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

$E\,|\,X - \mu\,|$

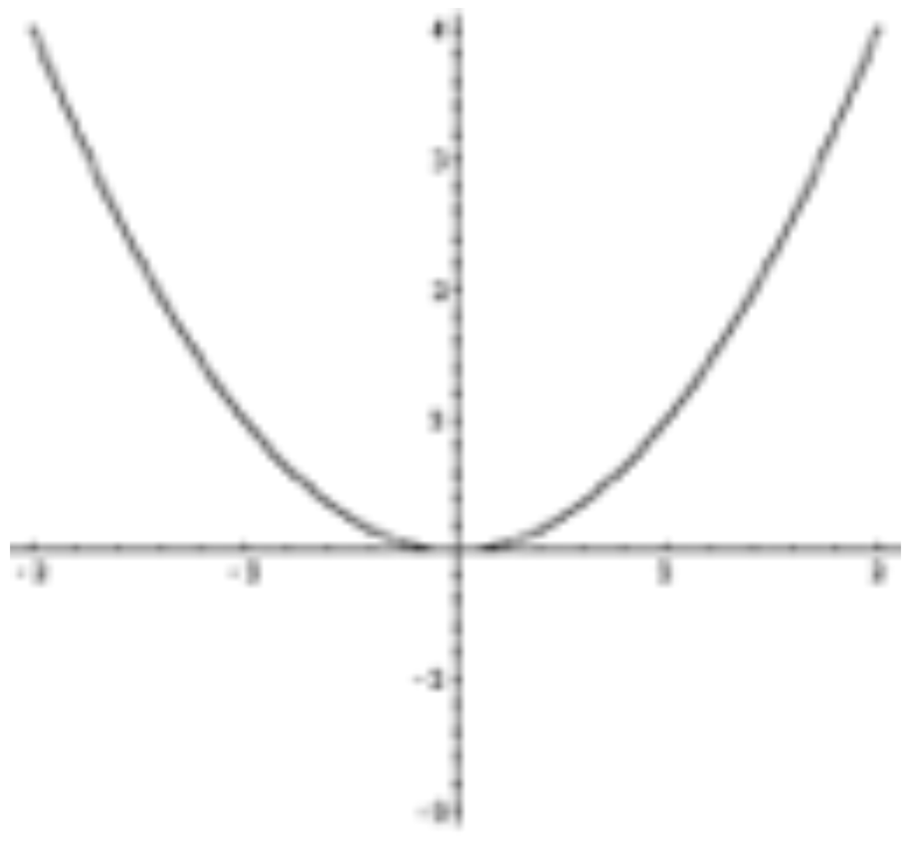
Mean absolute difference

Not commonly used

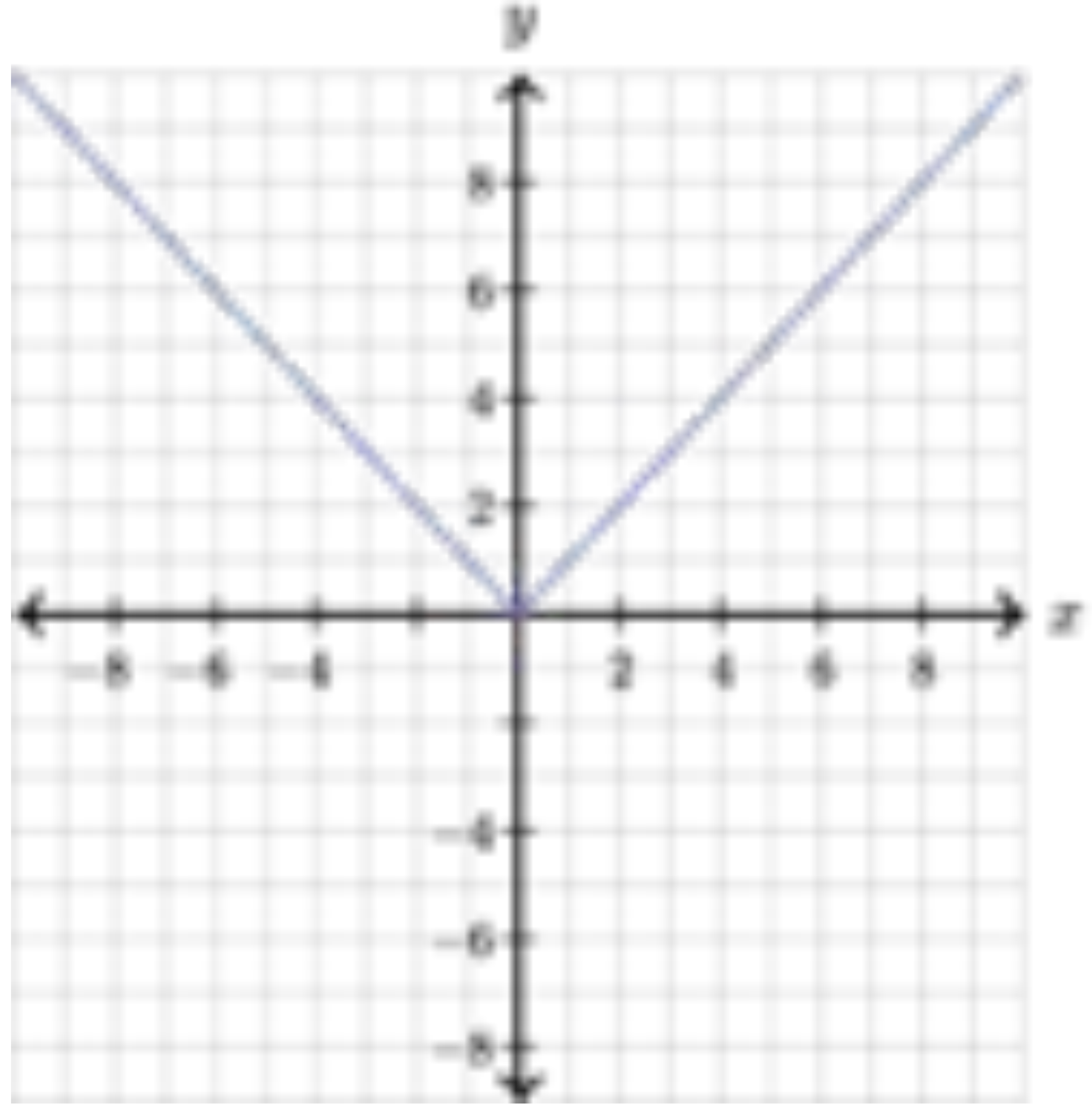
Absolute value function hard to analyze

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)

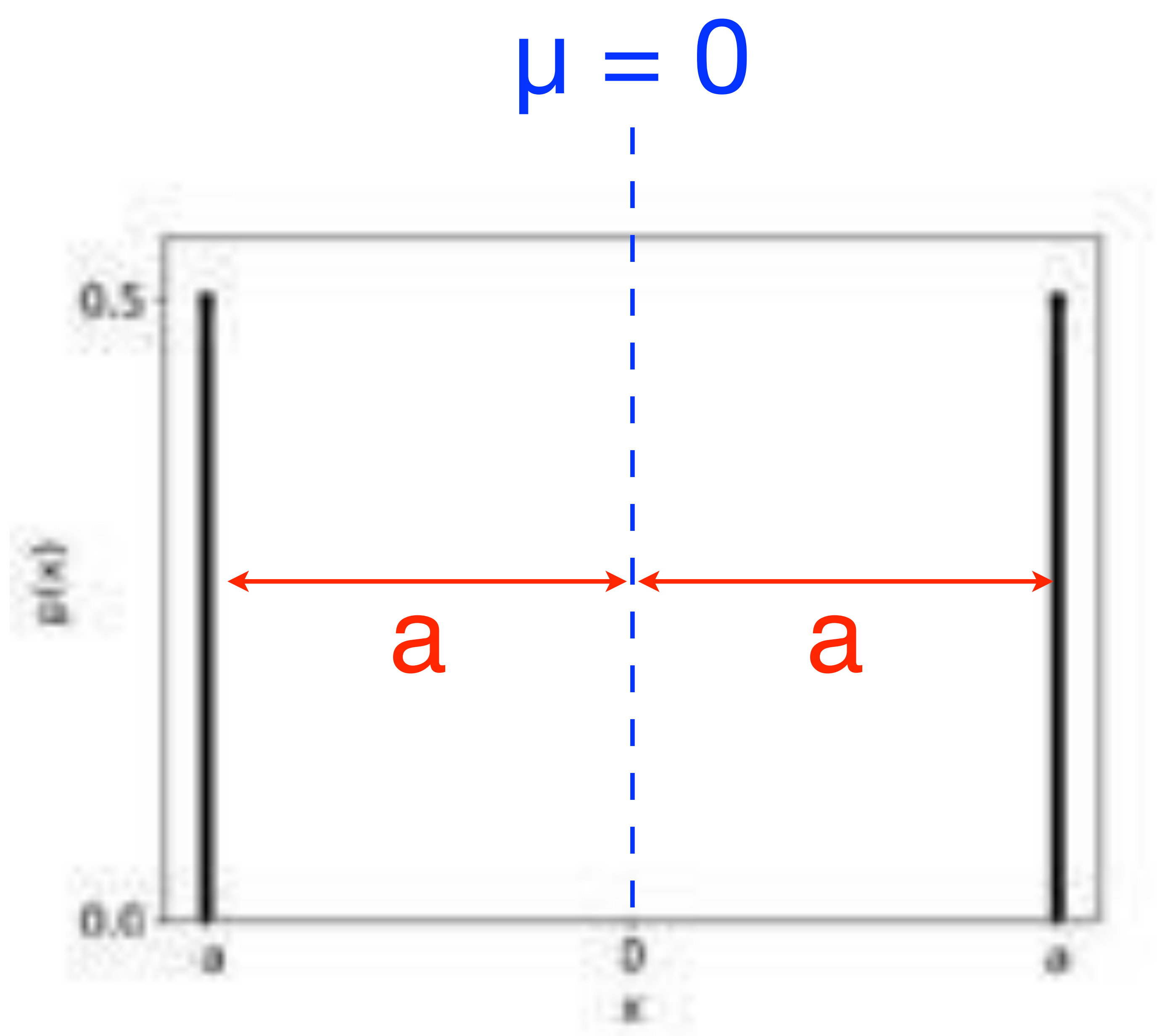
Constants

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

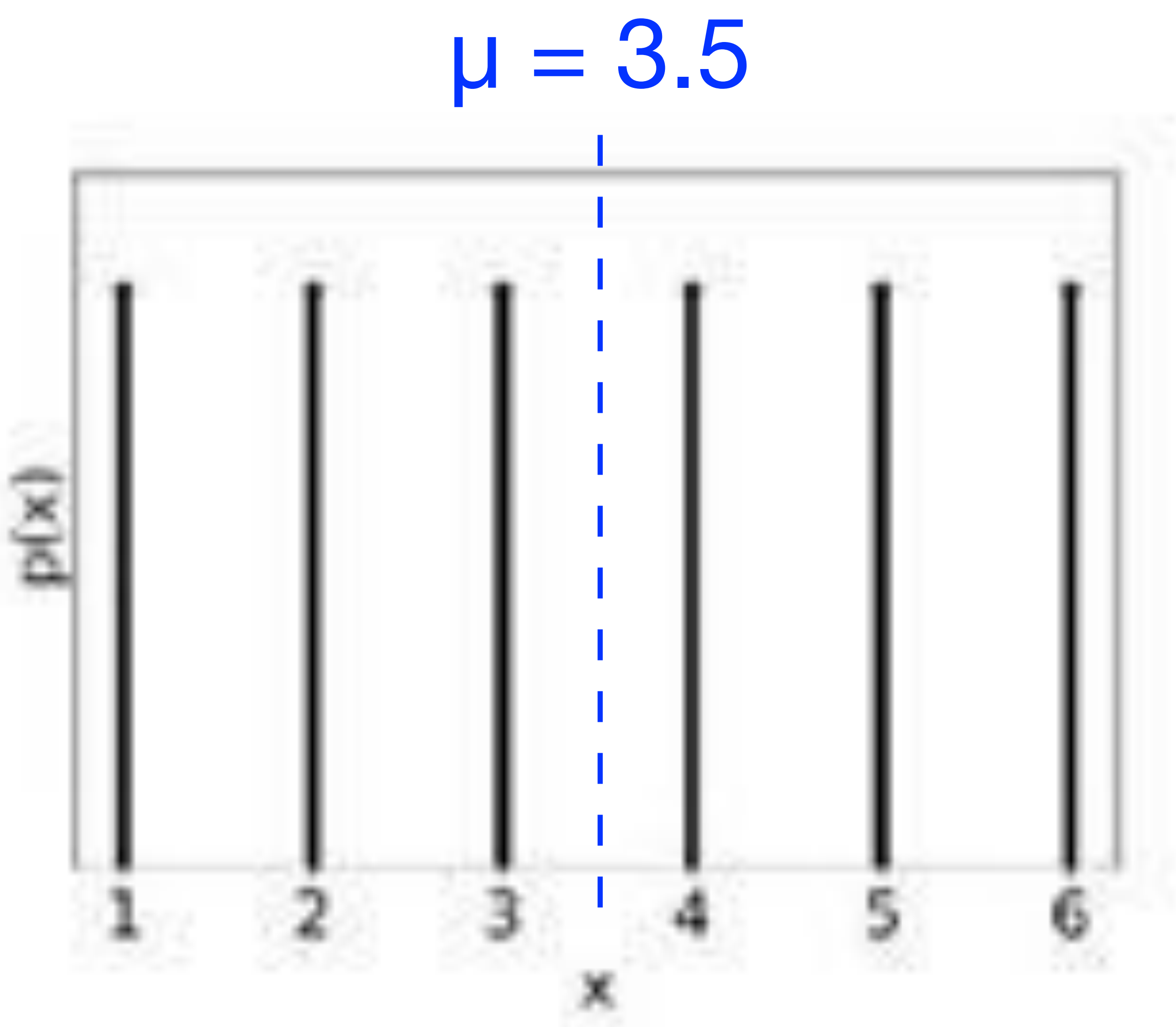
Fair Die

$\mu = 3.5$

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

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

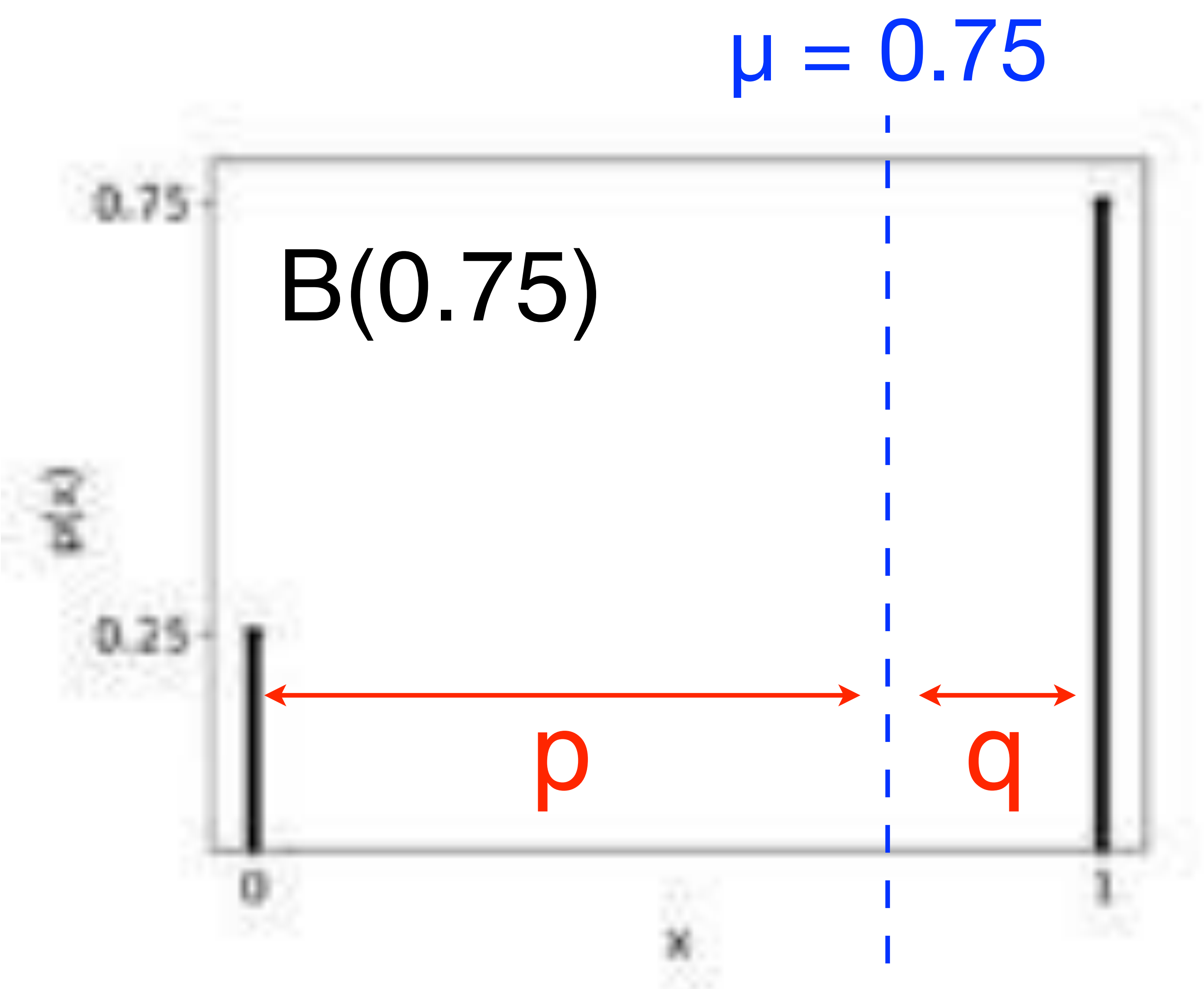
$$\sigma = \sqrt{2.92\ldots} = 1.71\ldots$$



Bernoulli p

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

$\mu = p$



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

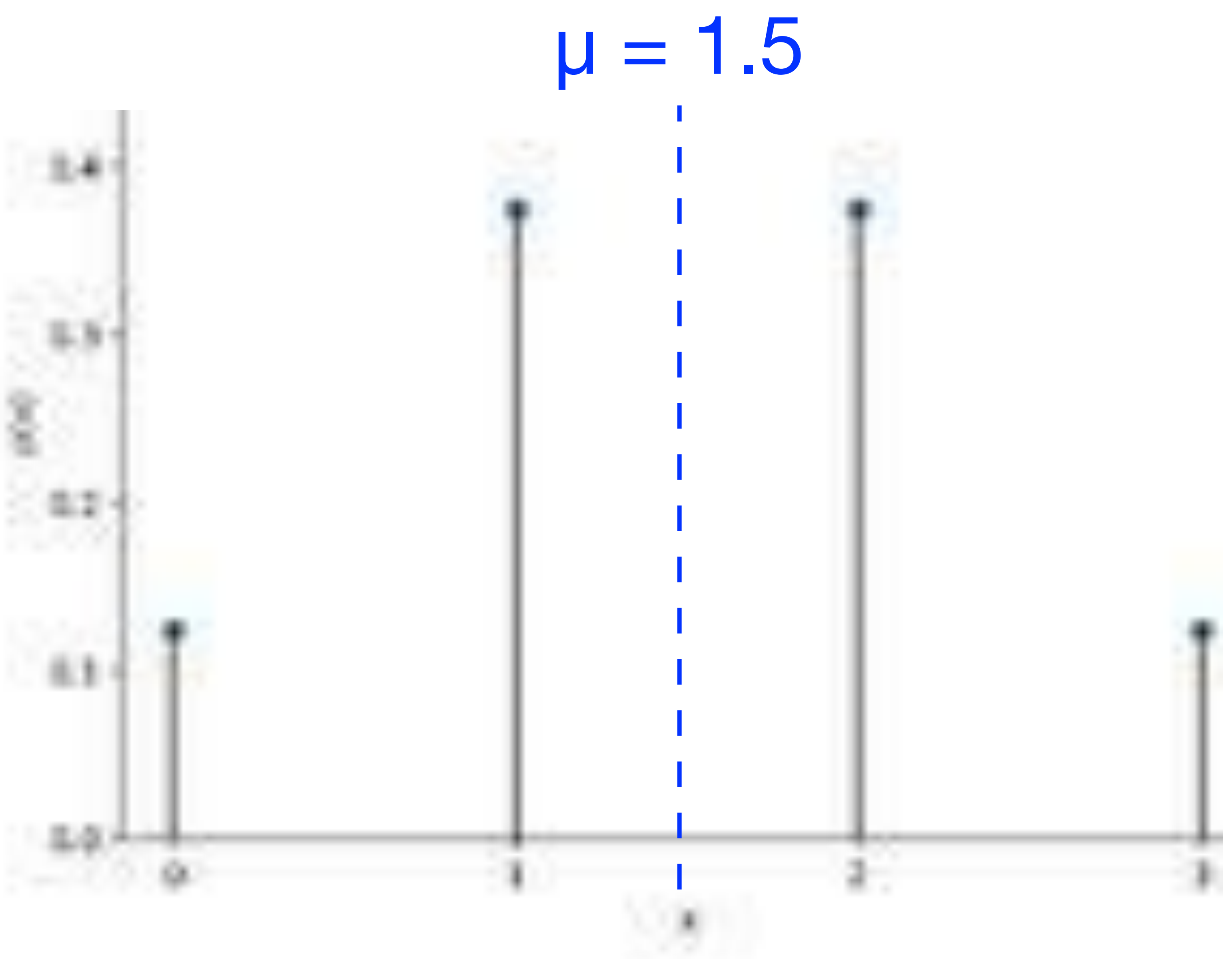
3 Coins

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

$\mu = 1.5$



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

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

Shortly: simpler derivation

Different Formula

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

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

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

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

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

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

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

$E(X)=\mu$

2, μ - constants

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

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

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

=
X is a
constant

=
X constant or
takes two values
with equal prob.

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

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

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

Properties

How simple modification affect V and σ

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

Addition

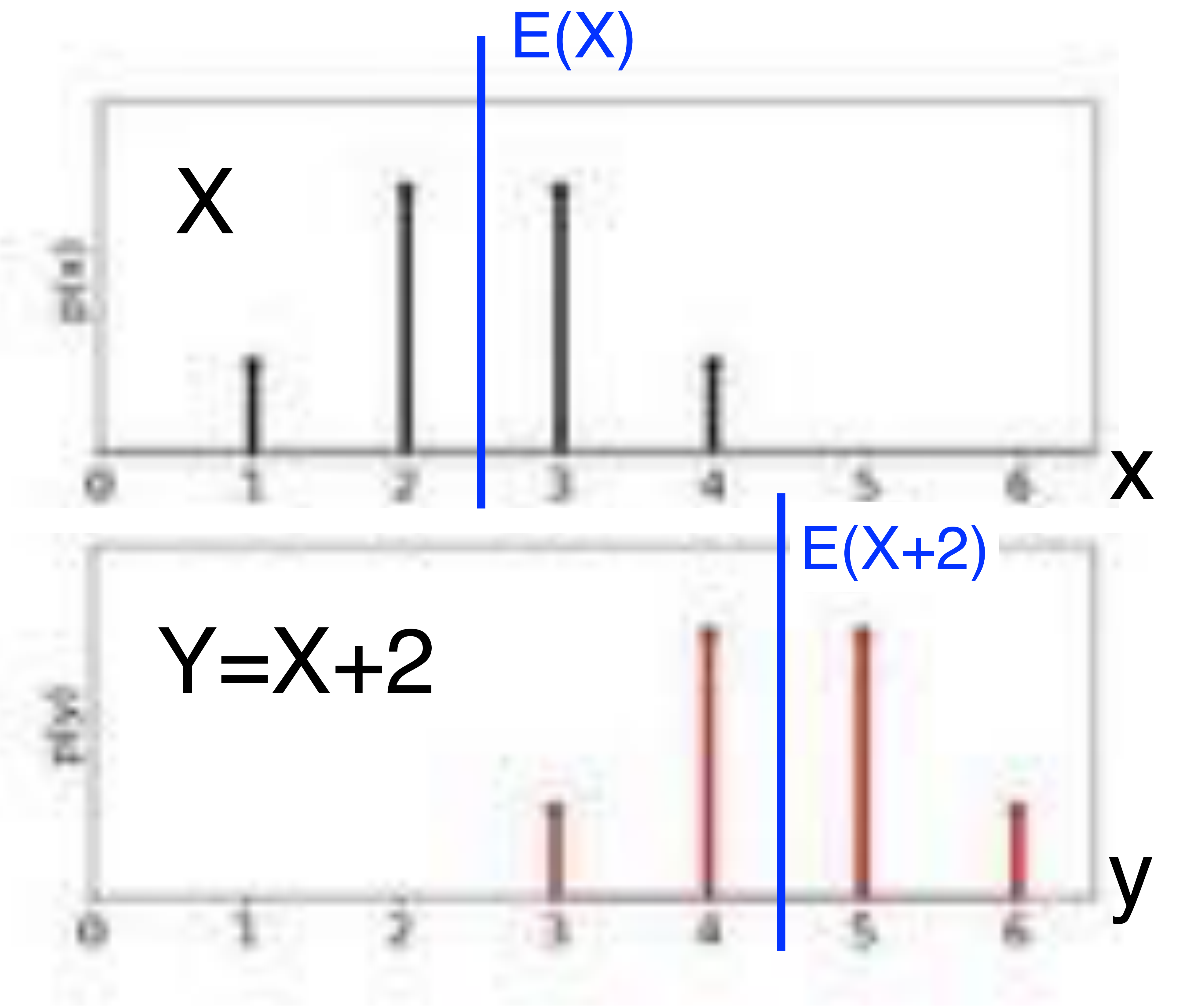
X - random variable

b - constant (e.g. 2)

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

Linearity of expectation

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

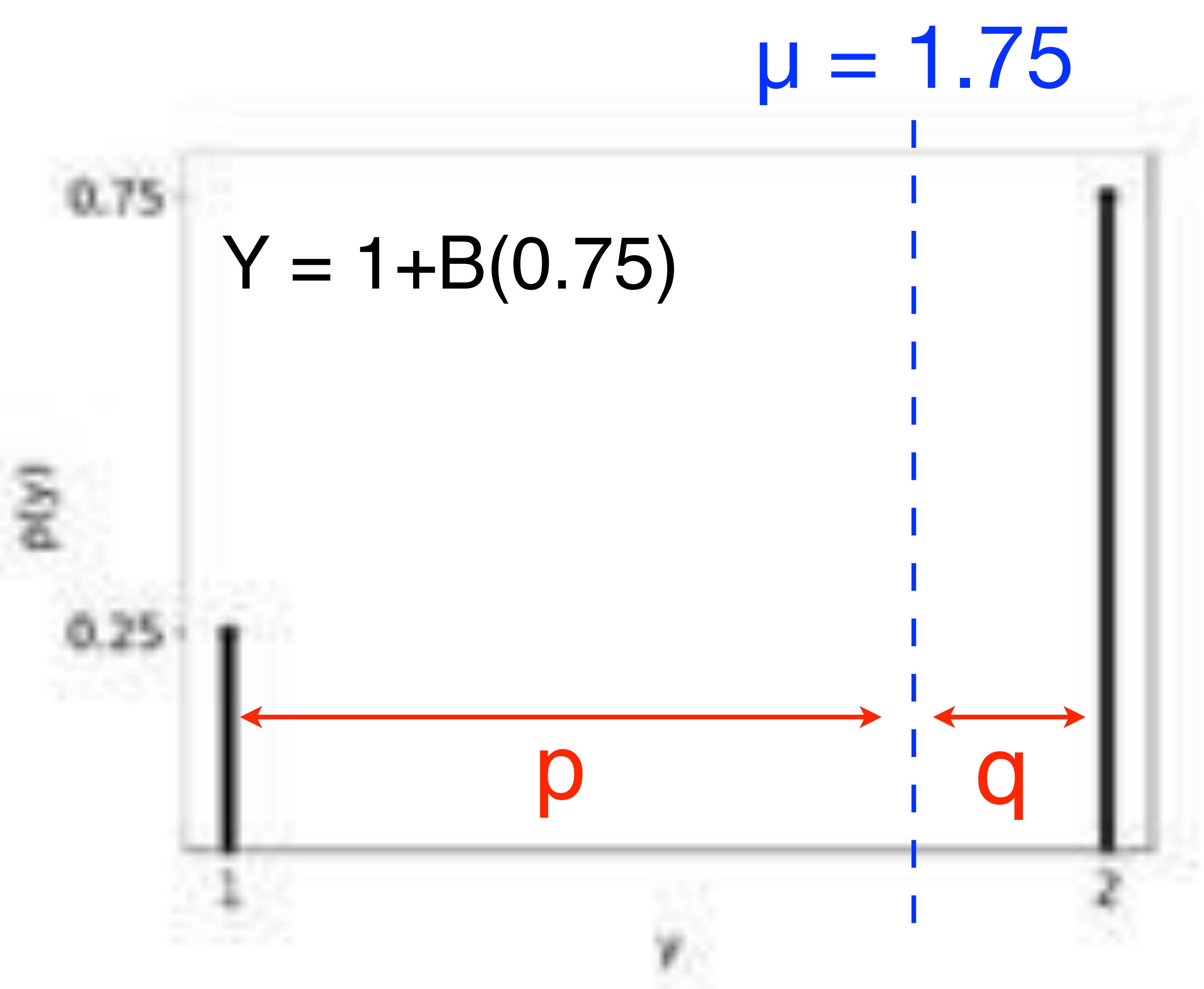


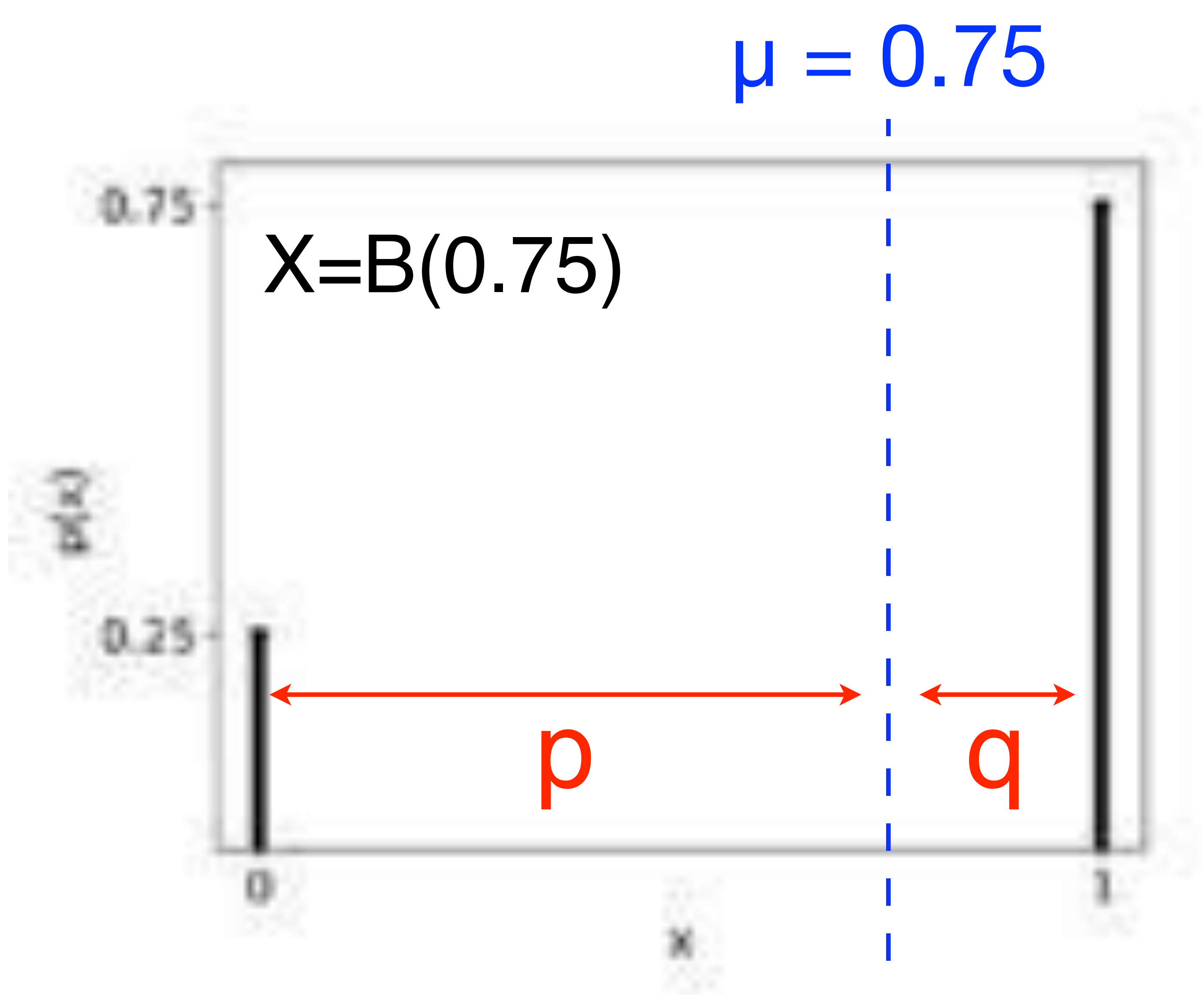
Translated B(p)

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

$Y = X + 1$

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

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

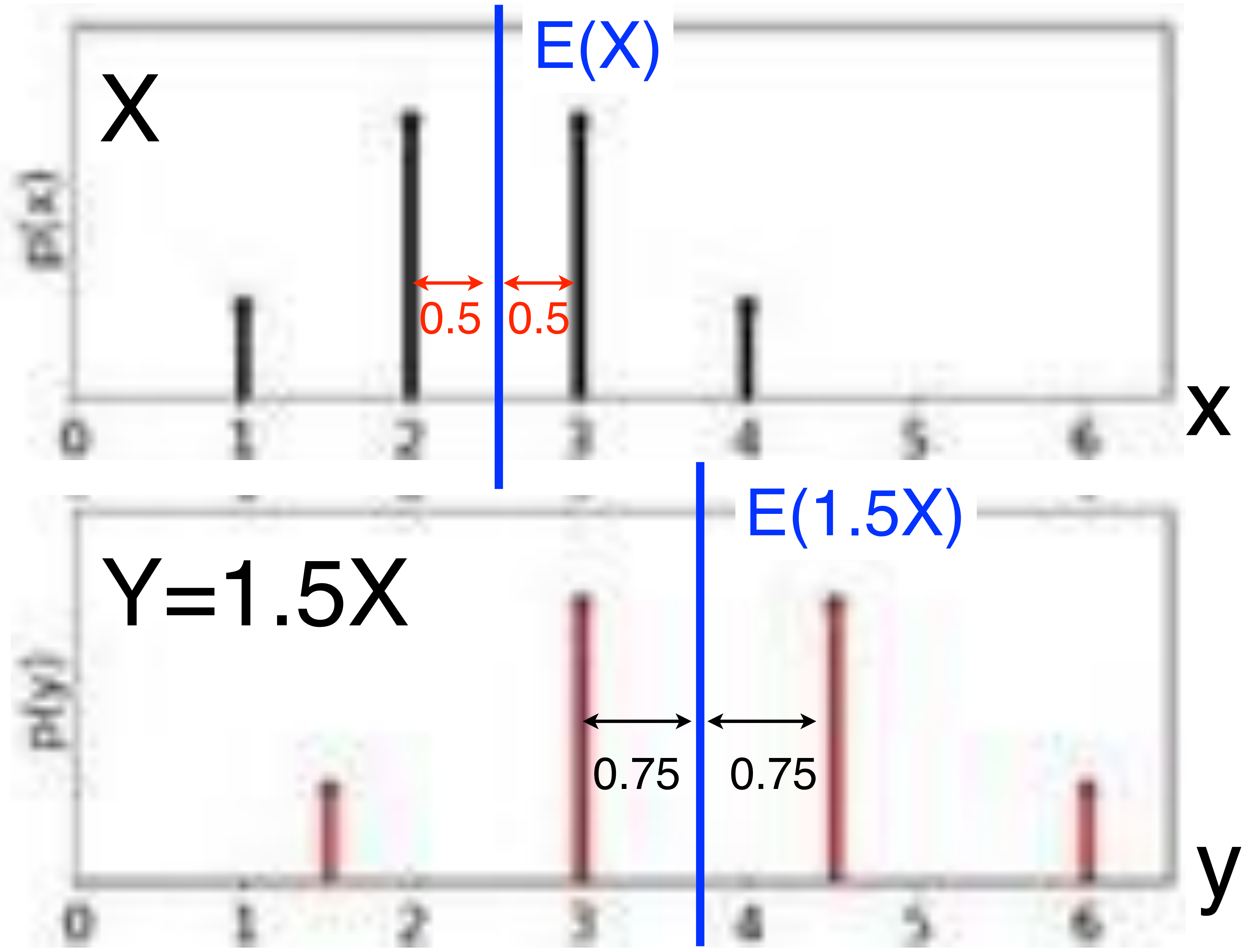
$$\mu_{ax} = a\mu_X$$

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

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

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

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



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, salary, happiness, 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
0	1	0	1
1	0	0	1
1	1	1	1

¼

½

¼

y = max

x = min

	0	1
0	¼	½
1	0	¼

Product - Sum

$X = U \cdot V$

$Y = U + V$

x \ y	0	1	2
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)

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

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

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

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

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	$\leftarrow P(X = 0) = .3$
1	0.3	0.4	$\leftarrow P(X = 1) = .7$

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

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

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

$\leftarrow P(X = 0) = 0.3$

$\uparrow P(Y = 0) = 0.4$

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

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

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

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

Independence

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

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

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

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

$$\begin{array}{cc|c|c|l} & & \multicolumn{2}{c}{\overbrace{\qquad\qquad}^{P(y)}} & \\ & & 0.2 & 0.8 & y \\ \cline{3-4} \multirow{2}{*}{$P(x)$} & 0.6 & 0.12 & 0.48 & \\ \cline{3-4} & 0.4 & 0.08 & 0.32 & \\ \cline{3-4} & x & & & \end{array} \qquad \perp\!\!\!\perp$$

$$\begin{array}{cc|c|c|l} & & \multicolumn{2}{c}{\overbrace{\qquad\qquad}^{P(y)}} & \\ & & 0.4 & 0.6 & y \\ \cline{3-4} \multirow{2}{*}{$P(x)$} & 0.3 & 0.1 & 0.2 & \\ \cline{3-4} & 0.7 & 0.3 & 0.4 & \\ \cline{3-4} & x & & & \end{array} \qquad \cancel{\perp\!\!\!\perp}$$

Independence Checks

Independent → rows proportional to each other

→ columns proportional to each other

$X \sim B(½)$

$Y = X$

		y = 0	y = 1
x	0	½	0
	1	0	½

$\cancel{\perp\!\!\!\perp}$

$Y = 1 - X$

		y = 0	y = 1
x	0	0	½
	1	½	0

$\cancel{\perp\!\!\!\perp}$

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:

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

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

$$\rho_{aX+b,cY+d} = \text{sign}(ac) \cdot \rho_{X,Y} \qquad \text{sign}(x) = \begin{cases} 1 & x > 0 \\ 0 & 0 \\ -1 & x < 0 \end{cases}$$

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 $\to$ uncorrelated

Uncorrelated $\overset{?}{\to}$ independent

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

$$X = -1 \to \ Y = 0$$

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

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

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

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

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