P(X=2)$$

$$= \binom{n}{2} p^2 (1-p)^{n-2}$$

$$= \binom{2}{2}\left(\frac{1}{2}\right)^{2}\left(1-\frac{1}{2}\right)^{2-2}$$

$$= 1 \cdot \frac{1}{4} \cdot 1$$

$$= \frac{1}{4}$$

$$X \sim Bin\left(2, \frac{1}{2}\right)$$

k	0	1	2
$P_X(k)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$

Ex + proof: (Binomial PMF Theorem).

Suppose that for each student, the probability to pass the course test is is 60%, Suppose that 4 students are randomly selected and each takes the test,

X : # of 4 students who pass the test.

(a). P(none of 4 passes the test)

i.e. $P(X=0)$

(b) P(three of 4 pass the test)

i.e. $P(X=3)$

(c). Derive the PMF for RV X

We know $n=4$, $p=0.6$

$X \sim Bin(n, p)$

$X=0$	$X=1$	$X=2$	$X=3$	$X=4$
FFFF	SFFF	FFSS	FSSS	SSSS
	FSFF	FSFS	SFSS	
	FFSF	FSSF	SSFS	
	FFFS	SFFS	SSSF	
		SFSF		
		SSFF		
		$\frac{4!}{2!(4-2)!}=6$		
$\binom{4}{0}=1$	$\binom{4}{1}=4$	$\binom{4}{2}=6$	$\binom{4}{3}=4$	$\binom{4}{4}=1$

(a) $P(X=0)$

$$= P(FFFF)$$

$$= P(F)\times P(F)\times P(F)\times P(F)$$

$$= (1-p)(1-p)(1-p)(1-p)$$

$$= (1-p)^4$$

(b) $P(X=3) = P(FSSS \cup SFSS \cup SSFS \cup SSSF)$

$= P(FSSS) + P(SFSS) + P(SSFS) + P(SSSF)$

$= (0.6)^3(0.4)^1 + (0.6)^3(0.4)^1 + (0.6)^3(0.4)^1 + (0.6)^3(0.4)^1$

$= \overset{\binom{4}{3}}{4} \times (0.6)^3 \times (0.4)^1$

$= (\text{\# of pts. for which } X=3) \times (0.6)^{\text{\# of } S} \times (0.4)^{\text{\# of } F}$

$= \begin{pmatrix} \text{\# of ways of selecting} \\ 3\ S \text{ in the 4 trials} \end{pmatrix} \times p^3 \times (1-p)^1$

$= \binom{4}{3} \times p^3 \times (1-p)^1$

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

$n=4$, $p=0.6$, $k=3$

$= p_X(3)$

Therefore, in general.

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

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

If $n=4$, $p=0.6$

$$P_X(0) = P(X=0) = \binom{4}{0} 0.6^0\, 0.4^{4-0}$$

$$= \frac{4!}{0!\,(4-0)!}\, 0.6^0 \cdot 0.4^4 = 1 \cdot (0.6)^0 (0.4)^4$$

$$P_X(1) = P(X=1) = \binom{4}{1} 0.6^1\, 0.4^{4-1}$$

$$= \frac{4!}{1!\,(4-1)!}\, 0.6^1 \cdot 0.4^3 = 4 \cdot (0.6)^1 \cdot (0.4)^3$$

$$P_X(2) = P(X=2) = \binom{4}{2} 0.6^2\, 0.4^{4-2}$$

$$= \frac{4!}{2!\,(4-2)!}\, 0.6^2 \cdot 0.4^2 = 6 \cdot (0.6)^2 \cdot (0.4)^2$$

$$P_X(3) = P(X=3) = \binom{4}{3} 0.6^{3}\, 0.4^{4-3}$$

$$= \frac{4!}{3!(4-3)!} 0.6^{3} \cdot 0.4^{1} = 4 \cdot (0.6)^{3} \cdot (0.4)^{1}$$

$$P_X(4) = P(X=4) = \binom{4}{4} 0.6^{4}\, 0.4^{4-4}$$

$$= \frac{4!}{4!(4-4)!} 0.6^{4} \cdot 0.4^{0} = 1 \cdot (0.6)^{4} \cdot (0.4)^{1}$$

$$= (0.6)^{4}$$