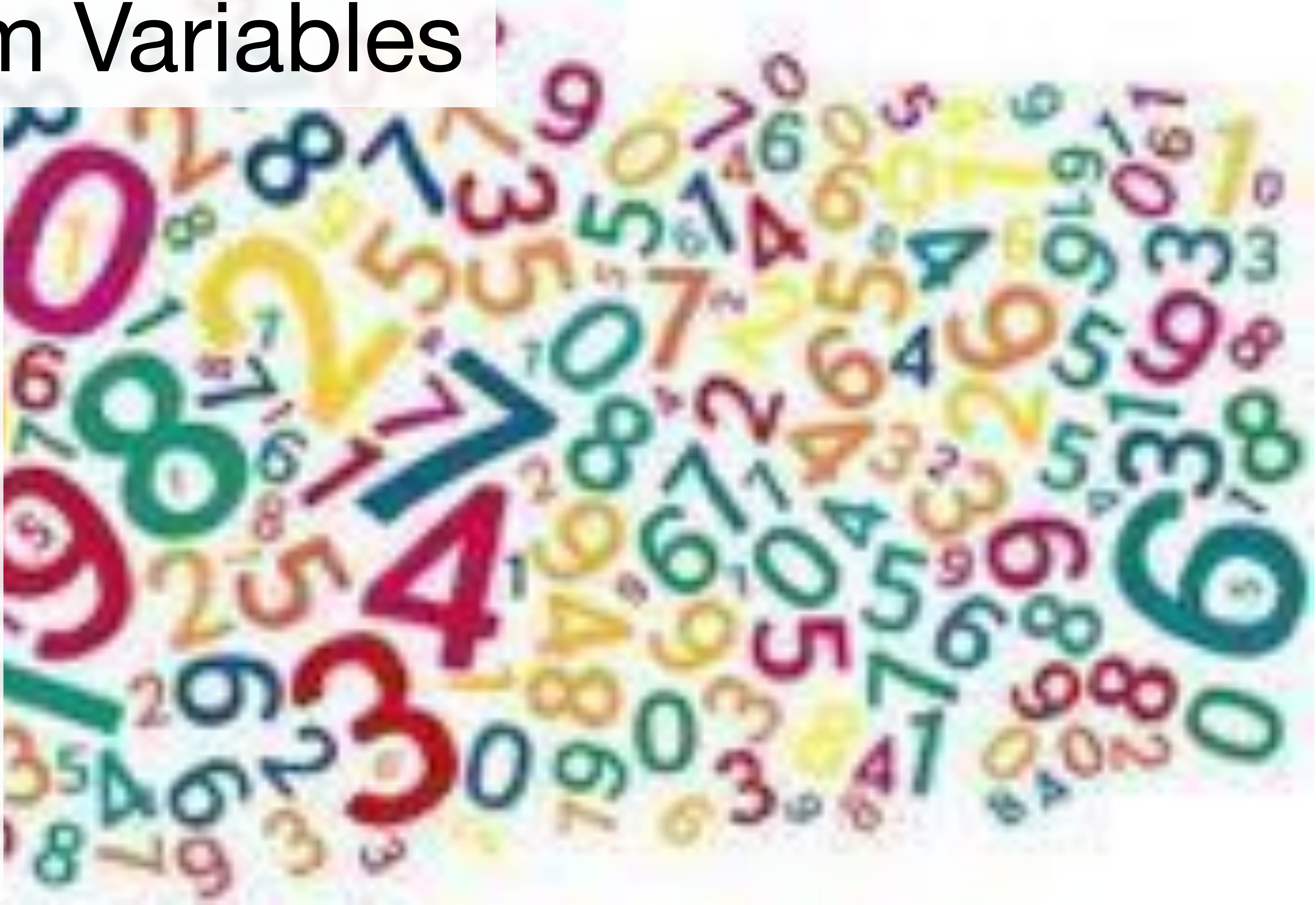
Random Variables

Motivation

Examples

Definitions

Visualization



What Matters

So far coins, dice, cards, dominoes, marbles, ….

Often subscribers clicks viewers yield weight sales

time congestion delay age temperature heart rate

GPA tuition assignment income cost

Numbers!

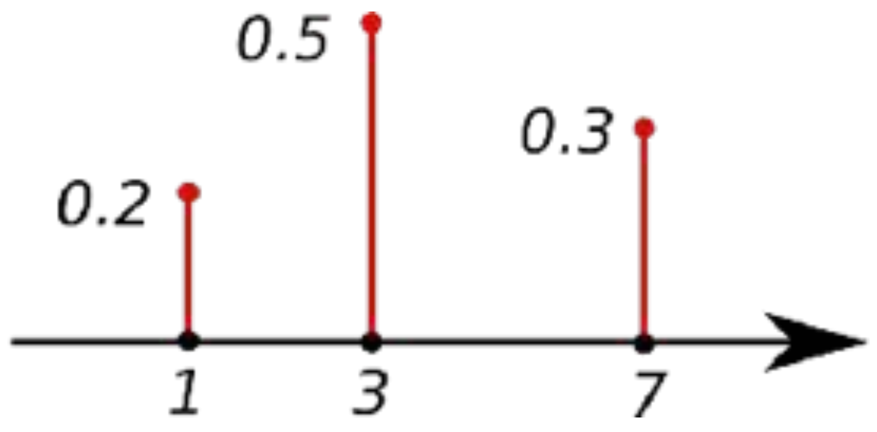
Random variable Number-valued random outcome

Xtra with Numbers

Distri-
bution

$p(x)$

View on a line

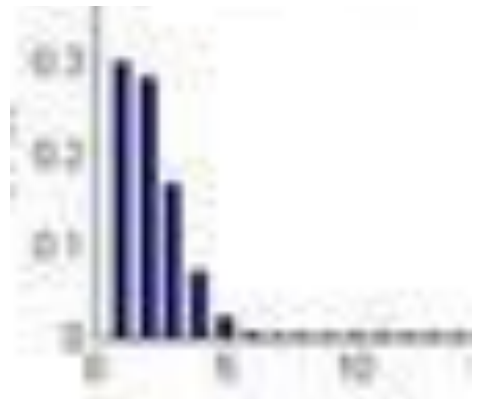


Express as function

$p(x)=1/x^2$

Consider properties

Decreasing



concentrated

Random
Variable

X

Perform operations

$X+1$ X^2

Combine variables

$X+Y$

Consider properties

average value of X

Two Types

Size of sample space Ω

Ω is finite $\{1, 2, 3\}$ $\{e, \pi\}$ or countably infinite $\mathbb{N}$ $\mathbb{Z}$ Discrete

Ω is uncountably infinite $[0,2]$ $(-1,3) \cup [4,5)$ $\mathbb{R}$ Continuous

Combination $[0,2] \cup \{e, \pi\}$ Mixed

Begin with discrete

Been There

Several past examples had number outcomes

Outcome of a die roll {1,…,6}

Number of heads in 3 coin tosses {0,…,3}

Values of a domino tile {0,…,6}

Did not use numerical features → Use extensively

Familiar and new examples General formulation

Heads

3 fair coins

Ω = { ttt, tth, tht, thh, htt, hth, hht, hhh }

$|\Omega| = 8$

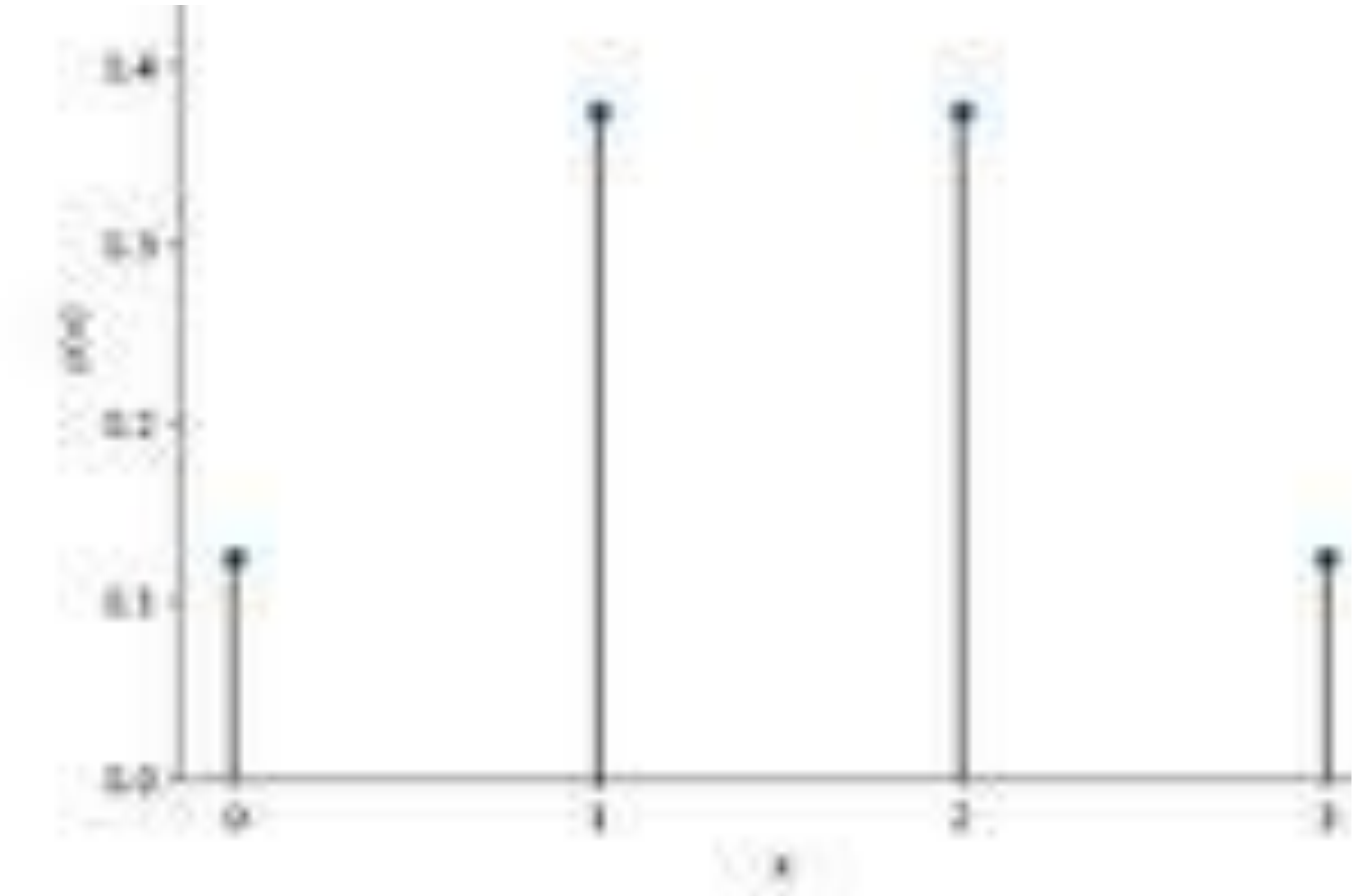
Equiprobable

$p = 1/8$

X

heads

x	Outcomes	p(x)
0	ttt	⅛
1	tth, tht, htt	⅜
2	thh, hth, hht	⅜
3	hhh	⅛



Specification

As before

Explicit $p(1)=.1 \quad p(2)=.2 \quad p(3)=.3 \quad p(4)=.4$

Table

x	1	2	3	4
p(x)	.1	.2	.3	.4

With numbers

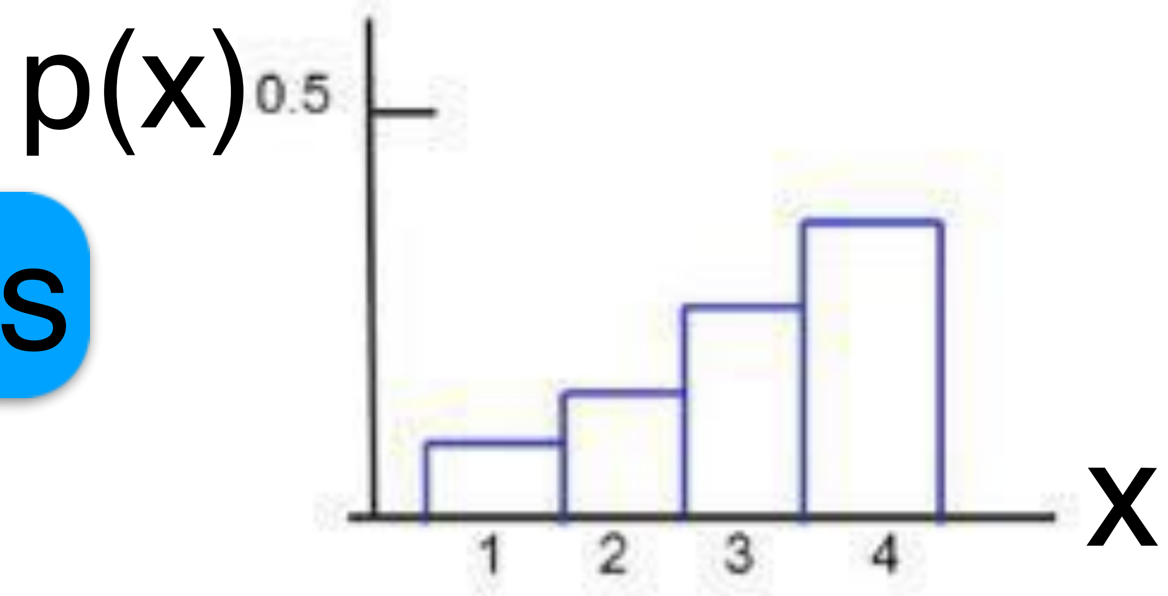
Function $p(x) = x / 10$ $\quad x \in \{1,2,3,4\}$

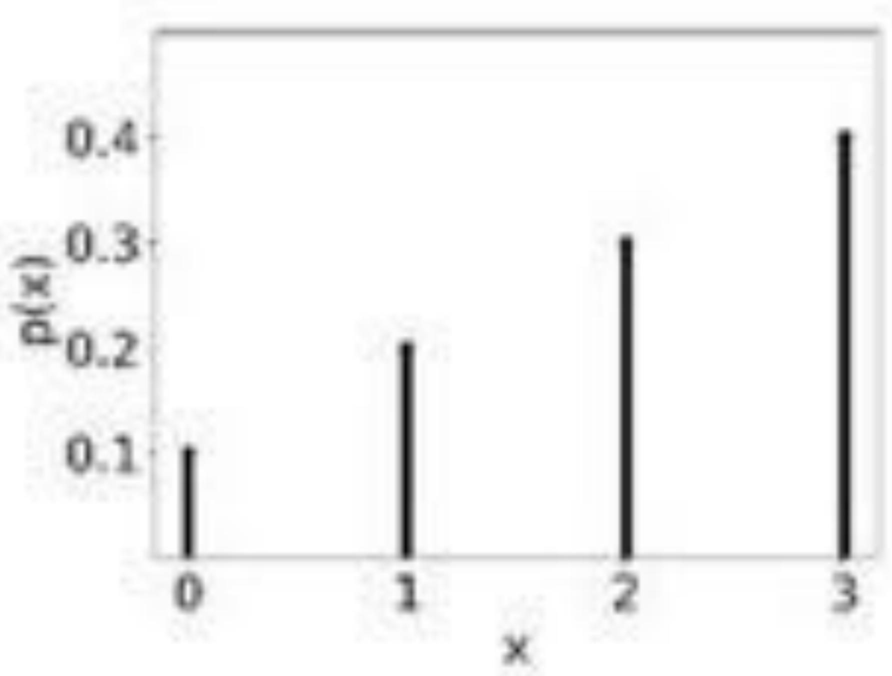
Histogram

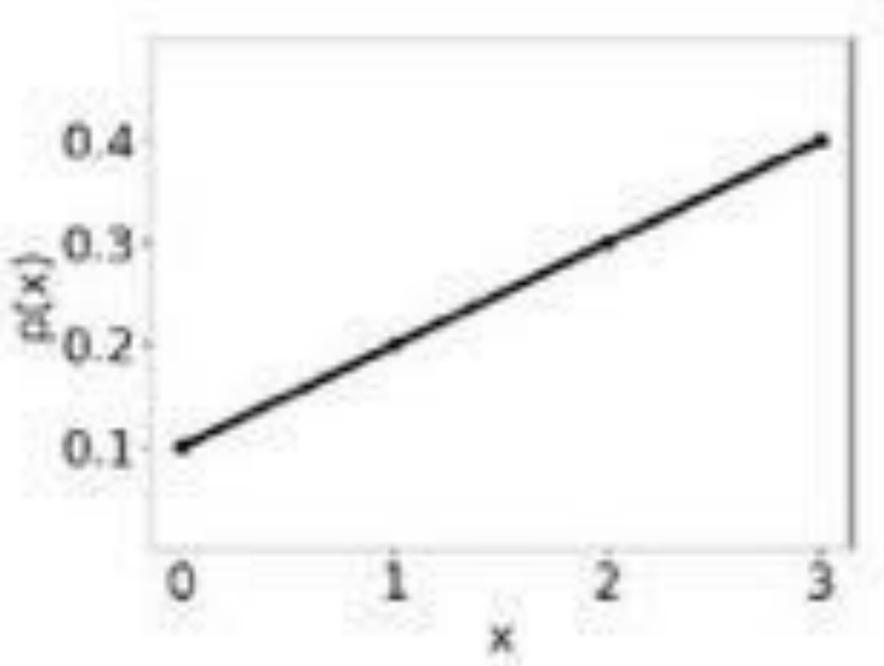
Stem plot

Plot

Graphs







Probability Mass Function

As before

pmf

p : $\Omega \to \mathbb{R}$

Specify Ω and p

Ω

Random variable → $\subseteq \mathbb{R}$

Discrete → finite or countably infinite

p

$p(x) \geq 0 \quad \forall x \in \Omega$

$$\sum_{x \in \Omega} p(x) = 1$$

If X is distributed according to p, we write X ~ p

Alternative Notation

Discrete

$\Omega \subseteq \mathbb{R}$ Often $\mathbb{Z}$ $\mathbb{N}$ $\mathbb{P}$ $\{1,\ldots n\}$

$p(x) \rightarrow p_x$ p_i

$p_i \geq 0$ $\sum_i p_i = 1$

Types of Discrete Distributions

Finite $|\Omega| = n \in \mathbb{P}$

Infinite $|\Omega| = \infty = \aleph_0$

Finite Distributions

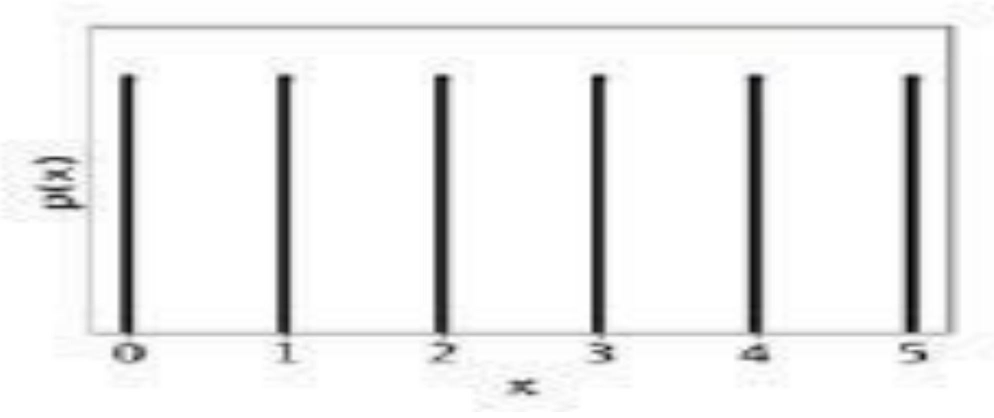
$|\Omega| = n$

Specify pmf

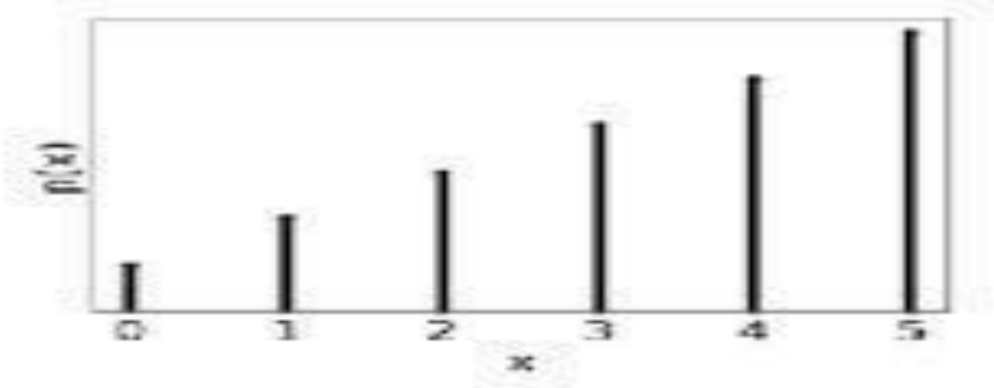
$p_1, p_2, \ldots, p_n$

$\forall_{1 \le i \le n} \quad p_i \ge 0 \qquad \sum_{i=1}^{n} p_i = 1$

Uniform $p_1 = p_2 = \ldots = p_n$ $= 1/n$



Increasing $p_1 \le p_2 \le \ldots \le p_n$



Decreasing $p_1 \ge p_2 \ge \ldots \ge p_n$

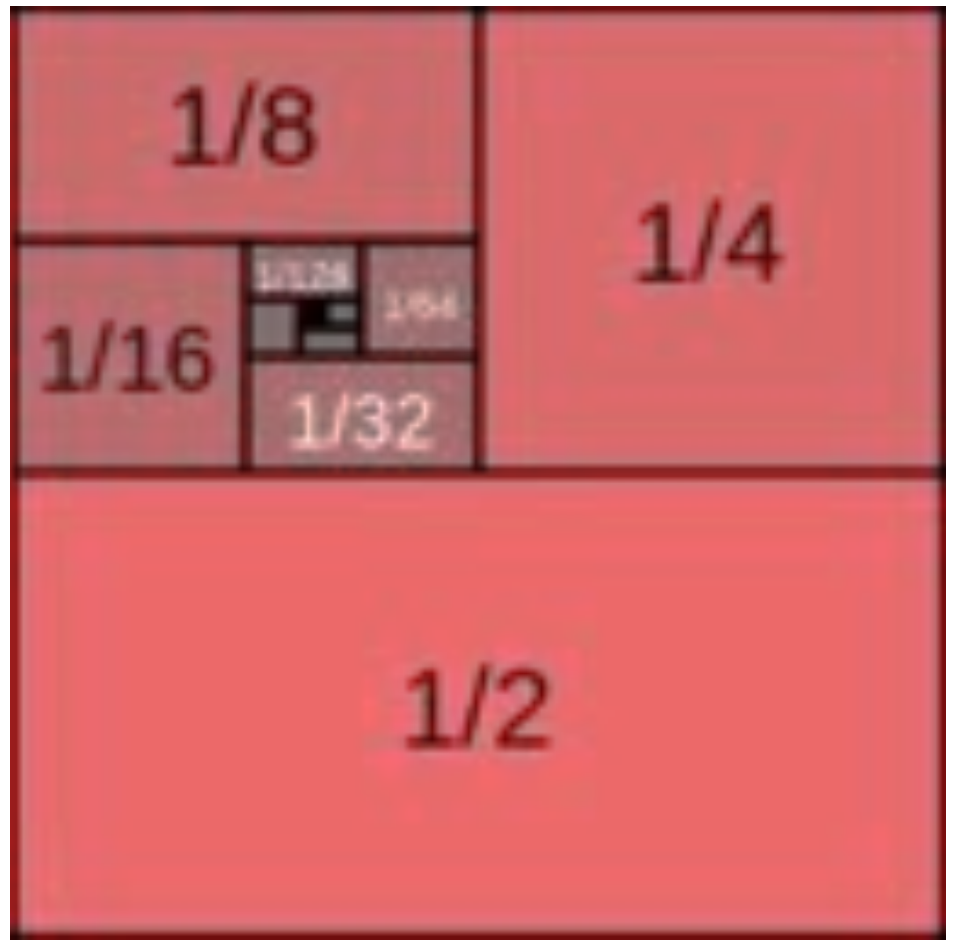
Infinite Distributions

$|\Omega| = \infty$

One-sided infinite $p_1, p_2, p_3, \ldots$

Cannot be uniform $p = 0 \rightarrow \sum = 0$ $p > 0 \rightarrow \sum = \infty$

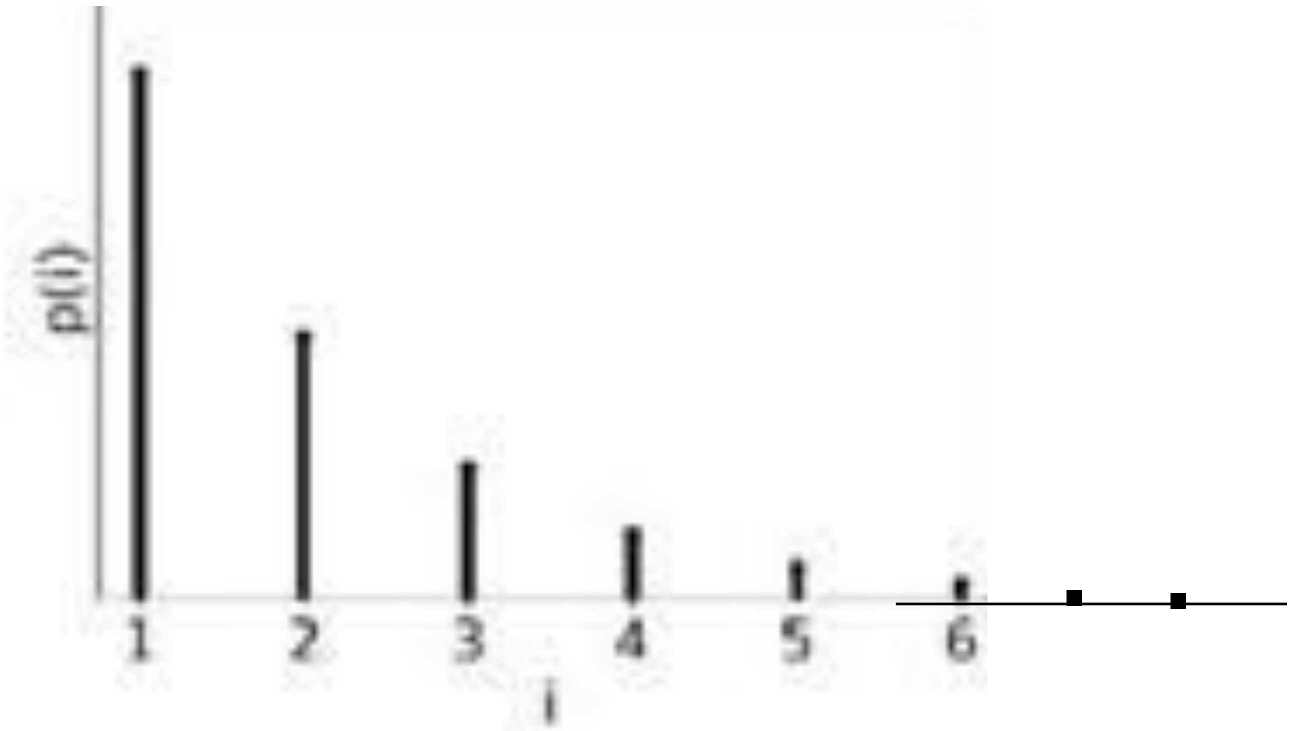
Cannot increase $p_i > 0 \rightarrow p_{i+1}, p_{i+2}, \ldots > 0 \rightarrow \sum = \infty$



Can decrease $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$

$$\sum_{i=1}^{n} \frac{1}{2^i} = 1 - \frac{1}{2^n}$$

$$\sum_{i=1}^{\infty} \frac{1}{2^i} = 1$$



Doubly infinite $\ldots, p_{-2}, p_{-1}, p_0, p_1, p_2, \ldots$

$\ldots, \frac{1}{8}, \frac{1}{4}, 0, \frac{1}{4}, \frac{1}{8}, \ldots$

Formal Definition

Random variable is a mapping $f : \Omega \rightarrow \mathbb{R}$

Simplify terminology, focus on math

Number-valued random experiment

Random Variables

Motivation

Numbers

Operations

Examples

Definitions

pmf

Visualization

Histogram

Plot

Stem

Cumulative Distribution Functions

Areas of Interest

For random varialbe, often, interested in probability of intervals

Temperature between 20 and 80

Salary > 80K

GPA < 3.0

One function helps determine all interval probabilities

Cumulative Distribution Function

Probability mass function (pmf)

p: $\Omega \rightarrow \mathbb{R}$

Cumulative distribution function (cdf)

F: $\mathbb{R} \rightarrow \mathbb{R}$

$$F(x) \stackrel{\text{def}}{=} P(X \in (-\infty, x])$$

$$\stackrel{\text{def}}{=} P(X \leq x)$$

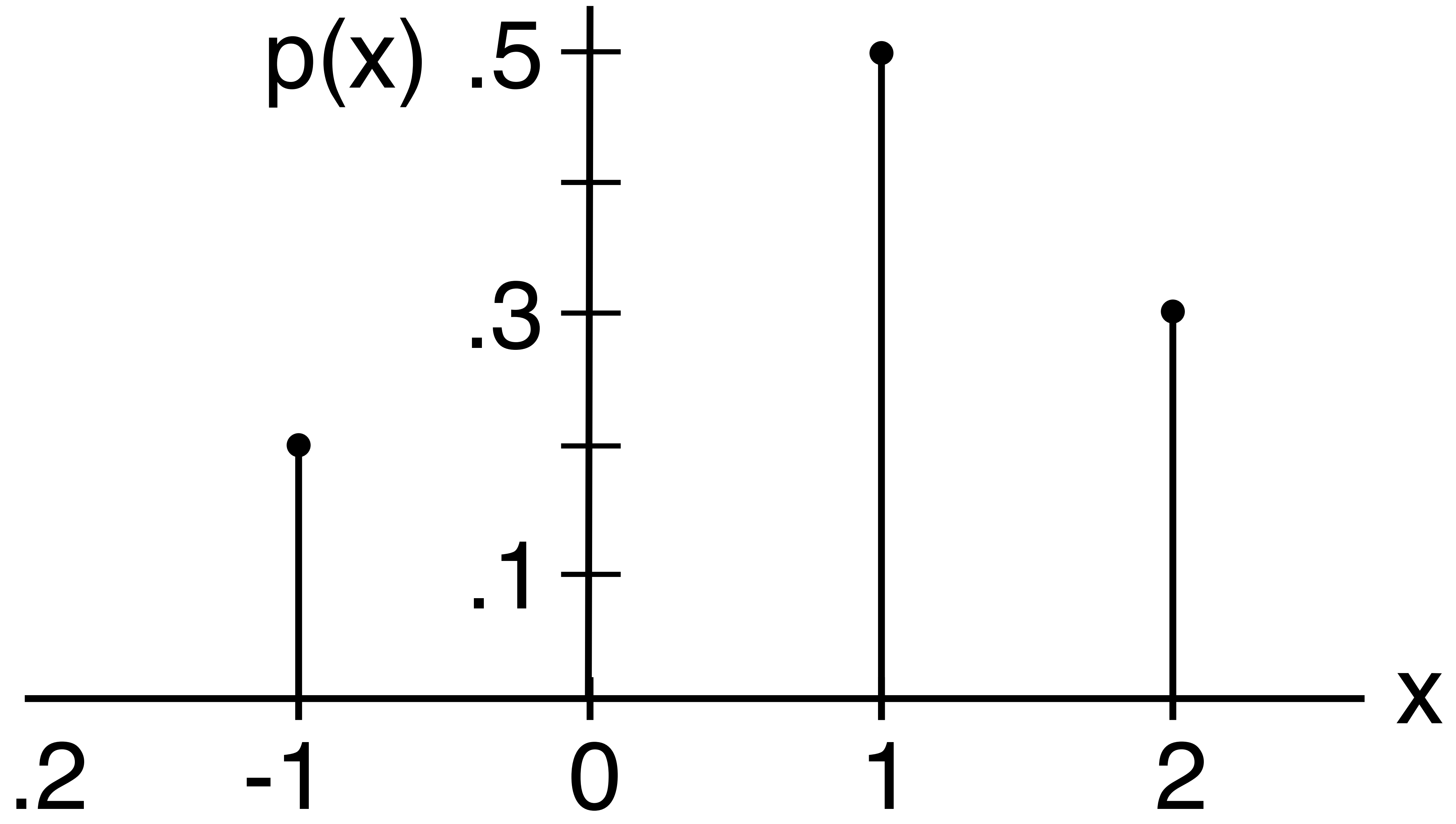
$$= \sum_{u \leq x} p(u)$$

X discrete, still F defined over $\mathbb{R}$

Example

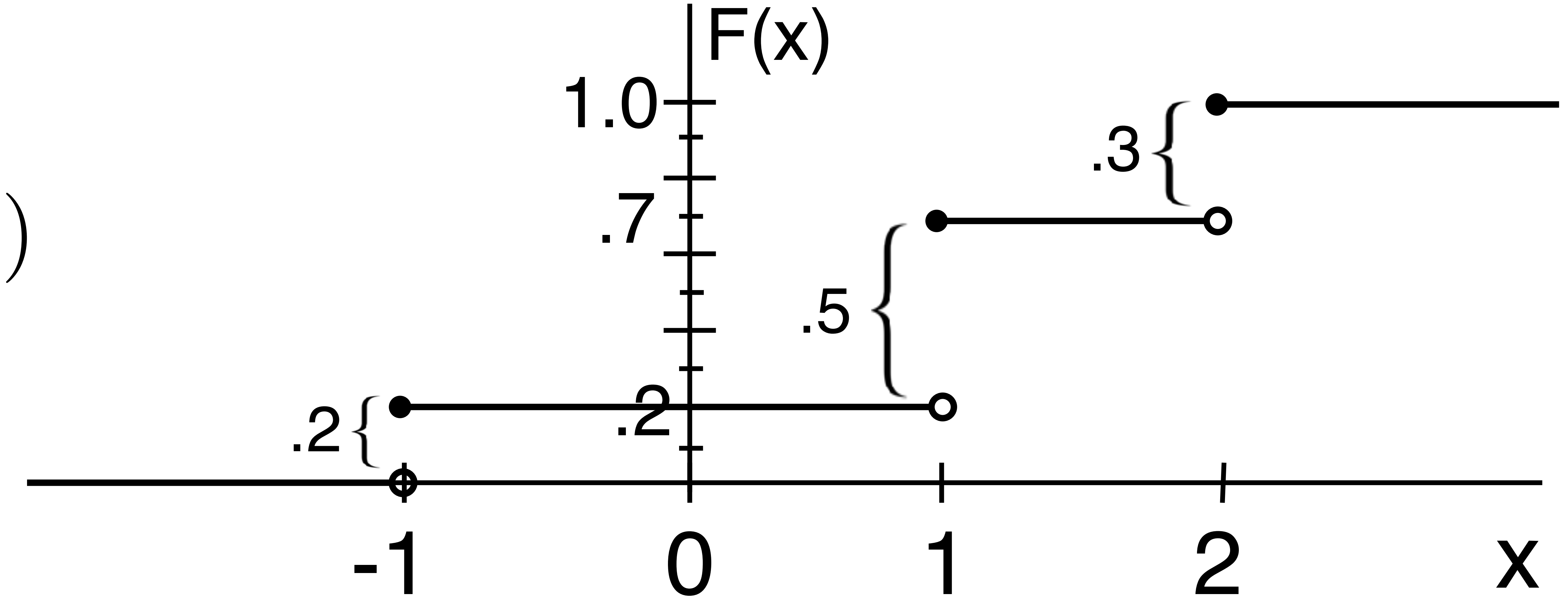
PMF

$$p(x) = \begin{cases} .2 & -1 \\ .5 & 1 \\ .3 & 2 \end{cases}$$



CDF

$$F(x) = P(X \leq x)$$
$$= \sum_{u \leq x} p(u)$$

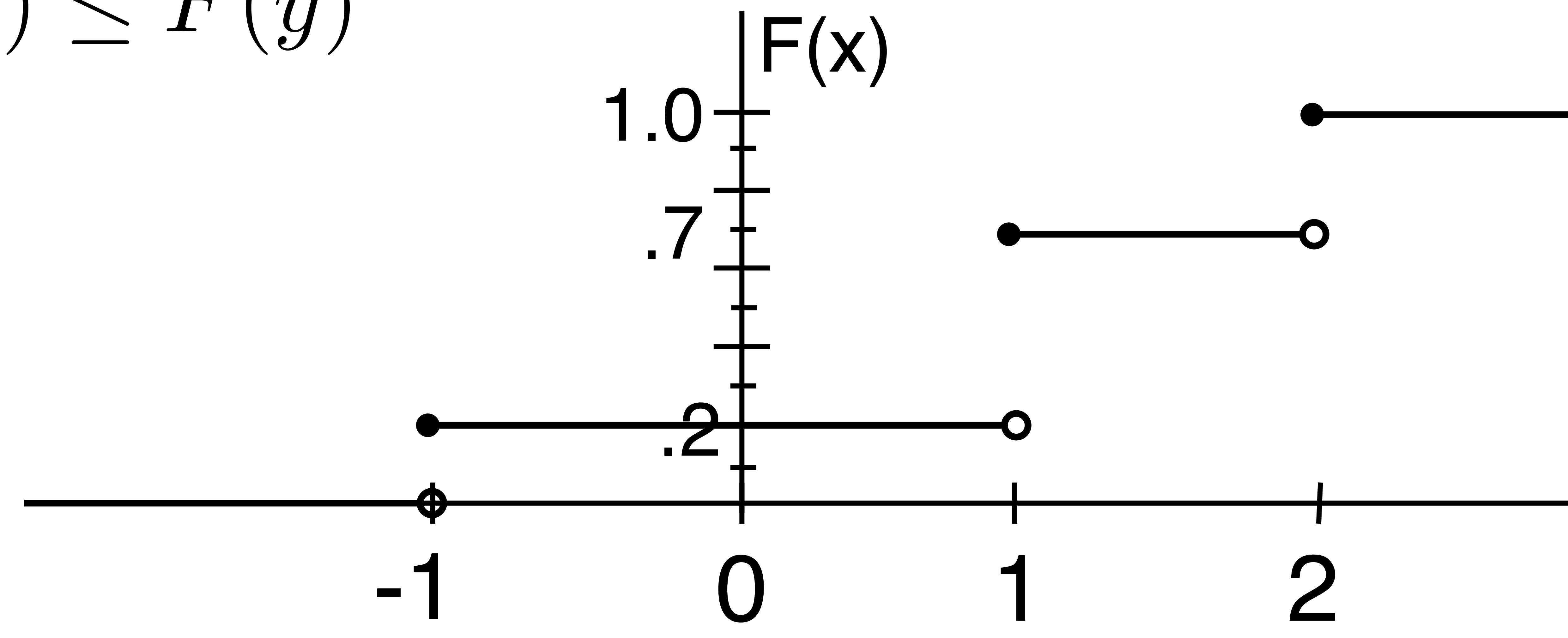


Properties

Nondecreasing $\quad x \leq y \quad \rightarrow \quad F(x) \leq F(y)$

Limits $\quad \lim_{x \to -\infty} F(x) = 0$

$$\lim_{x \to \infty} F(x) = 1$$



Right-continuous $\quad \lim_{x \searrow a} F(x) = F(a)$

Interval Probabilities

$$P(X \leq a) = F(a) \quad - \text{by definition}$$

$$P(X > a) = 1 - P(X \leq a) = 1 - F(a)$$

$$P(a < X \leq b) = P((X \leq b) - (X \leq a))$$

$$= P(X \leq b) - P(X \leq a)$$

$$= F(b) - F(a)$$

Expectations

What Matters

Important random-variable properties?

Range

Min & max values of X

Lowest & highest temperature / salary

$x_{min} = \min \{ x \in \Omega \mid p(x) > 0 \}$

$x_{max} = \max \{ x \in \Omega \mid p(x) > 0 \}$

Average

Average temperature / salary

Range average

$$\frac{x_{\min} + x_{\max}}{2} \ ?$$

Element average

$$\frac{1}{|\Omega|} \sum_{x \in \Omega} x \ ?$$

or over x s.t. $p(x)>0$

Sample Mean

$\Omega = \{0,\ldots,100\}$ $p(0)=.8$ $p(90)=.1$ $p(100)=.1$ all other $p(x)= 0$

Range average $(x_{min}+x_{max})/2$ $(0+100)/2 = 50$

Element average positive probabilities $(0+90+100)/3 = 63.3$

Ten samples Typical 0, 0, 0, 0, 90, 0, 0, 0, 100, 0

Sample mean $(8\cdot 0+1\cdot 90+1\cdot 100)/10 = 190/10 = 19$

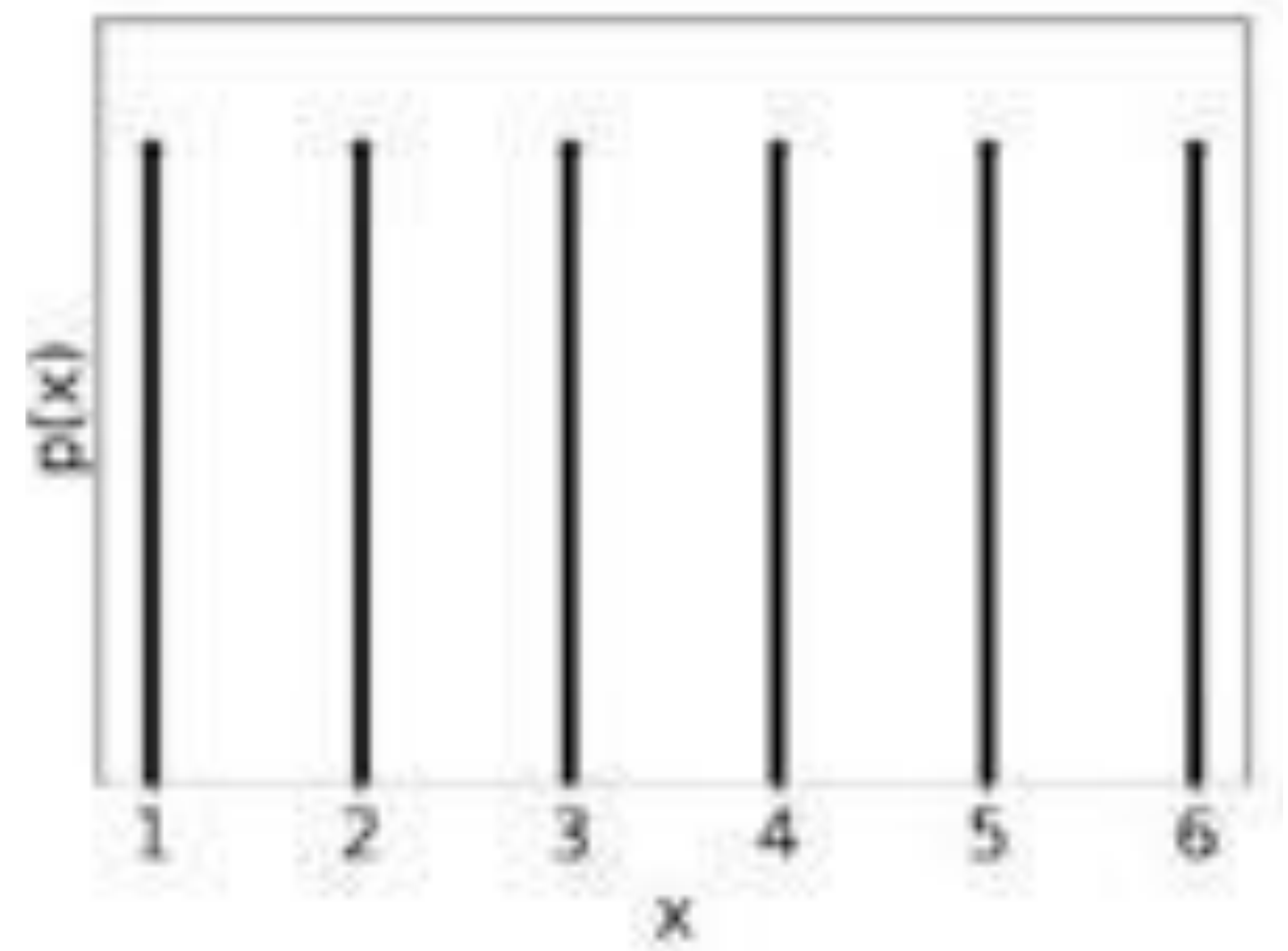
More representative of what we will observe

Fair Die

Roll a fair die $n \to \infty$ times

Average of the observed values = ?

Each value ~n/6 times



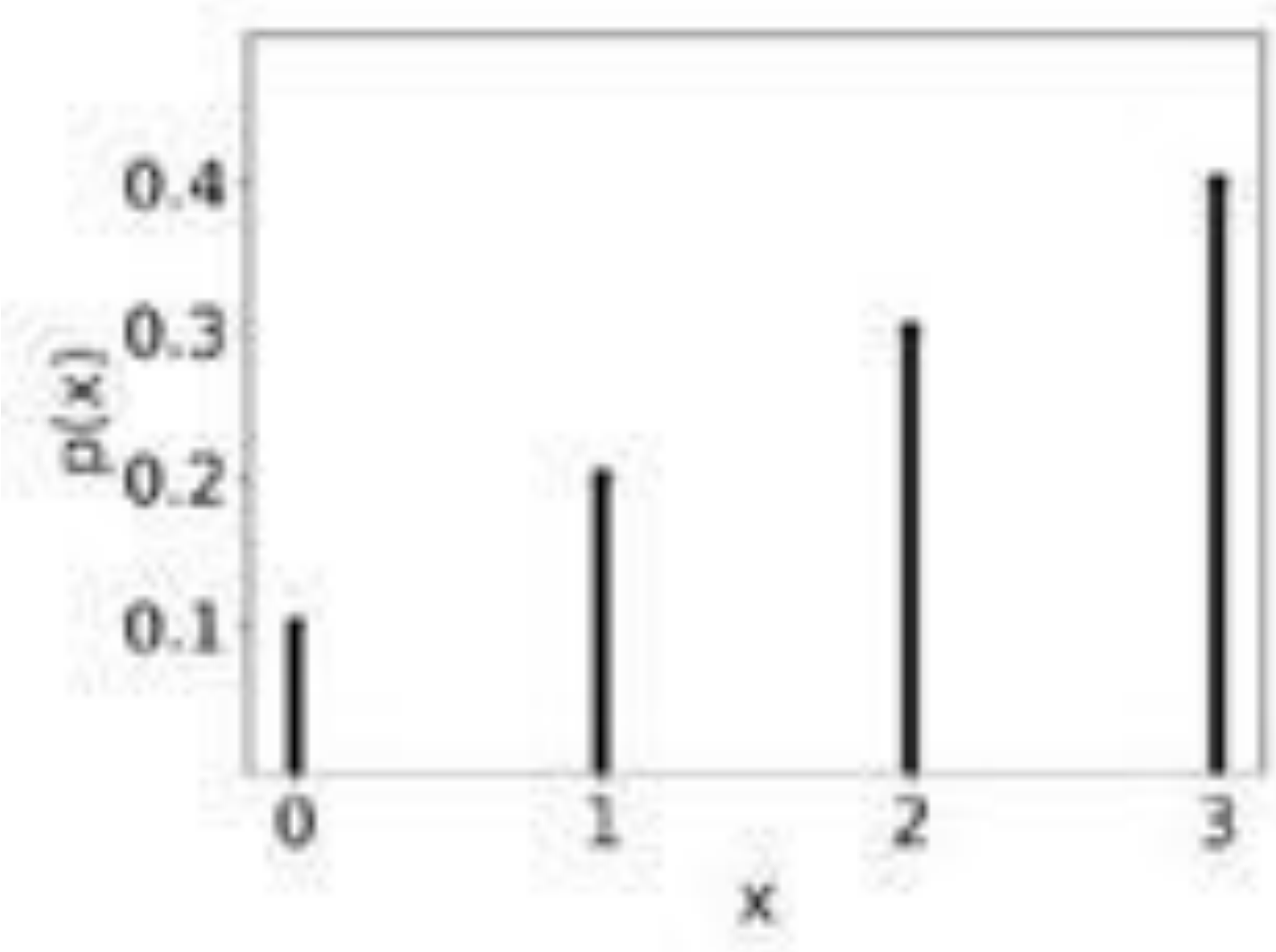
Average

$$\frac{\frac{n}{6}\cdot 1 + \frac{n}{6}\cdot 2 + \ldots + \frac{n}{6}\cdot 6}{n} = \frac{1+\ldots+6}{6} = \frac{1}{6}\,\frac{(1+6)\cdot 6}{2} = 3.5$$

1,…,6 → Average = 3.5

4-Sided Die

Side	Prob	Appear
1	.1	.1n
2	.2	.2n
3	.3	.3n
4	.4	.4n
	1	n



Average $= \dfrac{.1n \cdot 1 + .2n \cdot 2 + .3n \cdot 3 + .4n \cdot 4}{n}$

$= 0.1 \cdot 1 + 0.2 \cdot 2 + 0.3 \cdot 3 + 0.4 \cdot 4$ $= 3$

Arithmetic average $(1+2+3+4)/4 = 2.5$

Probabilities skew to the right

Expectation

In n → ∞ samples   x will appear   p(x)·n times

$$\text{Average} = \frac{\sum_x [P(x)\cdot n]\cdot x}{n} = \boxed{\sum_x P(x)\cdot x \stackrel{\text{def}}{=} E(X)}$$

Expectation

Mean

E(X) also denoted   EX μ_X μ

Not random   constant   property of the distribution

Fair Die

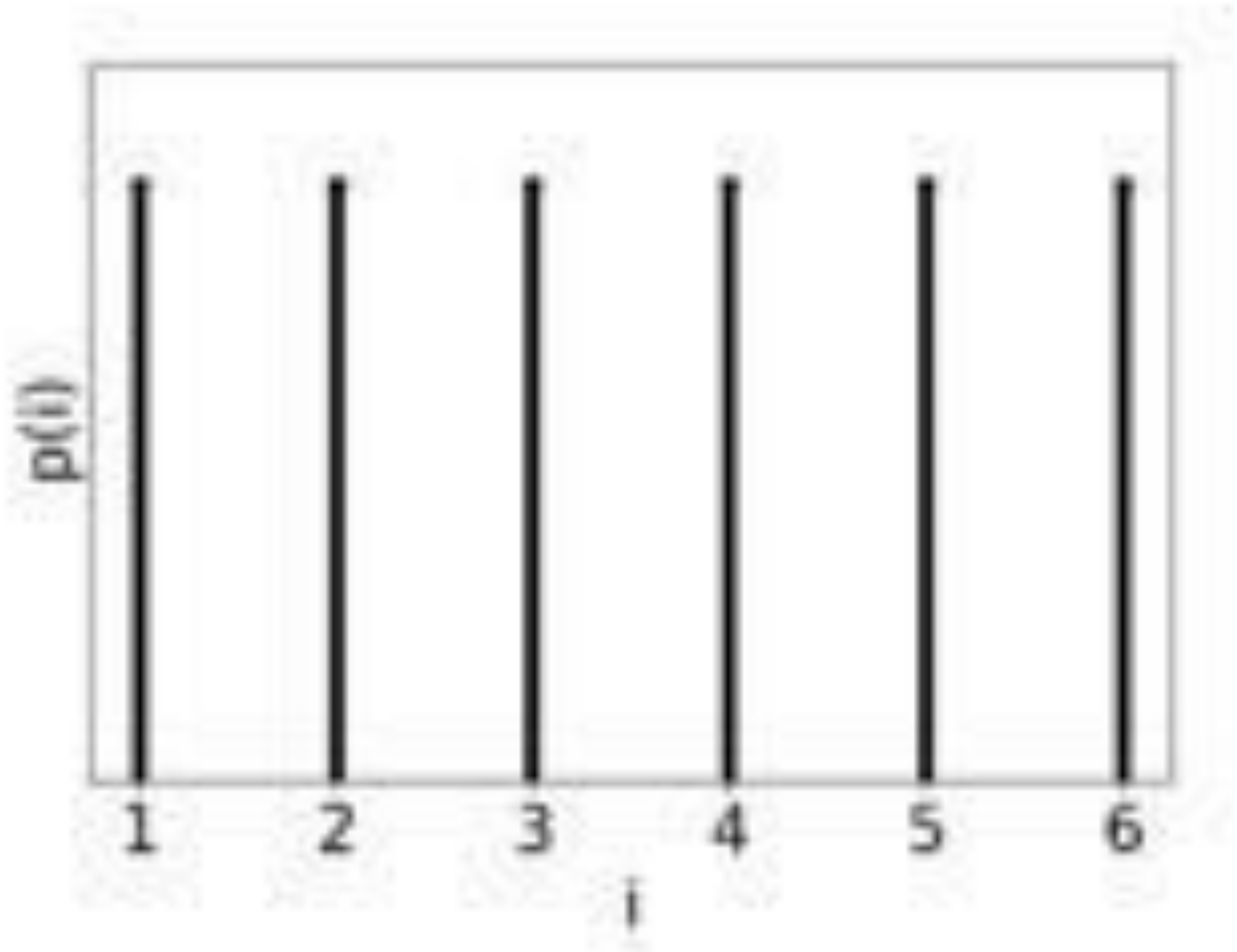
$$E(X) = \sum_{i=1}^{6} P(i) \cdot i$$

$$= \sum_{i=1}^{6} \frac{1}{6} \cdot i$$

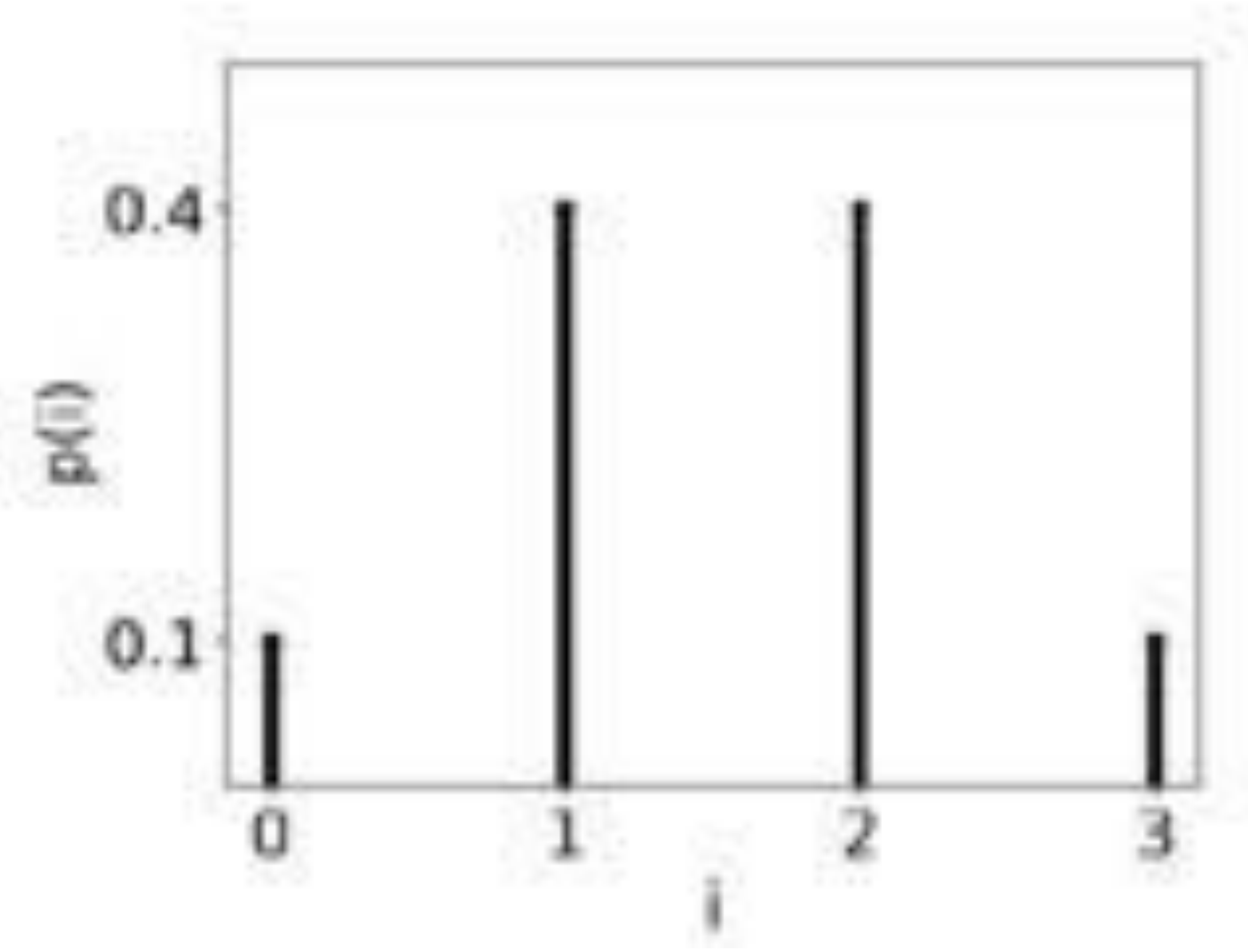
$$= \frac{1 + 2 + \ldots + 6}{6}$$

$$= \frac{1}{6} \frac{(1 + 6) \cdot 6}{2}$$

$$= \frac{7}{2} = 3.5$$ ✅



4 Sided- Die



$$E(X) = \sum_{i=1}^{4} p_i \cdot i$$

$$= 0.1 \cdot 1 + 0.2 \cdot 2 + 0.3 \cdot 3 + 0.4 \cdot 4$$

$$= 3$$ ✅

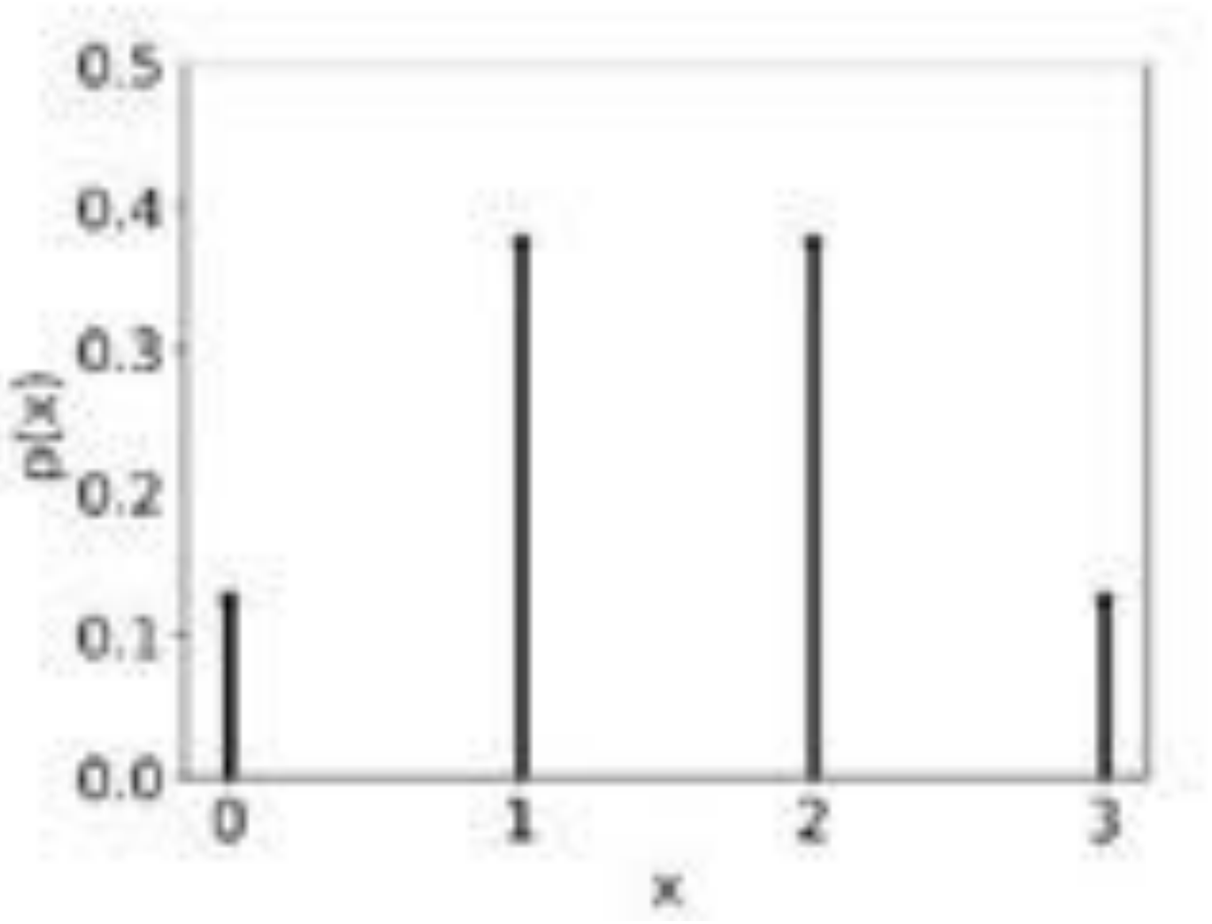
3 Coins

Toss a coin 3 times

X - # heads

E(X) = ?

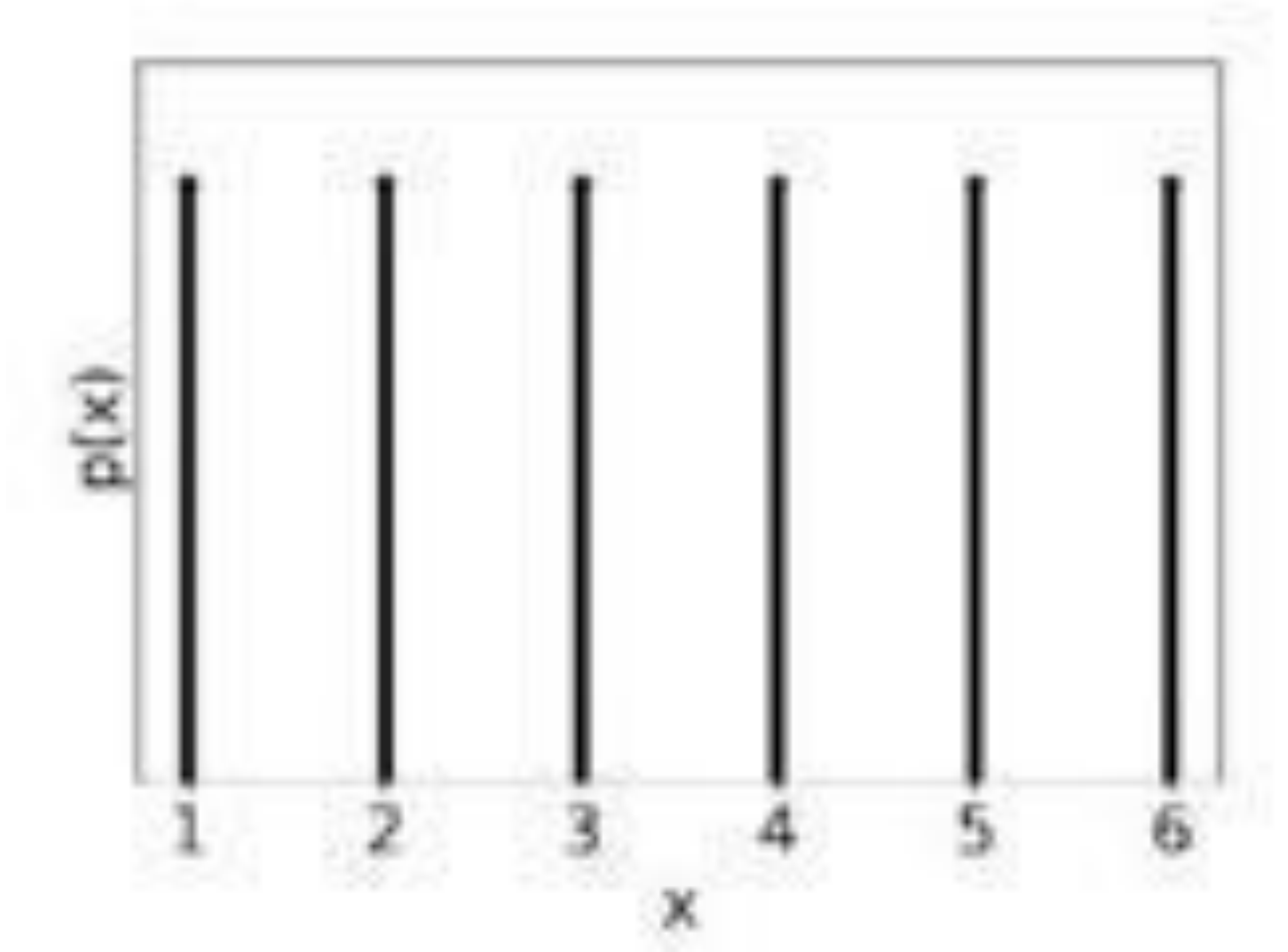
x	outcomes	p(x)
0	ttt	⅛
1	tth,tht,htt	⅜
2	thh, hth, hht	⅜
3	hhh	⅛



$\sum P(x) \cdot x$ $= ⅛ \cdot 0 + ⅜ \cdot 1 + ⅜ \cdot 2 + ⅛ \cdot 3$ $= 1.5$

heads ranges from 0 to 3, on average 1.5 ✅

Uniform Variables



X uniform over Ω

$$p(x) = \frac{1}{|\Omega|}$$

$$E(X) \;=\; \sum_{x \in \Omega} p(x) \cdot x \;=\; \sum_{x \in \Omega} \frac{1}{|\Omega|} \cdot x \;=\; \frac{1}{|\Omega|} \sum_{x \in \Omega} x$$

E(X) is the arithmetic average of elements in Ω

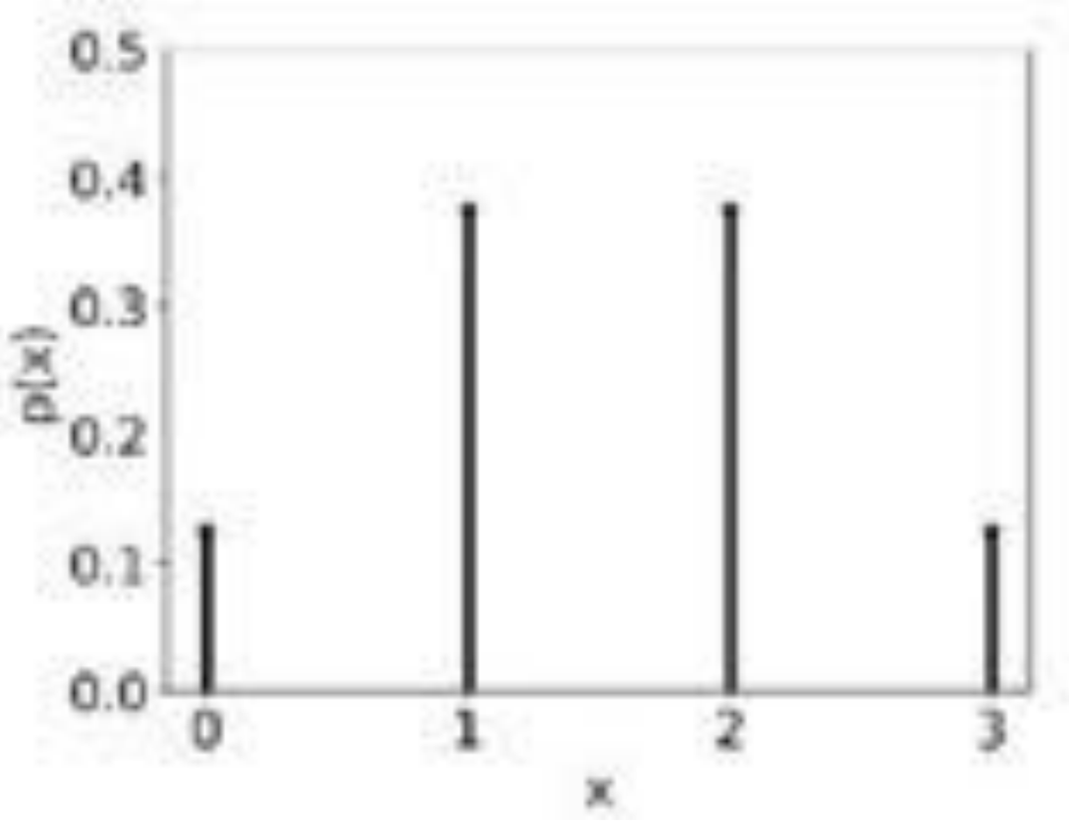
$$E(X) = \frac{1+2+\ldots+6}{6} = 3.5$$

Symmetry

A distribution p is symmetric around a if for all $x > 0$, $p(a+x) = p(a-x)$

If p is symmetric around a, then $E(X) = a$

x	outcomes	P(x)
0	ttt	⅛
1	tth,tht,htt	⅜
2	thh, hth, hht	⅜
3	hhh	⅛



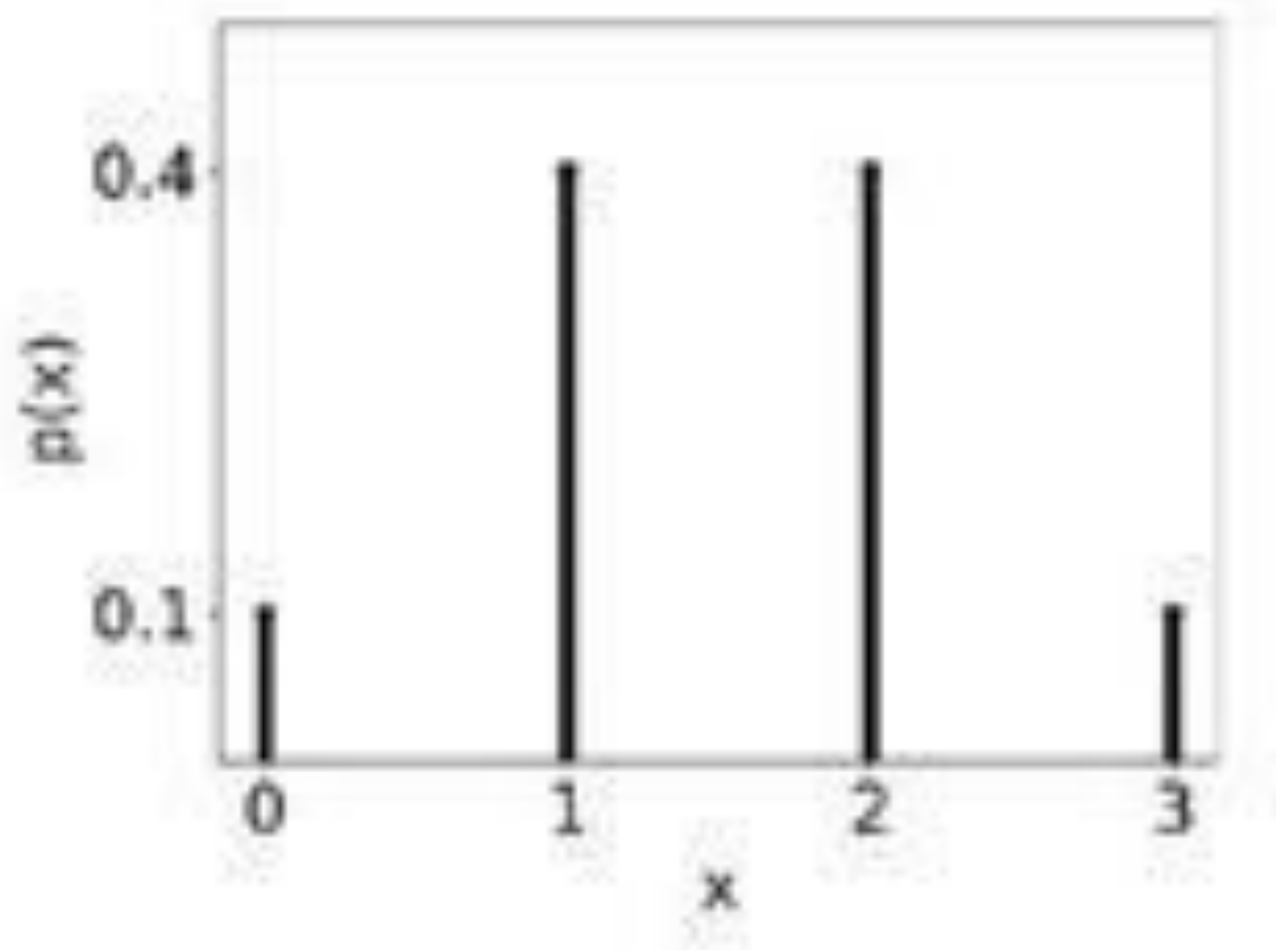
Symmetric around 1.5

$E(X) = 1.5$

Properties

E(X) | Despite notation | Not random | Number | Property of distribution

$E(X) = 1.5$



$x_{min} \leq E(X) \leq x_{max}$ | $=$ iff $X = c$

$0 \leq E(X) \leq 3$

X is a constant, namely $X=c$ → $E(X)=c$

$E(E(X)) = E(X)$

Is Expectation Expected?

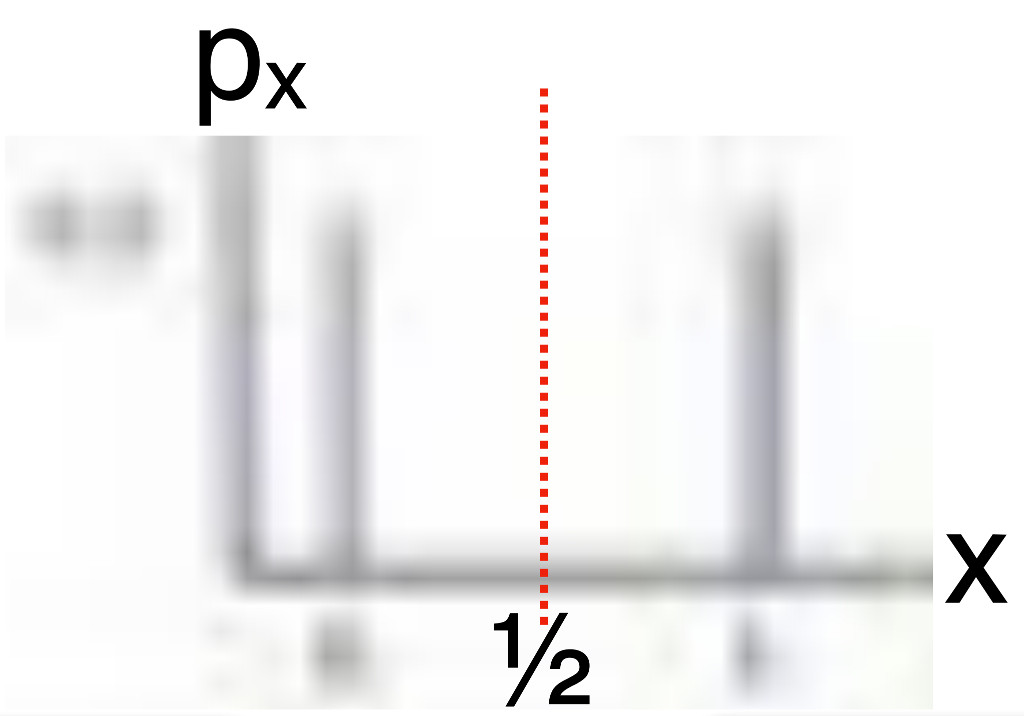
$\mu = EX$ - expectation of X

Do we expect to see it?

Is p_μ high?

Not necessarily

We may never see it!



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

$p_0 = p_1 = 0.5$

$EX = 0 \cdot p_0 + 1 \cdot p_1 = 0 \cdot ½ + 1 \cdot ½ = ½$

Symmetric around ½

½ will never happen!

Many samples → average = ½

EX - average of large sample

Not necessarily likely

May not be observed at all

Infinite Expectation

$$\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$$

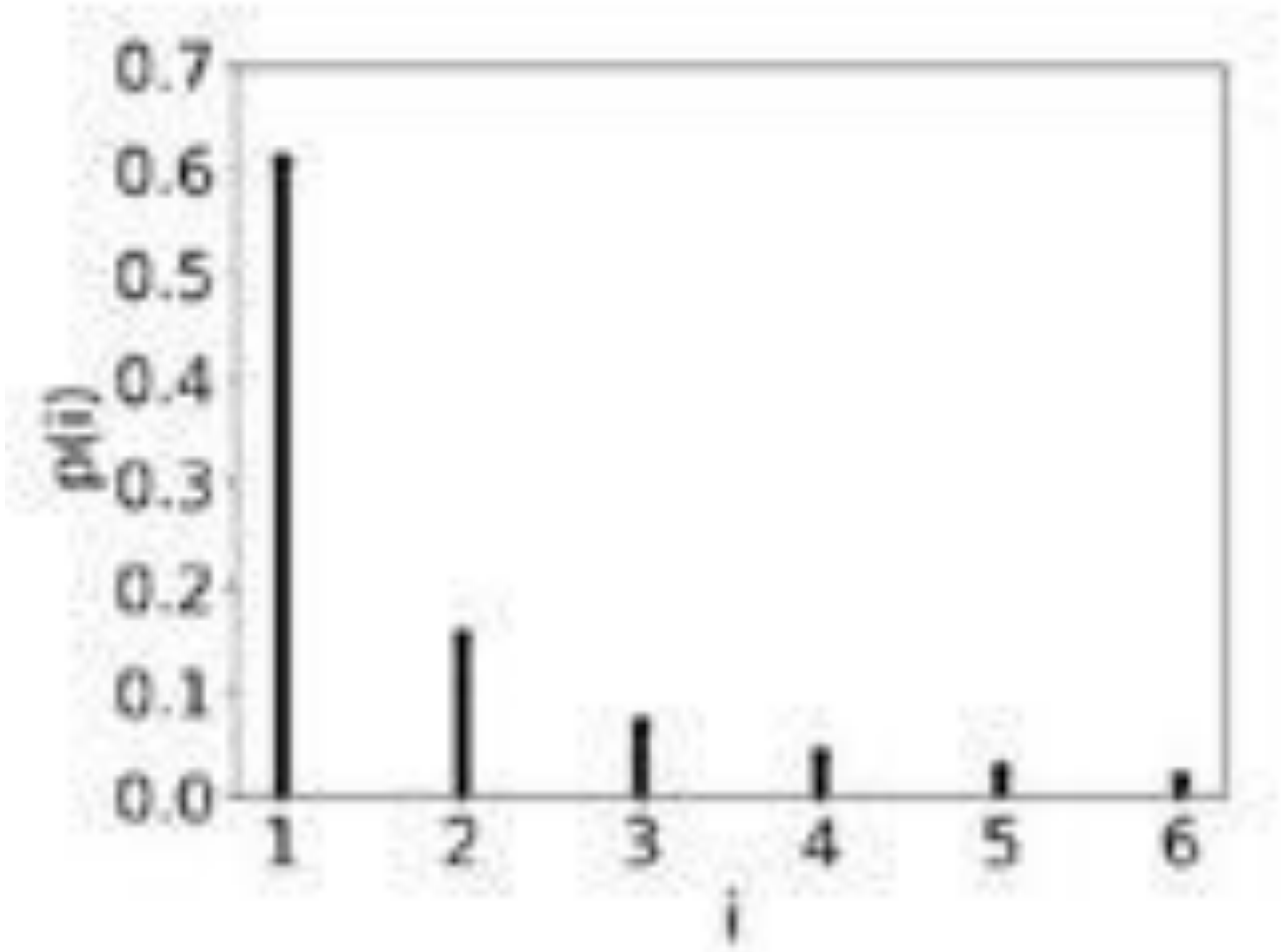
Basel problem

Euler → famous

$$\frac{6}{\pi^2} \sum_{i=1}^{\infty} \frac{1}{i^2} = 1$$

$$p_i = \frac{6}{\pi^2} \cdot \frac{1}{i^2}$$

probability distribution over $\mathbb{P}$



$$E(X) = \sum_{i=1}^{\infty} i \cdot p_i \;\; = \frac{6}{\pi^2} \sum_{i=1}^{\infty} \frac{1}{i} \;\; = \infty$$

Many samples

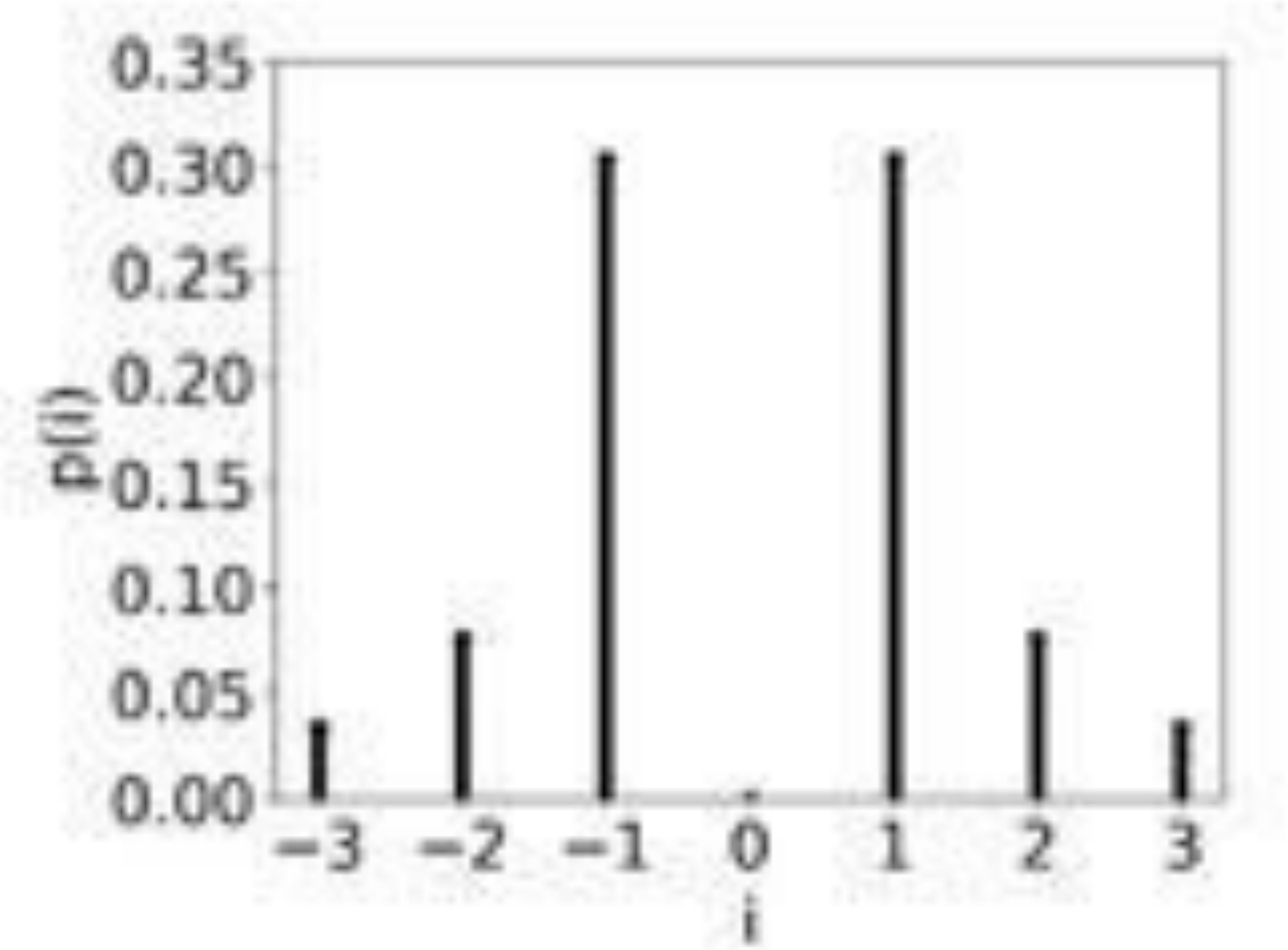
Average will go to ∞

Undefined Expectation

$$p_i = \frac{3}{\pi^2} \cdot \frac{1}{i^2} \quad \text{for} \quad i \neq 0$$

E(X) = ∞ - ∞

Undefined

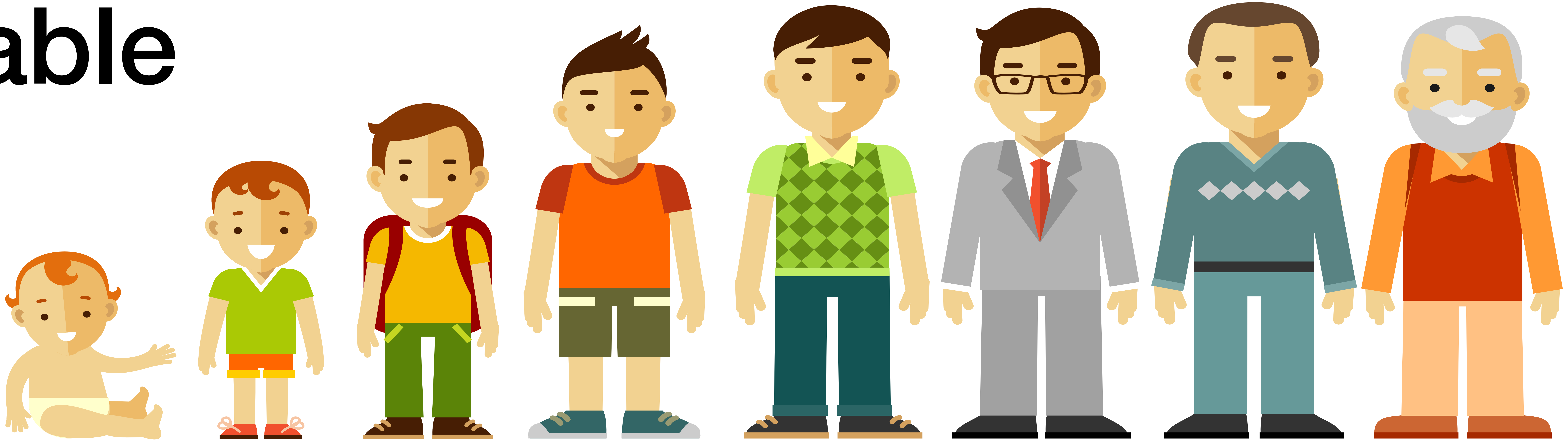


Life Expectancy

Expectation

Expectations of Functions of Random Variables

Variable



Modifications

(aka functions)

Functions of a Random Variable

Random variables X take values in $\mathbb{R}$

Often interested in related variable

$Y = g(X)$

$g: \mathbb{R} \rightarrow \mathbb{R}$ is a fixed function

X

Random salary in $

$10 raise

$Y = X + 10$

10% raise

$Y = 1.1X$

→ CEO

$Y = X^2$

Deterministic Functions

$Y = g(X)$

g is a deterministic function over $\mathbb{R}$ (or Ω)

$Y = X + 3$

All randomness in Y derives from X

Deterministically modified by g

$X = 5$

$Y = 8$

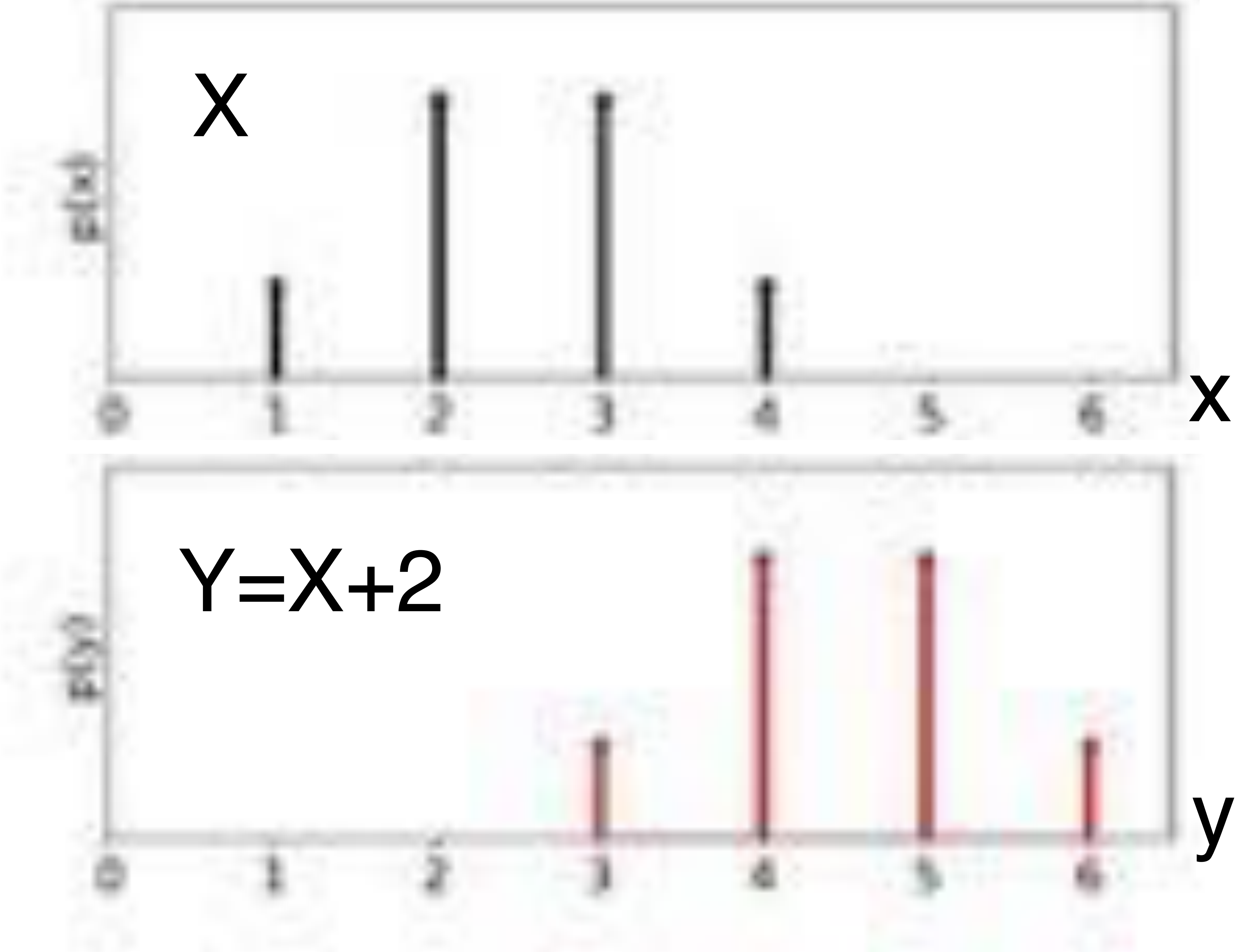
Translation

Add constant b to X

Translate X by b

Y = X + b

P(Y=y) = P(X+b=y) = P(X=y-b)



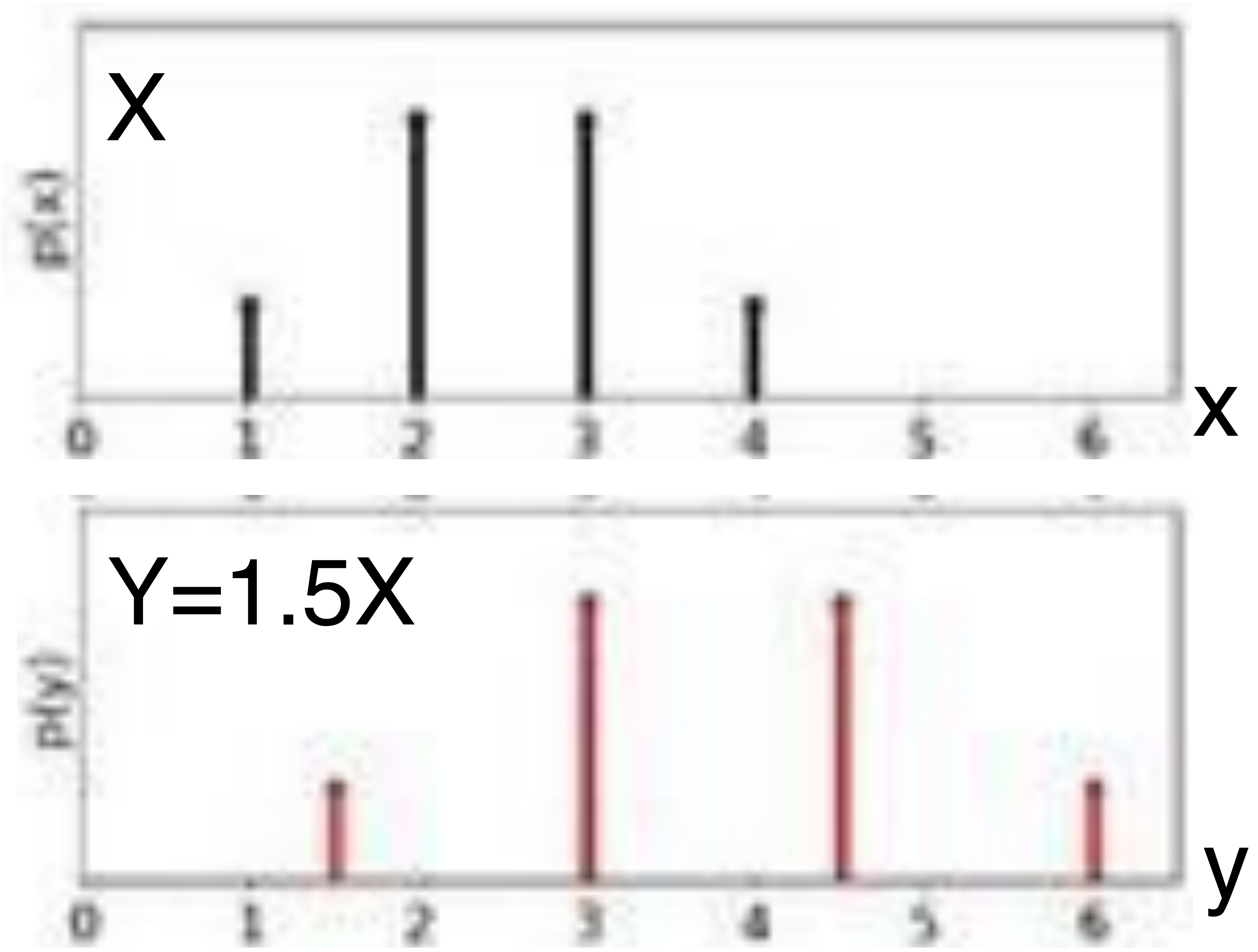
Scaling

Multiply X by a constant b

Scale X by a factor b

$Y = b \cdot X$

$P(Y=y) = P(bX=y) = P(X=y/b)$



Two Square Examples

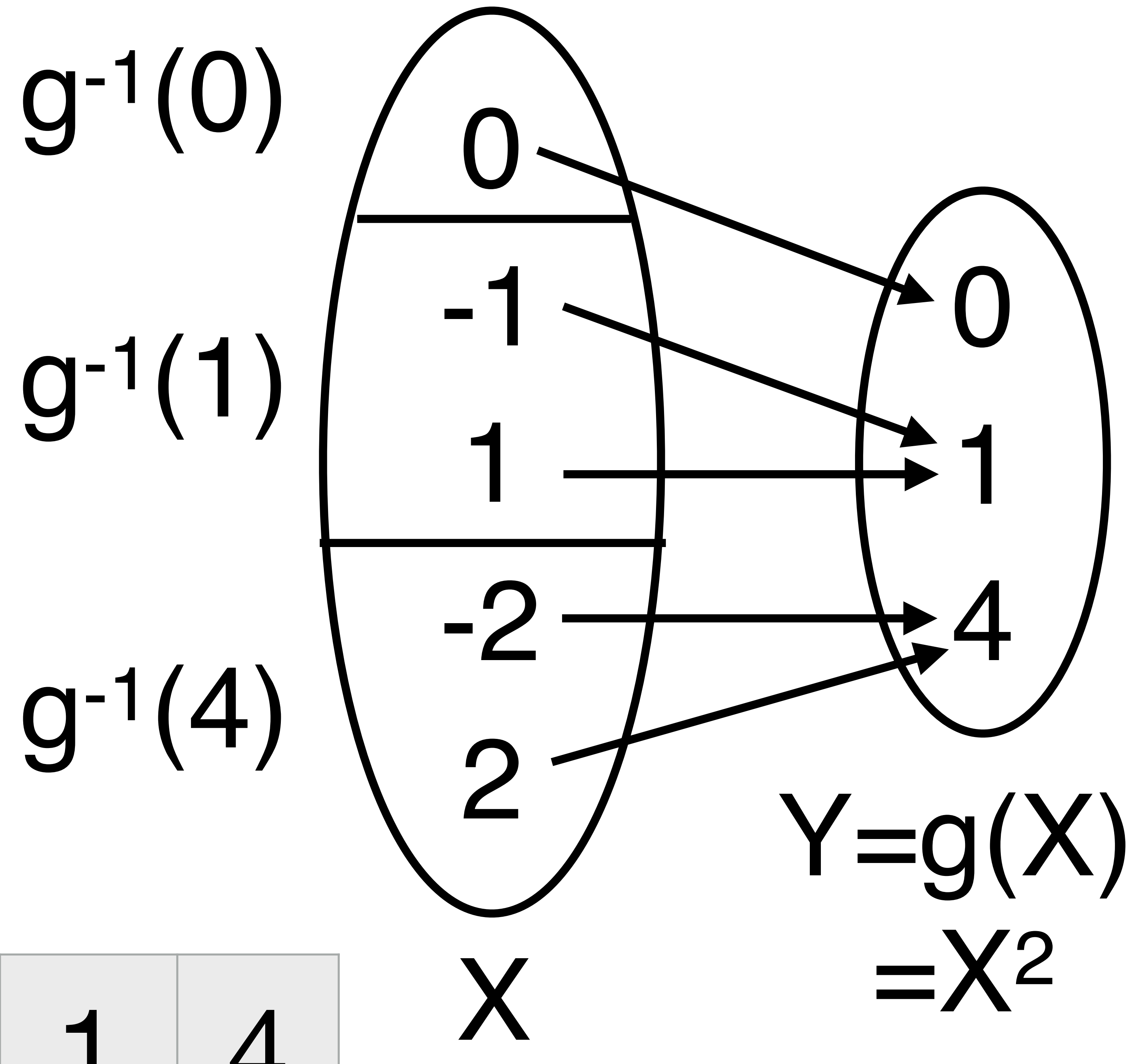
Square is 1-1

X

x	0	1	2
p(X = x)	½	⅓	⅙

$Y = X^2$

y	0	1	4
p(Y = y)	½	⅓	⅙



Square is many to 1

X

x	-2	-1	0	1	2
p(X = x)	⅕	⅕	⅕	⅕	⅕

$Y = X^2$

y	0	1	4
p(Y = y)	⅕	⅖	⅖

$$P(Y=y) = P(g(X)=y) = P(X \in g^{-1}(y)) = \sum_{x \in g^{-1}(y)} P(X=x)$$

Expectation of Functions of Random Variables

Expectation Reminder

X

x	-2	-1	0	1	2
p(x)	⅕	⅕	⅕	⅕	⅕

$$E(X) = \sum_x p(x) \cdot x$$

$$= \underbrace{-2 \cdot \tfrac{1}{5} + \underbrace{-1 \cdot \tfrac{1}{5} + 0 \cdot \tfrac{1}{5} + 1 \cdot \tfrac{1}{5}}_{0} + 2 \cdot \tfrac{1}{5}}_{0} = 0$$

"By Symmetry"

Expectation of a Square

X

x	-2	-1	0	1	2
p(x)	⅕	⅕	⅕	⅕	⅕

$Y = X^2$

y	0	1	4
p(y)	⅕	⅖	⅖

$P(Y = 0) = P(X^2 = 0) = P(X = 0) = 1/5$

$P(Y = 1) = P(X^2 = 1) = P(X \in \{-1, 1\}) = 2/5$

$P(Y = 4) = P(X^2 = 4) = P(X \in \{-2, 2\}) = 2/5$

$E(Y) = 1/5 \cdot 0 + 2/5 \cdot 1 + 2/5 \cdot 4 = 10/5 = 2$

Alternative Formulation

$$
\begin{aligned}
E(Y) &= \sum_{y} y \cdot P(Y=y) \\
&= \sum_{y} y \cdot P(X \in g^{-1}(y)) \\
&= \sum_{y} \; y \sum_{x \in g^{-1}(y)} p(x) \\
&= \sum_{y} \; \sum_{x \in g^{-1}(y)} y \cdot p(x) \\
&= \sum_{y} \; \sum_{x \in g^{-1}(y)} g(x) \cdot p(x) \\
&= \sum_{x} g(x) \cdot p(x)
\end{aligned}
$$

Example

Visualize

Square Again

X

x	-2	-1	0	1	2
p(x)	⅕	⅕	⅕	⅕	⅕

$Y = X^2$

y	0	1	4
p(y)	⅕	⅖	⅖

$$E(Y) = \sum_{y=0,1,4} y \cdot p(Y{=}y) = \tfrac{1}{5} \cdot 0 + \tfrac{2}{5} \cdot 1 + \tfrac{2}{5} \cdot 4 = 10/5 = 2$$

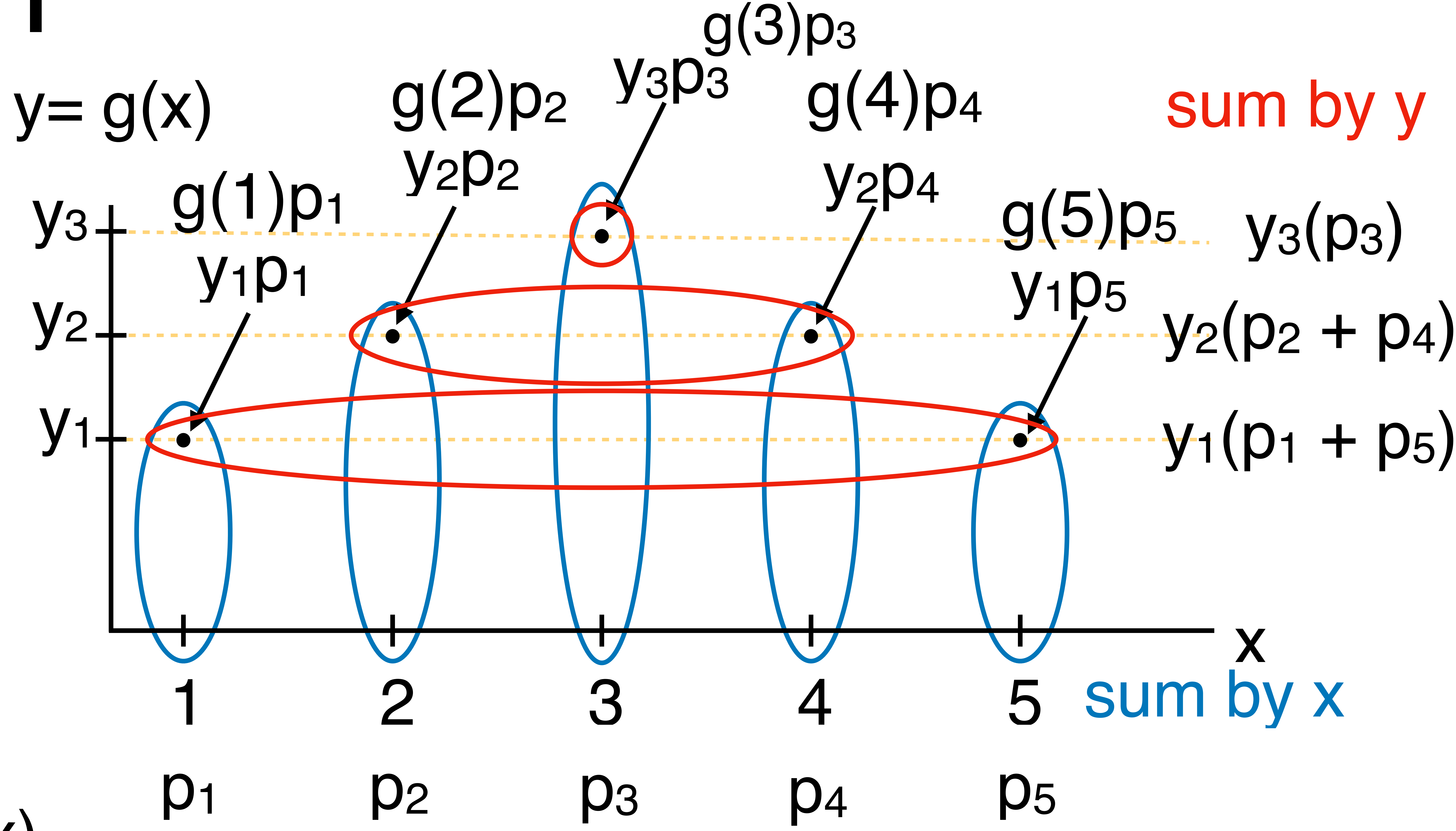
$$E(Y) = \sum_{x} x^2 \cdot p(x)$$

$$= (-2)^2 \cdot \tfrac{1}{5} + (-1)^2 \cdot \tfrac{1}{5} + 0^2 \cdot \tfrac{1}{5} + 1^2 \cdot \tfrac{1}{5} + 2^2 \cdot \tfrac{1}{5}$$

$$= \tfrac{4}{5} + \tfrac{1}{5} + \tfrac{1}{5} + \tfrac{4}{5} = 2$$

Visualization

$$E(Y) = \sum_y y \cdot P(Y{=}y)$$
$$= \sum_y y \cdot P(X \in g^{-1}(y))$$
$$= \sum_y y \sum_{x \in g^{-1}(y)} p(x)$$
$$= \sum_y \sum_{x \in g^{-1}(y)} y \cdot p(x)$$
$$= \sum_y \sum_{x \in g^{-1}(y)} g(x) \cdot p(x)$$
$$= \sum_x g(x) \cdot p(x)$$



y: Fewer multiplication

x: Properties (next)

General Formulas

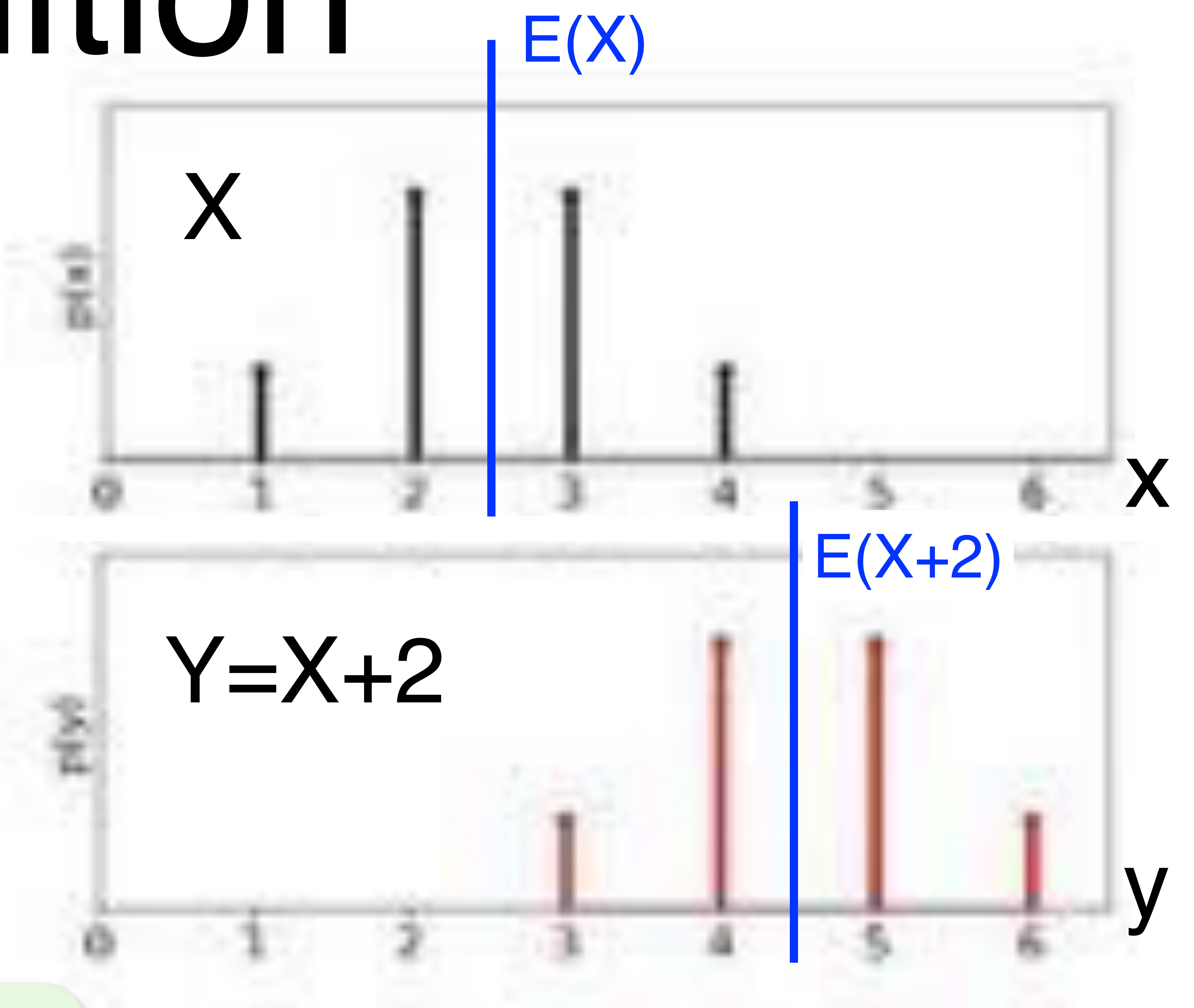
Constant Addition

$E(X + b)$ $= \sum p(x) \cdot (x + b)$

$= \sum p(x) \cdot x + \sum p(x) \cdot b$

$= E(X) + b \cdot \sum p(x)$

$= E(X) + b$



x	0	1
p(x)	1 - p	p

Bernoulli p

$E(X) = (1 - p) \cdot 0 + p \cdot 1 = p$

$E(X + 2) = (1 - p) \cdot (0 + 2) + p \cdot (1 + 2)$

$= 2 - 2p + 3p$

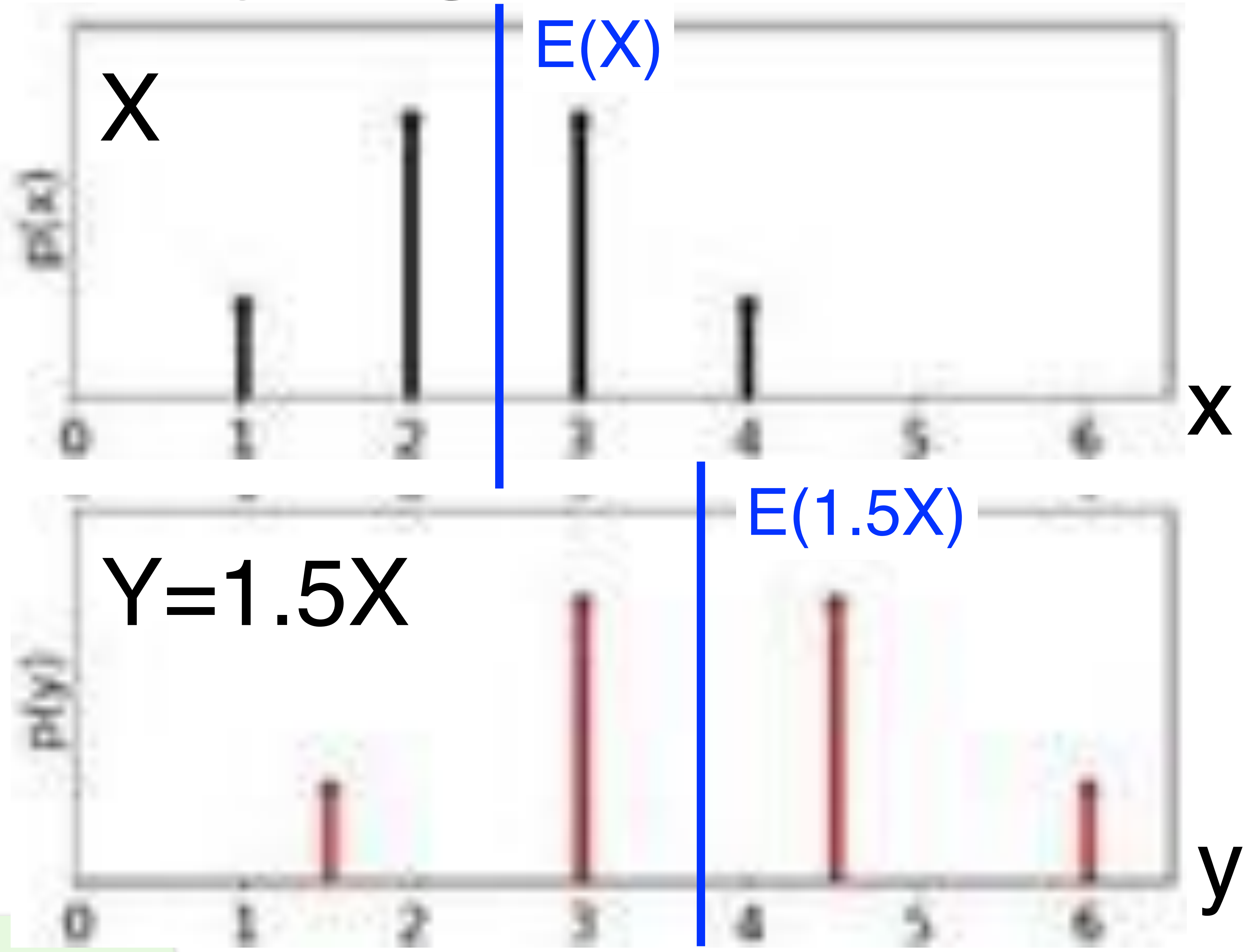
$= p + 2 = E(X) + 2$

Constant Multiplication

$$E(aX) = \sum p(x) \cdot (ax)$$

$$= a \sum p(x) \cdot x$$

$$= aE(X)$$



Bernoulli p

x	0	1
p(x)	1 - p	p

$$E(X) = (1 - p) \cdot 0 + p \cdot 1 = p$$

$$E(3X) = (1 - p) \cdot (3 \cdot 0) + p \cdot (3 \cdot 1)$$

$$= 3p = 3E(X)$$

Linearity of Expectation

$E(aX + b) = E(aX) + b$

$= a\,E(X) + b$

Bernoulli p

x	0	1
p(x)	1 - p	p

$E(X) = (1 - p) \cdot 0 + p \cdot 1 = p$

$E(2X + 3) = (1 - p)(2 \cdot 0 + 3) + p(2 \cdot 1 + 3)$

$= 3 - 3p + 5p$

$= 2p + 3 = 2E(X) + 3$

Expectation of Functions of Variables

Next: Variance

Variance

Distribution Properties

Deterministic functions of distribution

Not random

⬅ Long-term average

Expectation EX, μ

Die: μ=3.5

➡ Consistency

Variation from the mean

Money Matters

Two companies, each with 1,000 employees

Both same mean salary: $100K But

C1: Every employee makes $100K

100M total

C2: Every employee makes $1, CEO $99,999,001

100M-999

Which will you join?

Same mean Very different distributions

Mean ain’t all Variation matters!

Difference from Mean

X r.v. with mean μ

How much X differs from μ on average?

Candidate

$E\,|X - \mu|$

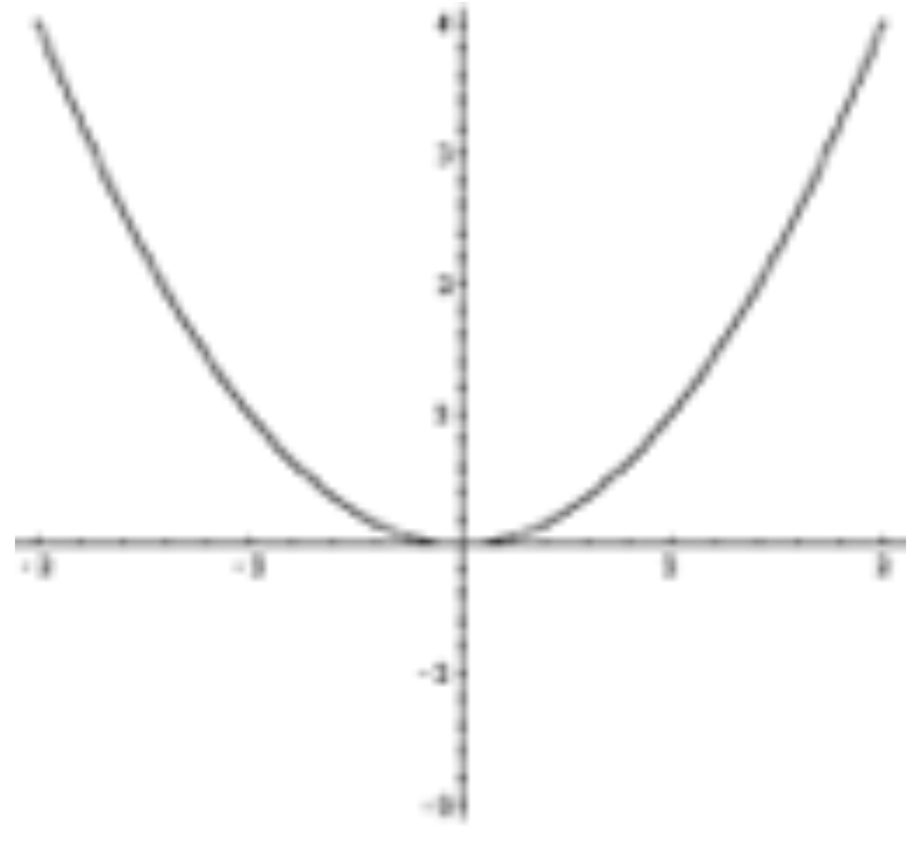
Mean absolute difference

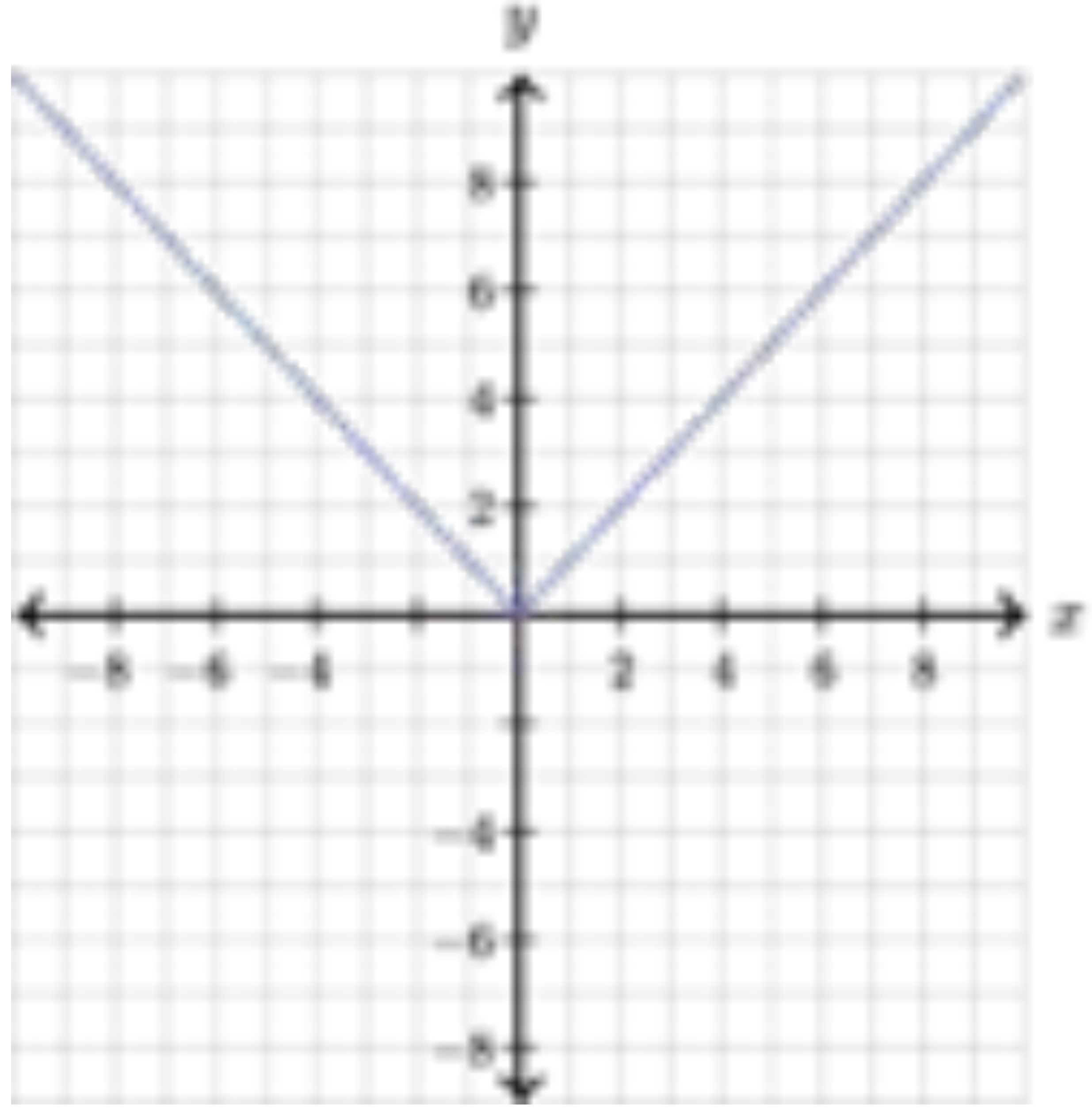
Not commonly used

Absolute value function hard to analyze

Instead

$E(X - \mu)^2$


$y = x^2$



Variance

Expected squared difference between X and its mean

$$\boxed{V(X) = E[(X - \mu)^2]}$$

$V(X) = E(X - \mu)^2$

Standard deviation

$$\boxed{\sigma_X = \sqrt{V(X)}}$$

(positive)

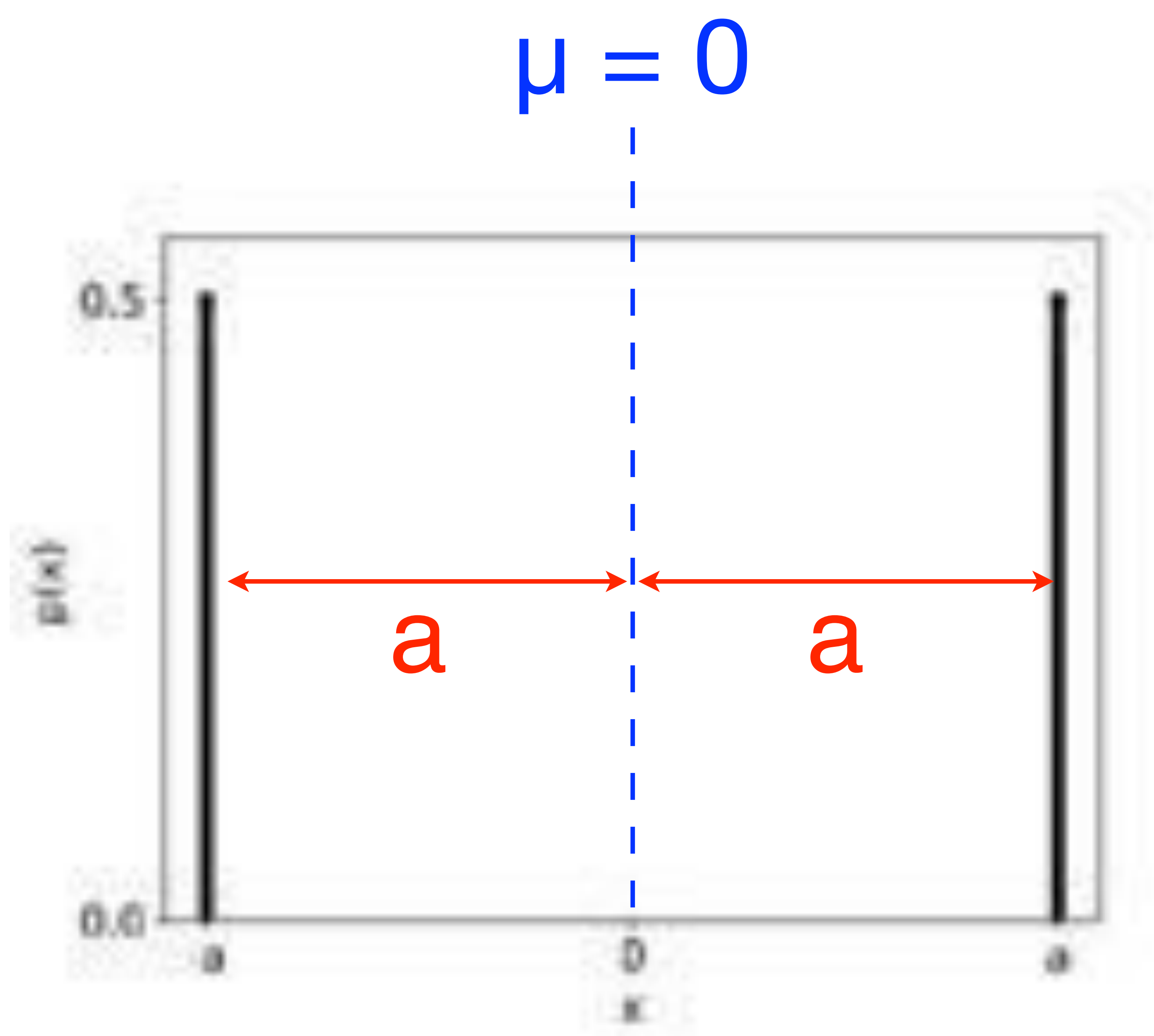
Constants

Properties of distribution

Examples

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

$\mu = 0$



$V(X) = \frac{1}{2} \cdot a^2 + \frac{1}{2} \cdot a^2 = a^2$

X^2 is always a^2

$(X - \mu)^2 = a^2$ always

$\sigma_x = a$

"average" distance from mean

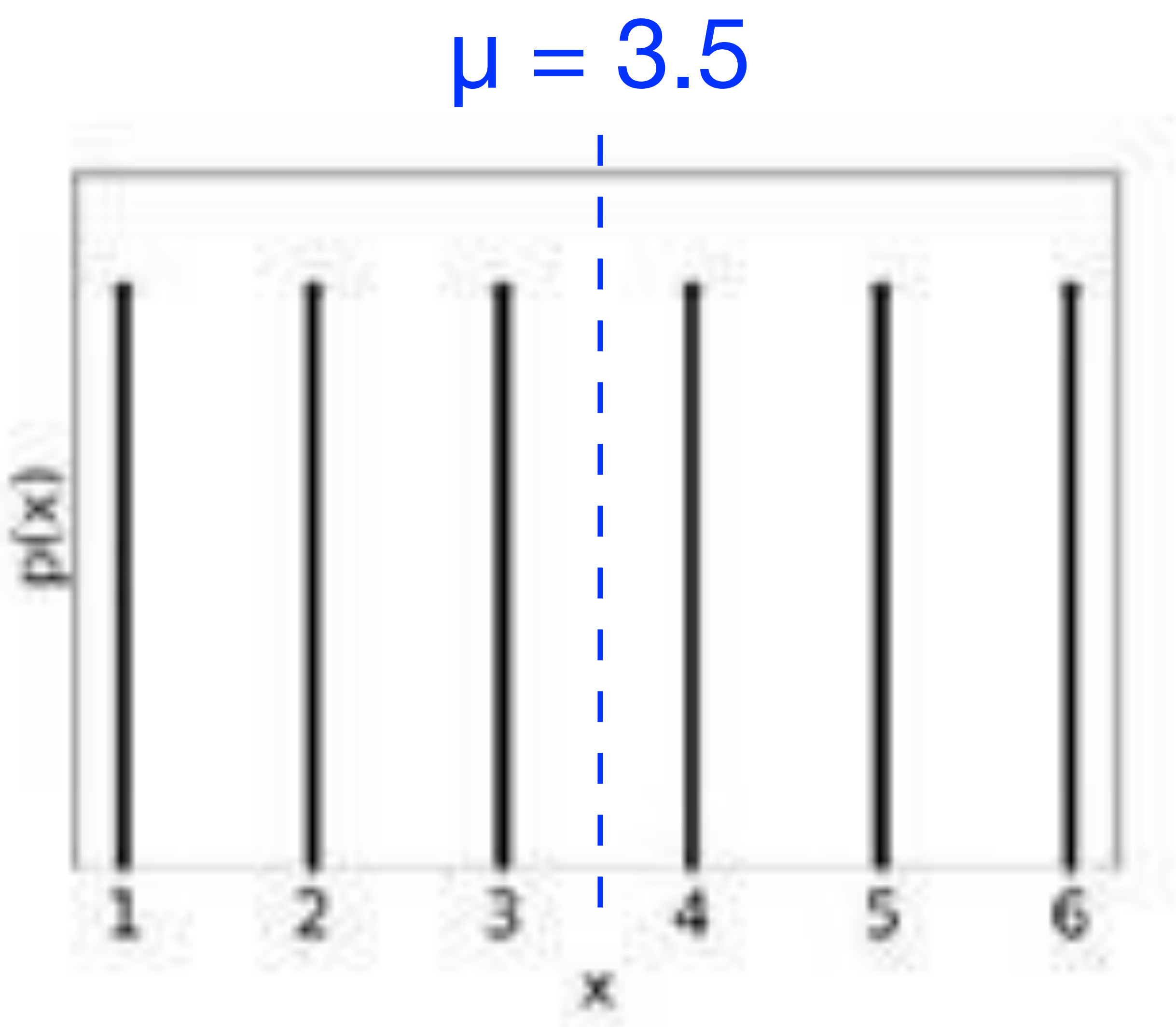
Fair Die

$\mu = 3.5$

$$V(X) = E(X - \mu)^2 = \frac{2(6.25 + 2.25 + 0.25)}{6} = \frac{8.75}{3} = 2.92..$$

$$\sigma = \sqrt{2.92\ldots} = 1.71\ldots$$

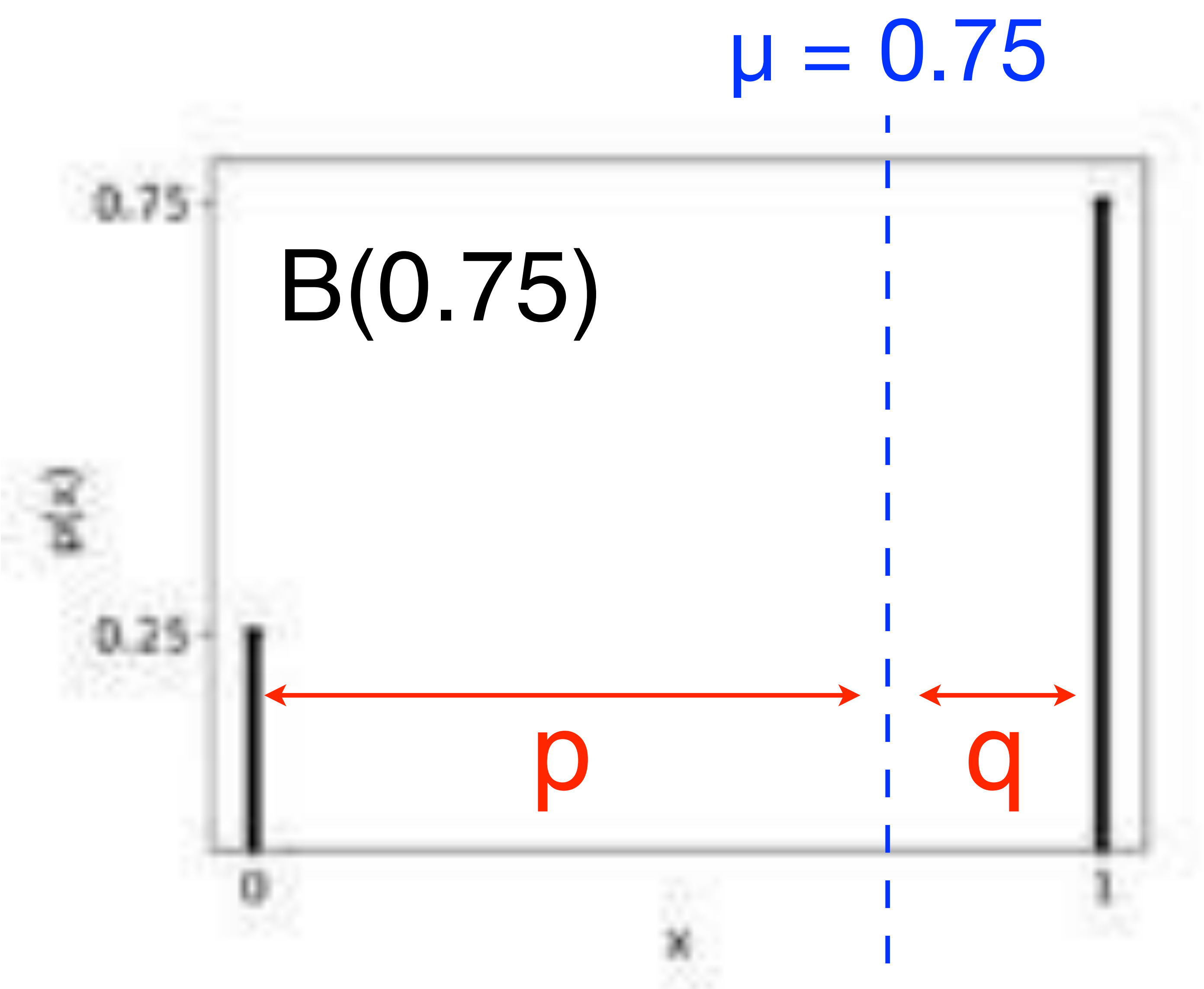
x	p_x	$x - \mu$	$(x - \mu)^2$
1	⅙	-2.5	6.25
2	⅙	-1.5	2.25
3	⅙	-0.5	0.25
4	⅙	0.5	0.25
5	⅙	1.5	2.25
6	⅙	2.5	6.25



Bernoulli p

x	p_x	x - μ	$(x - \mu)^2$
0	q	0-p = p	p^2
1	p	1-p = q	q^2

$\mu = p$



$$V(X) = q \cdot p^2 + p \cdot q^2 = pq(p+q) = pq$$

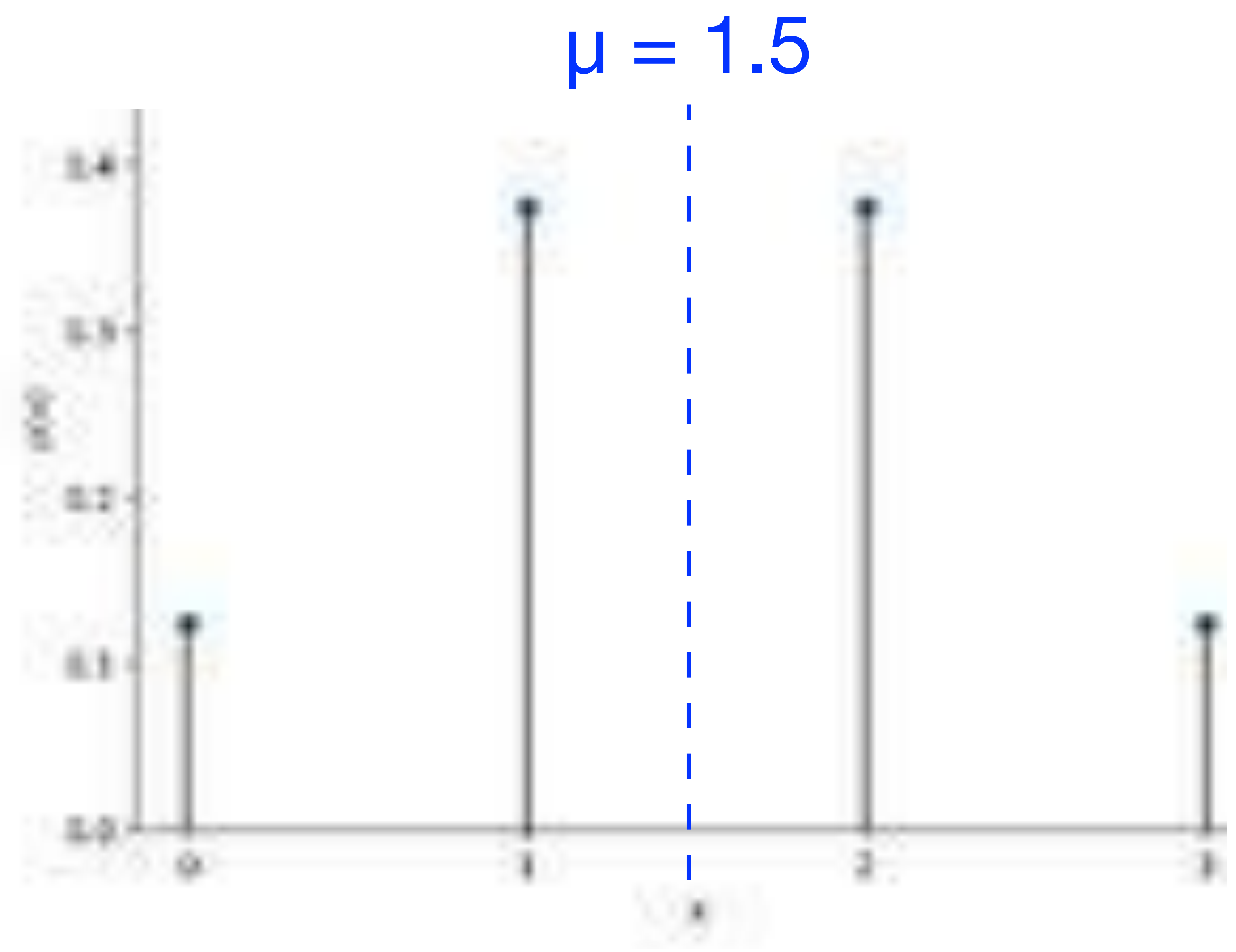
3 Coins

Toss 3 fair coins

X = # heads

x	p_x	$x - \mu$	$(x - \mu)^2$
0	⅛	-1.5	2.25
1	⅜	-0.5	0.25
2	⅜	0.5	0.25
3	⅛	1.5	2.25

$\mu = 1.5$



$V = 2(\tfrac{1}{8} \cdot 2.25 + \tfrac{3}{8} \cdot 0.25) = \tfrac{1}{4}(2.25 + 0.75) = \tfrac{3}{4}$

$\sigma = \sqrt{3}/2$

Shortly: simpler derivation

Different Formula

$V(X)$ $= E(X - \mu)^2$

$E(X)=\mu$

$= E(X^2 - 2\mu X + \mu^2)$

$= E(X^2) - E(2\mu X) + E(\mu^2)$

$2, \mu$ - constants

$= E(X^2) - 2\mu E(X) + \mu^2$

$= E(X^2) - 2\mu^2 + \mu^2$

$= E(X^2) - \mu^2$

$= E(X^2) - (E\,X)^2$

Bernoulli p Again

$X \sim B(p)$

Recall: $EX = p$ $V(X)=pq$

Re-derive using $V(X) = E\,X^2 - (EX)^2$

$E(X^2) = (1 - p) \cdot 0^2 + p \cdot 1^2 = p$

Even simpler $0^2=0,\ 1^2=1$ $\rightarrow$ $X^2=X$ $\rightarrow$ $EX^2 = EX = p$

$V(X) = E\,X^2 - (EX)^2 = p - p^2 = p(1 - p) = pq$ ✅

Observations

$V(X) = E(X - \mu)^2$

$$0 \leq V \leq \max (X-\mu)^2$$

=
X is a
constant

=
X constant or
takes two values
with equal prob.

$$0 \leq \sigma \leq \max |X-\mu|$$

$V(X) = EX^2 - \mu^2$

$$V(X) \leq E(X^2)$$

Properties

How simple modification affect V and σ

Addition	(translation)	$X + b$
Multiplication	(scaling)	$a \cdot X$
+ & x	(affine transformation)	$aX + b$

Addition

X - random variable

b - constant (e.g. 2)

$\mu_{x+b} = \mu_x + b$

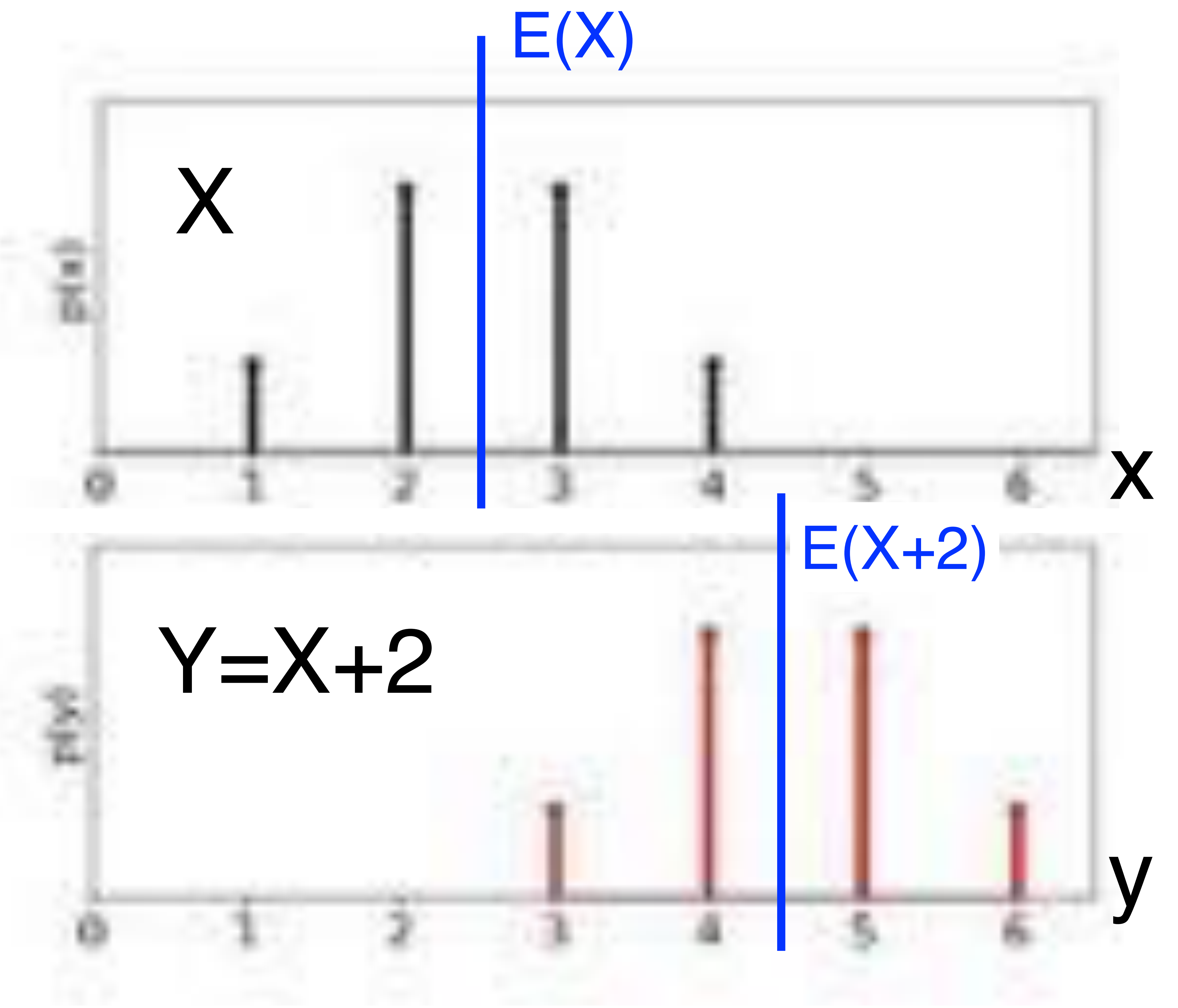
Linearity of expectation

$$V(X + b) = E[(X + b - \mu_{x+b})^2]$$

$$= E[(X + b - \mu_x - b)^2]$$

$$= E(X - \mu_x)^2$$

$$= V(X)$$

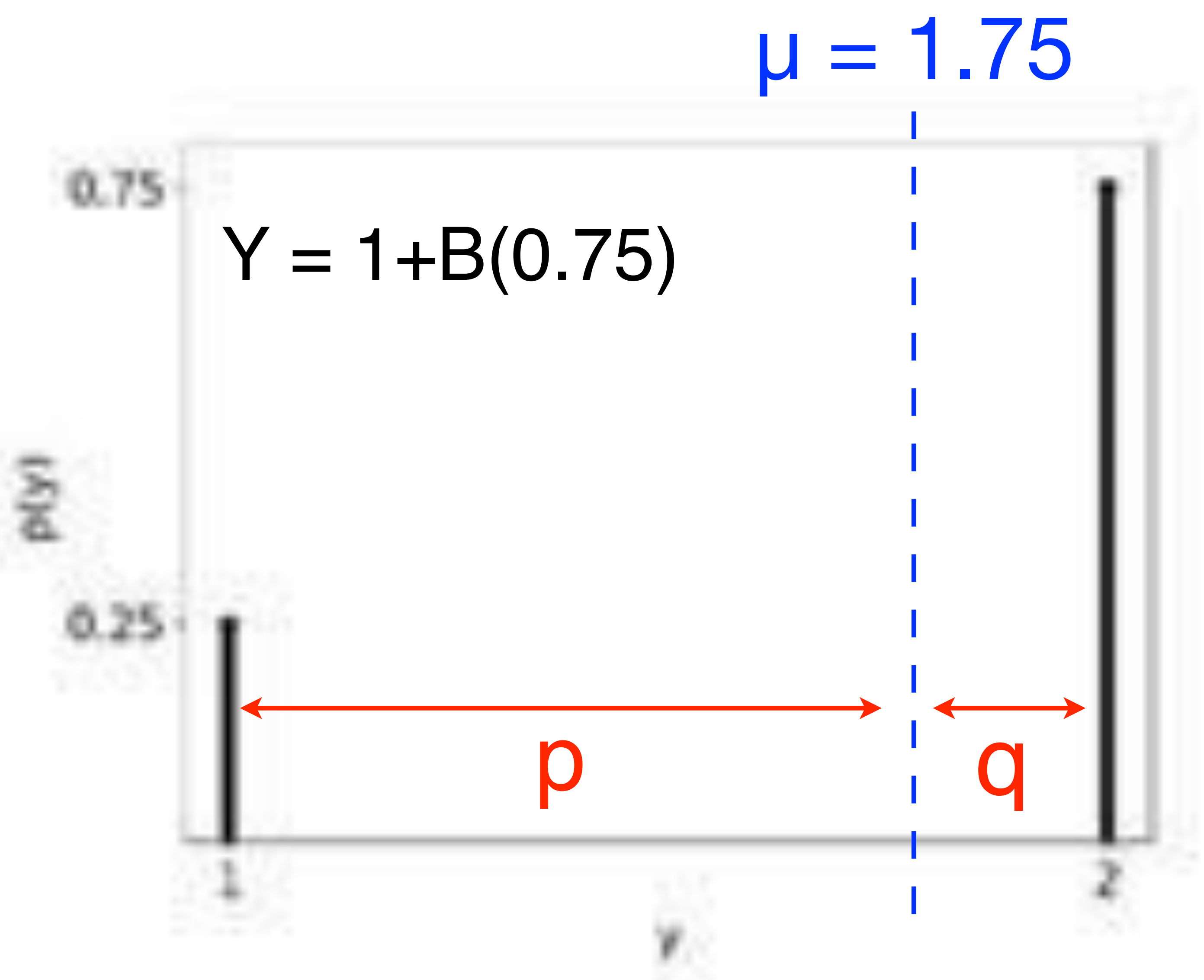


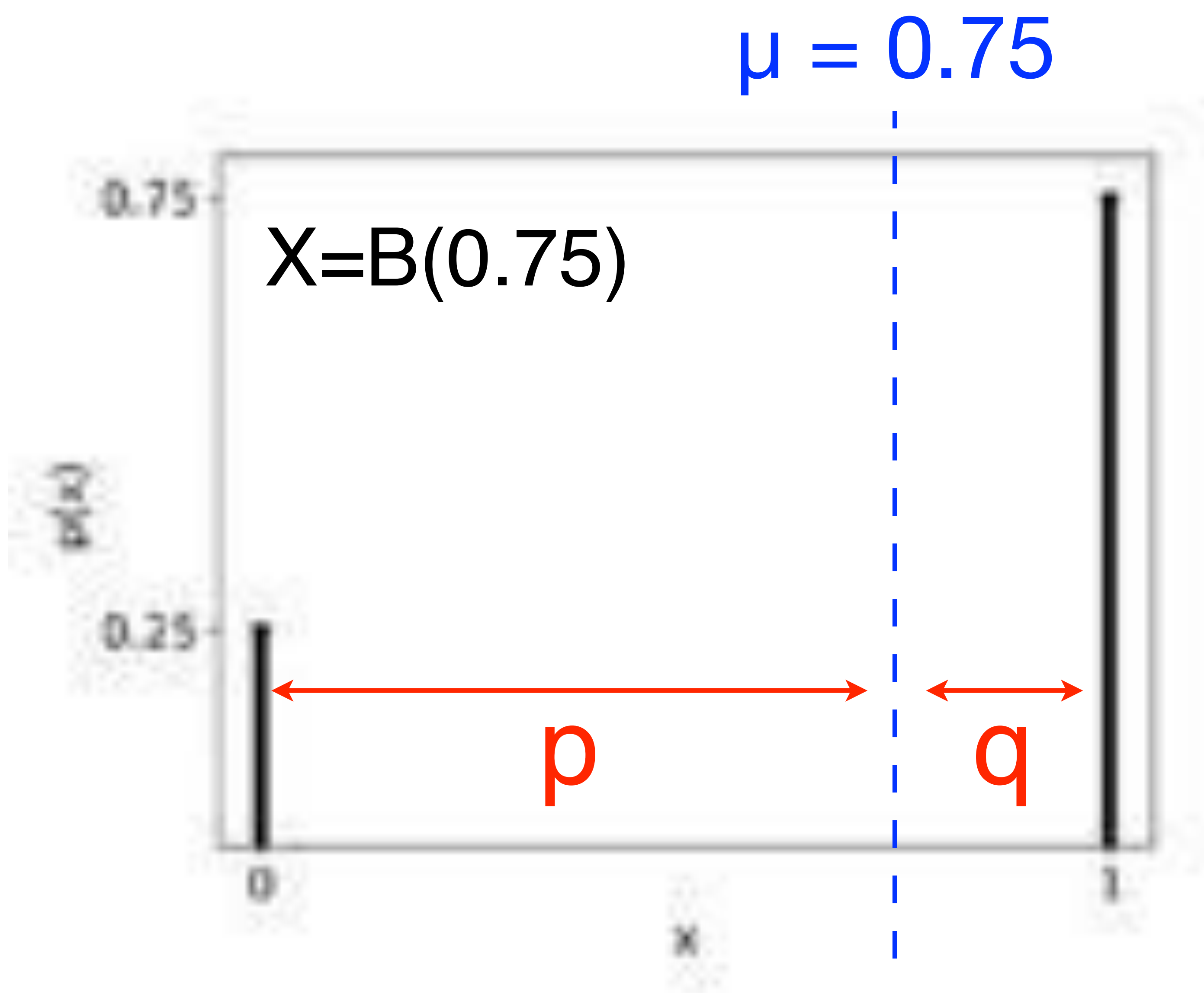
Translated B(p)

$X \sim B(p)$ $V(X) = p(1-p)$

$Y = X + 1$

y	p_y
1	$1-p$
2	p





$\mu_y = 1 + p$ (linearity of expectations)

$$V(Y) = E(Y - \mu_y)^2 = (1 - p)(1 - 1 - p)^2 + p(2 - 1 - p)^2$$
$$= (1 - p)p^2 + p(1 - p)^2 = p(1 - p)(p + 1 - p) = p(1 - p)$$
$$= V(X)$$ ✅

Scaling

$$V(aX) = E(aX - \mu_{ax})^2$$

$$\mu_{ax} = a\mu_X$$

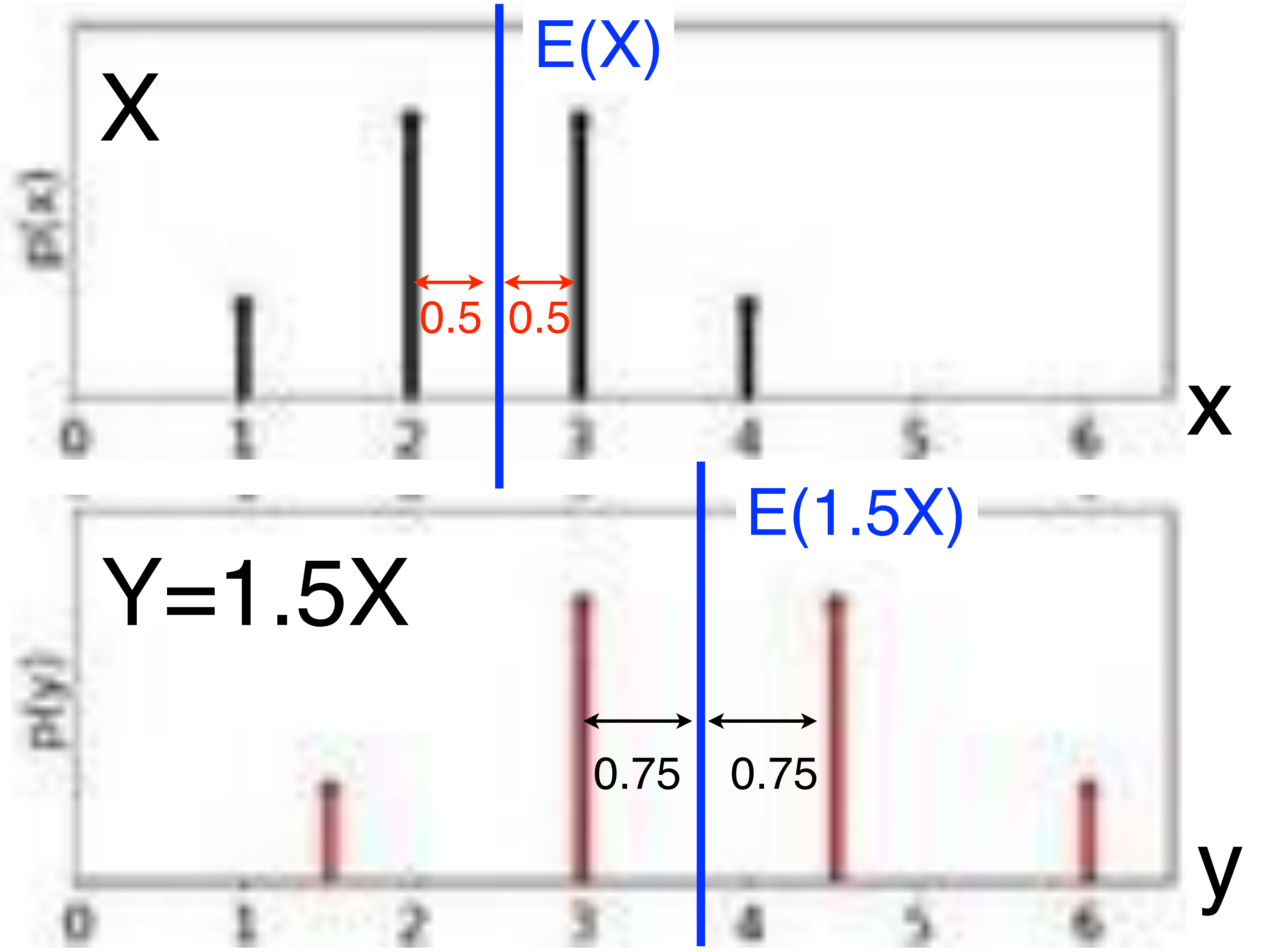
$$= E(aX - a\mu_x)^2$$

$$= E[a^2(X - \mu_x)^2]$$

$$= a^2 E(X - \mu_x)^2$$

$$= a^2 V(X)$$

Difference from mean grew by a^2



$$\sigma_{aX} = \sqrt{V(aX)} = \sqrt{a^2 V(X)} = |a|\sigma_x$$

"Average" difference from mean grew by a factor of $|a|$

Affine Transformation

$$V(aX + b) = V(aX) = a^2V(X)$$

$$\sigma_{ax+b} = |a|\sigma_x$$

This Lecture: Variance

Next: Two Variables

Two Variables

Why 2

Outcomes often result from multiple factors

Rain temperature and humidity

Economy unemployment and inflation

Hiring experience and salary

Student # classes GPA

Human condition profession age cholesterol happiness salary location dinner plans …

Two Fair Coins

$U,V \sim B(1/2)$ $\perp\!\!\!\perp$

Several ways to indicate distribution

Explicit $P(u,v) \stackrel{\text{def}}{=} P(U{=}u, V{=}v) = ¼ \quad \forall \{u,v\} \in \{0,1\}$

1-d table

u	v	P(u,v)
0	0	¼
0	1	¼
1	0	¼
1	1	¼

2-d table

u \ v	0	1
0	¼	¼
1	¼	¼

Use U, V, for several examples

Min - Max

$U,V \sim B(1/2)$ $\perp\!\!\!\perp$

$X = \min(U,V)$ $Y = \max(U,V)$

u	v	min	max	
0	0	0	0	¼
0	1	0	1	½
1	0	0	1	
1	1	1	1	¼

y = max

x = min

	0	1
0	¼	½
1	0	¼

Product - Sum

$X = U \cdot V$

$Y = U + V$

		y		
		0	**1**	**2**
x	**0**	¼	½	0
	1	0	0	¼

3 Coins

$U_1, U_2, U_3 \sim B(1/2)$ $\perp\!\!\!\perp$

$X = U_1 + U_2$ # heads among first 2

$Y = U_2 + U_3$ # heads among last 2

U_1	U_2	U_3	X	Y
0	0	0	0	0
0	0	1	0	1
0	1	0	1	1
0	1	1	1	2
1	0	0	1	0
1	0	1	1	1
1	1	0	2	1
1	1	1	2	2

X \ Y	0	1	2
0	⅛	⅛	0
1	⅛	¼	⅛
2	0	⅛	⅛

General B(p)

$U \sim B(p),\ V \sim B(q) \quad \perp\!\!\!\perp$

$X = \min(U,V) \qquad Y = \max(U,V)$

$$
\begin{array}{c|c|c|}
 & \multicolumn{2}{c}{v} \\
\hline
u & \bar{p}\bar{q} & \bar{p}q \\
\hline
 & p\bar{q} & pq \\
\hline
\end{array}
$$

$$
\begin{array}{cc|c|c|}
 & & \multicolumn{2}{c}{y = \max} \\
 & & 0 & 1 \\
\hline
x = \min & 0 & \bar{p}\bar{q} & p\bar{q}+\bar{p}q \\
\hline
 & 1 & 0 & pq \\
\hline
\end{array}
$$

General?

Joint Distribution

X, Y - random variables

Joint distribution: P: probability of every possible (x,y) pair

$$p(x,y) \stackrel{\text{def}}{=} P(X = x, Y = y)$$

$$\forall x,y \; p(x,y) \geq 0$$

$$\sum_{x,y} p(x,y) = 1$$

Joint Distribution Tells All

Joint distribution determines probabilities of all events

x \ y	0	1
0	0.1	0.2
1	0.3	0.4

$$P(X \leq Y) = P(X = 0, Y = 0) + P(X = 0, Y = 1) + P(X = 1, Y = 1)$$

$$= P(0, 0) + P(0, 1) + P(1, 1)$$

$$= 0.1 + 0.2 + 0.4$$

$$= 0.7$$

Marginals

Marginal of X $\quad P(x) \stackrel{\text{def}}{=} P_X(x) \stackrel{\text{def}}{=} P(X = x) = \sum_y p(x,y)$

Rule of total probability

Marginal of Y $\quad P(y) \stackrel{\text{def}}{=} P_Y(y) \stackrel{\text{def}}{=} P(Y = y) = \sum_x p(x,y)$

x \ y	0	1	
0	0.1	0.2	← P(X = 0) = .3
1	0.3	0.4	← P(X = 1) = .7

$P(X = 0) = P(X = 0, Y = 0) + P(X = 0, Y = 1)$

$= P(0,0)+P(0,1) \quad = .1 + .2 = .3$

Conditionals

$$P(x \mid y) = \frac{p(x,y)}{p(y)}$$

$$P(y \mid x) = \frac{p(x,y)}{p(x)}$$

x \ y	0	1
0	0.1	0.2
1	0.3	0.4

$P(X = 0) = 0.3$

$P(Y = 0) = 0.4$

$$P(Y = 0 \mid X = 0) = \frac{P(X = 0\ ,\ Y = 0)}{P(X = 0)} = \frac{0.1}{0.3} = \frac{1}{3}$$

$$P(Y = 1 \mid X = 0) = \frac{P(X = 0\ ,\ Y = 1)}{P(X = 0)} = \frac{0.2}{0.3} = \frac{2}{3}$$

$$P(X = 0 \mid Y = 0) = \frac{P(X = 0\ ,\ Y = 0)}{P(Y = 0)} = \frac{0.1}{0.4} = \frac{1}{4}$$

$$P(X = 1 \mid Y = 0) = 1 - P(X = 0\ ,\ Y = 0) = 1 - \frac{1}{4} = \frac{3}{4}$$

Independence

X, Y independent $\qquad X \perp\!\!\!\perp Y$

$\forall x,y \qquad p(y \mid x) = p(y)$

$p(x \mid y) = p(x)$

$p(x,y) = p(x) \cdot p(y) \leftarrow$ more robust

$$\begin{array}{cc|cc|l} & & \multicolumn{2}{c}{\overbrace{\qquad\qquad}^{P(y)}} & \\ & & 0.2 & 0.8 & y \\ \hline P(x) & 0.6 & 0.12 & 0.48 & \\ & 0.4 & 0.08 & 0.32 & \\ \hline & x & & & \end{array} \qquad \perp\!\!\!\perp$$

$$\begin{array}{cc|cc|l} & & \multicolumn{2}{c}{\overbrace{\qquad\qquad}^{P(y)}} & \\ & & 0.4 & 0.6 & y \\ \hline P(x) & 0.3 & 0.1 & 0.2 & \\ & 0.7 & 0.3 & 0.4 & \\ \hline & x & & & \end{array} \qquad \not\perp\!\!\!\perp$$

Independence Checks

Independent $\rightarrow$ rows proportional to each other

$\rightarrow$ columns proportional to each other

$X \sim B(½)$

$Y = X$

$$\begin{array}{cc|c|c|} & & \multicolumn{2}{c}{y} \\ & & 0 & 1 \\ \hline \multirow{2}{*}{x} & 0 & ½ & 0 \\ \cline{3-4} & 1 & 0 & ½ \\ \hline \end{array}$$

$\not\perp\!\!\!\perp$

$Y = 1 - X$

$$\begin{array}{cc|c|c|} & & \multicolumn{2}{c}{y} \\ & & 0 & 1 \\ \hline \multirow{2}{*}{x} & 0 & 0 & ½ \\ \cline{3-4} & 1 & ½ & 0 \\ \hline \end{array}$$

$\not\perp\!\!\!\perp$

Linearity of Expectation

Expectation

$$
\begin{aligned}
Eg(X) &= \sum_z z \cdot P(g(x) = z) \\
&= \sum_z z \sum_{x \in g^{-1}(z)} p(x) \\
&= \sum_z \sum_{x \in g^{-1}(z)} z \cdot p(x) \\
&= \sum_z \sum_{x \in g^{-1}(z)} g(x)p(x) \\
&= \sum_x g(x)p(x)
\end{aligned}
$$

$$p(x) \to p(x, y)$$

$$g(x) \to g(x, y)$$

$$\sum_x \to \sum_{x,y}$$

Linearity of Expectation

$$
\begin{aligned}
E(X+Y) &= \sum_x \sum_y (x+y) \cdot p(x,y) \\
&= \sum_x \sum_y x \cdot p(x,y) + \sum_x \sum_y y \cdot p(x,y) \\
&= \sum_x x \sum_y p(x,y) + \sum_y y \sum_x p(x,y) \\
&= \sum_x x \cdot p(x) + \sum_y y \cdot p(y) \\
&= EX + EY
\end{aligned}
$$

Expectation of sum =
sum of expectations

The Hat Problem

	H_1	H_2	H_3	H
1 2 3	1	1	1	3
1 3 2	1	0	0	1
2 1 3	0	0	1	1
2 3 1	0	0	0	0
3 1 2	0	0	0	0
3 2 1	0	1	0	1

$\mathbb{1}_i$ - indicator function i^{th} student caught their own hat

H - # students who caught their own hat

$$H = \sum_{i=1}^{n} \mathbb{1}_i$$

$\mathbb{1}_i$ - Bernoulli

$$P(\mathbb{1}_i = 1) = \frac{\#\text{ permutations of }(\sigma_1, \ldots, \sigma_n)\text{ when }\sigma_i = i}{\#\text{ permutations of }(\sigma_1, \ldots, \sigma_n)} = \frac{(n-1)!}{n!} = \frac{1}{n}$$

$$E(\mathbb{1}_i) = P(\mathbb{1}_i = 1) = \frac{1}{n}$$

$$E(H) = E\left(\sum_{i=1}^{n} \mathbb{1}_i\right) = \sum_{i=1}^{n} E(\mathbb{1}_i) = \sum_{i=1}^{n} \frac{1}{n} = 1$$

Coupon Collector Problem

Variance

Expectations add $E(X+Y) = EX + EY$

Do variances? $V(X+Y) \stackrel{?}{=} V(X) + V(Y)$

$$\begin{aligned} V(X+Y) &= E(X+Y)^2 - (E(X+Y))^2 \\ &= E(X^2 + 2XY + Y^2) - (EX + EY)^2 \\ &= EX^2 + 2E(XY) + EY^2 - (E^2X + 2EX \cdot EY + E^2Y) \\ &= EX^2 - E^2X + EY^2 - E^2Y + 2(E(XY) - EX \cdot EY) \\ &= V(X) + V(Y) + 2(E(XY) - EX \cdot EY) \end{aligned}$$

$$E(XY) = EX \cdot EY?$$

Do expectations multiply?

Linearity of Expectation

Covariance

Do Expectations Multiply?

$$E(XY) = \sum_{x,y} xy \cdot p(x,y)$$

$$E(XY) \stackrel{?}{=} EX \cdot EY$$

$$X = Y = \begin{cases} -1 & \frac{1}{2} \\ 1 & \frac{1}{2} \end{cases}$$

	-1	1	y
-1	½	0	½
1	0	½	½
x	½	½	

$$EX = EY = 0$$

$$EX \cdot EY = 0$$

$$E(XY) = EX^2 = E(1) = 1$$

$$E(XY) \neq EX \cdot EY$$

Expectations don't always multiply! Satisfy any relation?

Wild World of Product Expectations

For any α, β, γ ∃ X, Y with: $EX = \alpha$ $\quad EY = \beta$ $\quad E(XY) = \gamma$

$$Y' = X' = \begin{cases} -1 & \frac{1}{2} \\ +1 & \frac{1}{2} \end{cases} \qquad EX' = EY' = 0 \qquad E(X'Y') = E[(X')^2] = 1$$

$$X = (\gamma - \alpha\beta)X' + \alpha \qquad Y = Y' + \beta$$

$$EX = \alpha \qquad EY = \beta$$

$$\begin{aligned} E(XY) &= E((\gamma - \alpha\beta)X' + \alpha)(Y' + \beta) \\ &= (\gamma - \alpha\beta)\underbrace{E(X'Y')}_{1} + \alpha \underbrace{EY'}_{0} + (\gamma - \alpha\beta)\beta \underbrace{EX'}_{0} + \alpha\beta \\ &= \gamma \end{aligned}$$

Can we still say something about E(XY)?

Covariance

Sufficient, and easier, to understand 0-mean variables

"Centralize" X, Y, consider expectation of centralized product

$$\sigma_{X,Y} \triangleq \mathrm{Cov}(X,Y) \triangleq E[(X-\mu_X)\cdot(Y-\mu_Y)]$$

$$= E(XY) - E(X\mu_Y) - E(\mu_X Y) + E(\mu_X\mu_Y)$$

$$= E(XY) - E(X)\mu_Y - \mu_X E(Y) + \mu_X\mu_Y$$

$$= E(XY) - \mu_X\mu_Y$$

If seems complex, think of E(XY) for 0-mean variables

Amount X and Y vary together

Properties

$$\mathrm{Cov}(X,X) = EX^2 - \mu_X^2 = V(X)$$

$$\mathrm{Cov}(X,Y) = E((X-\mu_X)(Y-\mu_Y)) = \mathrm{Cov}(Y,X)$$

$$\mathrm{Cov}(aX,Y) = E(aXY) - \mu_{aX}\mu_Y = aE(XY) - a\mu_X\mu_Y = a\mathrm{Cov}(X,Y)$$

$$\begin{aligned}\mathrm{Cov}(X+a,y) &= E[((X+a)-\mu_{X+a})(Y-\mu_Y)] \\ &= E(X-\mu_X)(Y-\mu_Y) = \mathrm{Cov}(X,Y)\end{aligned}$$

Intuitively if X changes by σ_X, Y grows by $\sigma_{X,Y} \cdot \sigma_X \cdot \sigma_Y$

Correlation Coefficient

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Properties:

$$\rho_{X,X} = 1 \qquad \rho_{X,-X} = -1$$

$$\rho_{X,Y} = \rho_{Y,X}$$

$$\rho_{aX+b,cY+d} = \text{sign}(ac) \cdot \rho_{X,Y}$$

$$\text{sign}(x) = \begin{cases} 1 & x > 0 \\ 0 & 0 \\ -1 & x < 0 \end{cases}$$

If X $\nearrow$ by σ_X, by how many σ_Y do we expect Y to $\nearrow$

Bounds on $\rho_{X,Y}$?

Cauchy-Schwarz Inequality

$E(X \cdot Y)$ can't take all possible values

$$|E(XY)| \leq \sqrt{EX^2} \cdot \sqrt{EY^2}$$

For any α

$$0 \leq E(\alpha X + Y)^2 = \alpha^2 EX^2 + 2\alpha E(XY) + EY^2$$

True for all α, so discriminant must be negative

$$4(EXY)^2 - 4EX^2 \cdot EY^2 \leq 0$$

$$(EXY)^2 \leq EX^2 \cdot EY^2$$

Correlation Coefficient

$$|E(X-\mu_X)(Y-\mu_Y)| \le \sqrt{E(X-\mu_X)^2 \cdot E(Y-\mu_Y)^2}$$

Namely

$$|\sigma_{X,Y}| \le \sigma_X \cdot \sigma_Y$$

$$\rho_{X,Y} \triangleq \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$$

$$|\rho_{X,Y}| \le 1$$

Examples

Uncorrelated: EXY =0

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

Correlation			
Positive	$X, X+Y$	$X, 2X+Y$	$\min(X,Y), \max(X,Y)$
Uncorrelated	X, Y	$3X, 4Y$	
Negative	$X, -Y$	$Y, -X$	$\lvert X-Y \rvert, \min(X,Y)$

$X = 3Y$ $\quad$ $\mathrm{Cov}(X,Y) = 3Var(X)$ $\quad$ $P = 1$

$$\perp\!\!\!\perp \rightarrow \perp$$

Independent implies uncorrelated

$$
\begin{aligned}
E(XY) &= \sum_x \sum_y xy \cdot p(x, y) \\
&= \sum_x \sum_y xy \cdot p(x)p(y) \\
&= \sum_x x \cdot p(x) \sum_y y \cdot p(y) \\
&= E(X) \cdot E(Y)
\end{aligned}
$$

$$\perp \not\leftrightarrow \perp\!\!\!\perp$$

Independent $\rightarrow$ uncorrelated

Uncorrelated $\stackrel{?}{\rightarrow}$ independent

$$X = \begin{cases} -1 & \frac{1}{2} \\ +1 & \frac{1}{2} \end{cases}$$

$$X = -1 \rightarrow \; Y = 0$$

$$X = +1 \rightarrow \; Y = \begin{cases} +1 \\ -1 \end{cases}$$

X \ Y	-1	0	1
-1		½	
1	¼		¼

Uncorrelated

$$EX = 0 \qquad EY = 0$$

$$E(XY) = \frac{1}{4} \cdot -1 + \frac{1}{2} \cdot 0 + \frac{1}{4} \cdot 1 = 0 = EX \cdot EY$$

Clearly dependent

Note: Uncorrelated binary random pairs are independent

Variance

$$V(X+Y) \stackrel{?}{=} V(X) + V(Y)$$

$$\begin{aligned}
V(X+Y) &= E(X+Y)^2 - (E(X+Y))^2 \\
&= E(X^2 + 2XY + Y^2) - (EX + EY)^2 \\
&= EX^2 + 2E(XY) + EY^2 - (E^2X + 2EX \cdot EY + E^2Y) \\
&= EX^2 - \ E^2X \ + EY^2 - \ E^2Y \ + 2(E(XY) - EX \cdot EY) \\
&= V(X) + V(Y) + 2(E(XY) - EX \cdot EY) \\
&= V(X) + V(Y) + 2\mathrm{Cov}(X,Y)
\end{aligned}$$

$$= \text{iff } \mathrm{Cov}(X,Y) = 0 \quad \text{Uncorrelated}$$