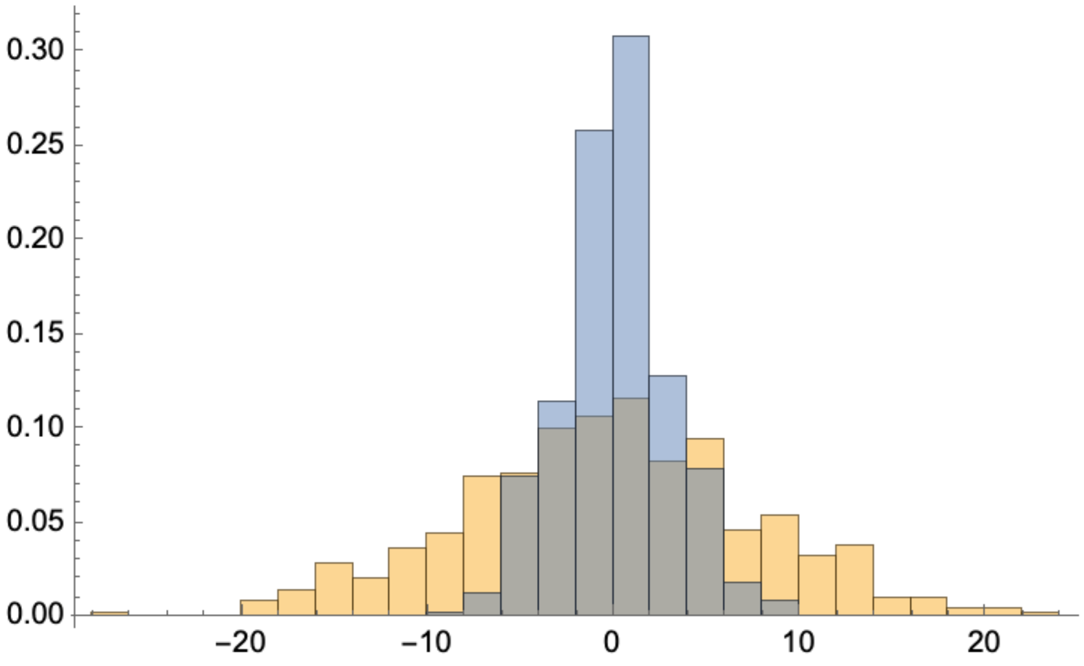
Lecture 18: Variance and Covariance

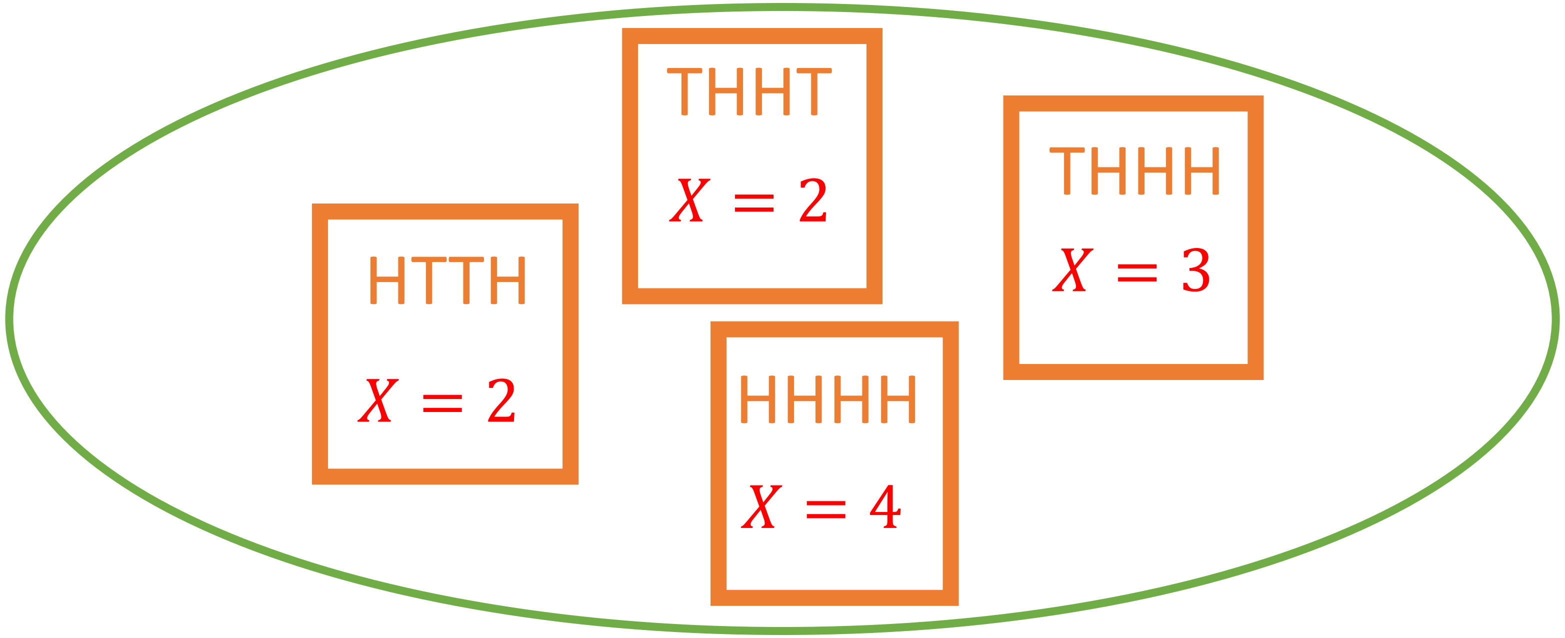


Recap: Random Variable

Definition (Random Variable)

A random variable X is a function $X: \Omega \to \mathbb{R}$.

For every outcome ω, it has a value $X(\omega)$.

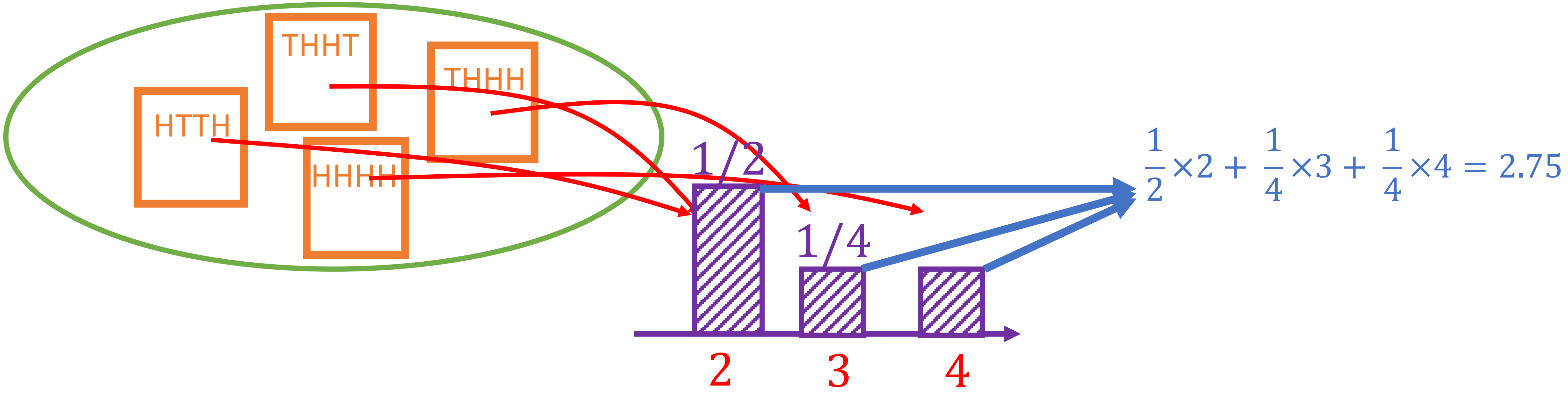


Recap: Expectation

Definition.

The expectation of a random variable X is defined as,

$$\mathbb{E}[X] = \sum_a \mathbb{P}[X = a] \cdot a\,.$$



Recap: Function of random variable

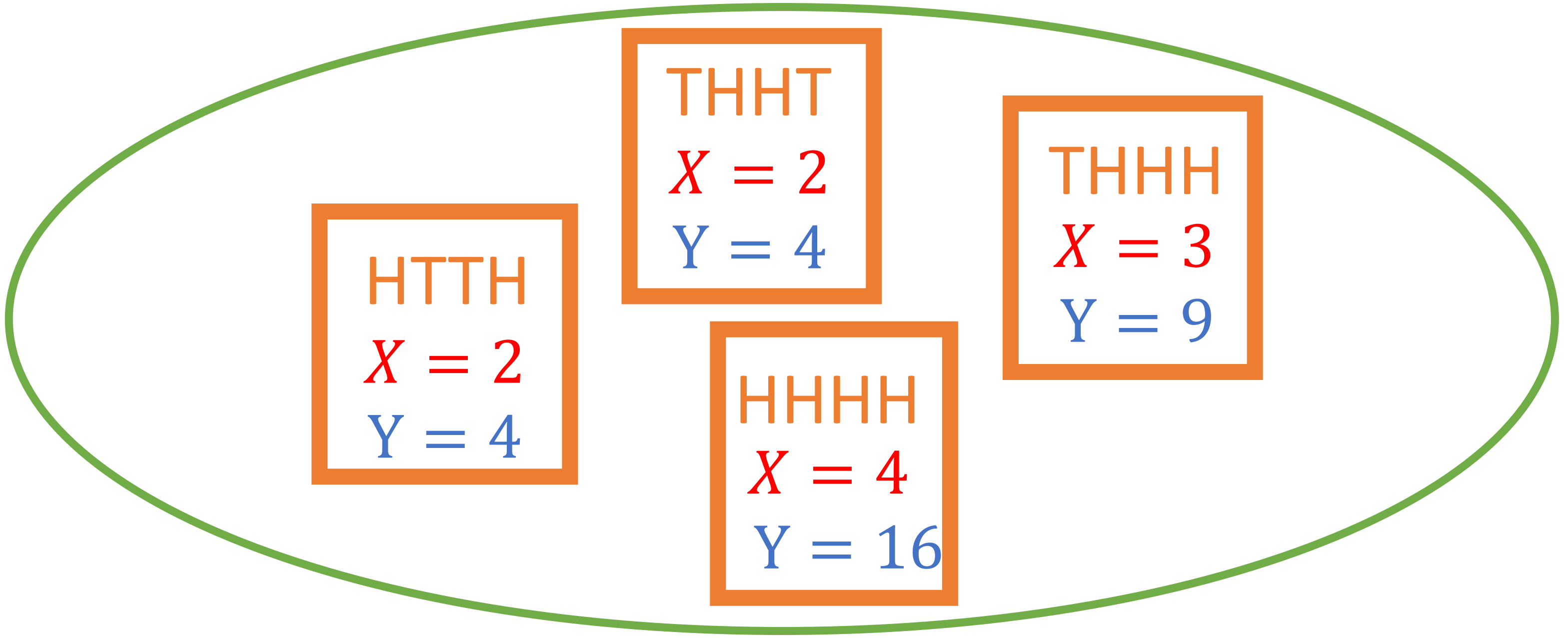
Definition.

If $X: \Omega \to \mathbb{R}$ is a random variable, then we can define another random variable $Y = f(X)$ for function f.

For every outcome ω, it has a value. $f(X(\omega))$.

Example.

$Y = X^2$



Recap: Function of random variable

Definition.

If $X: \Omega \rightarrow \mathbb{R}$ is a random variable, then we can define another random variable $Y = f(X)$ for function f.

We can also talk about the expectation of that random variable.

$$\mathbb{E}[f(X)] = \sum_a f(a) \cdot \mathbb{P}[X = a]$$

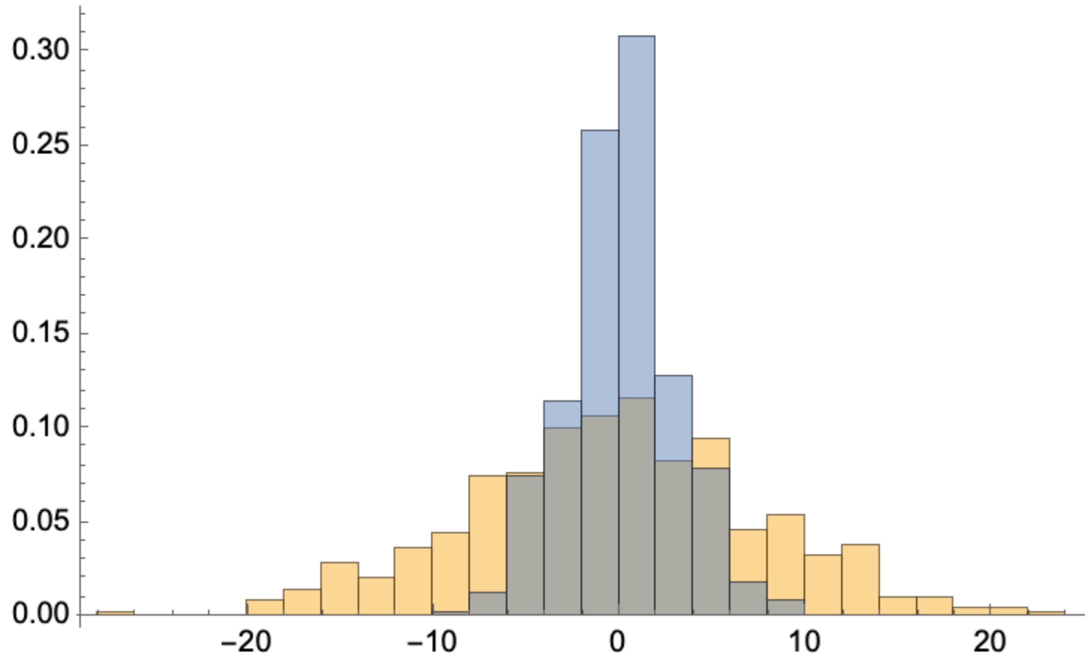
Variance

Intuition.

Say there are two distributions both of expectation 0.

The Blue distribution is more concentrated than the Orange one.

The Orange one is more uniform / spread out / uncertain….



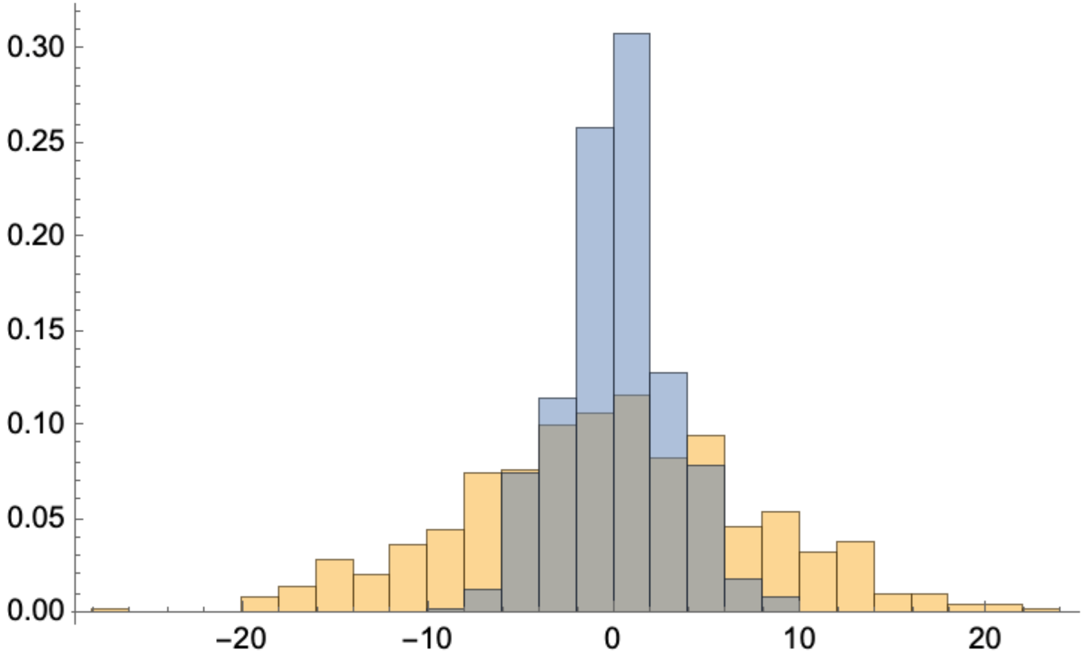
Variance

Intuition.

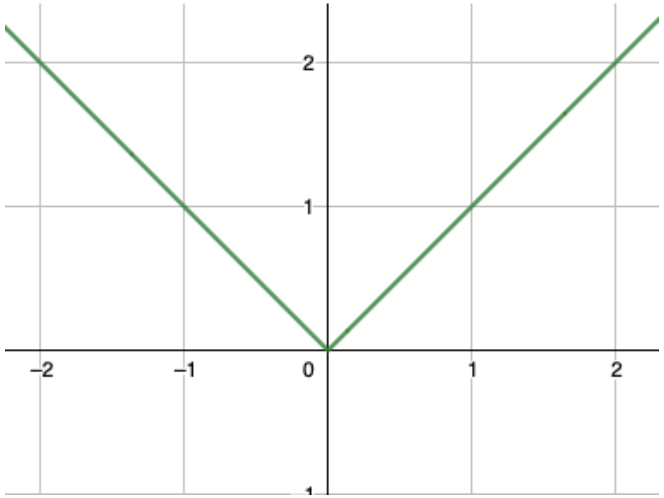
How do we characterize this?

deviation from expectation: $X - \mathbb{E}[X]$

$$\mathbb{E}[X - \mathbb{E}[X]] = 0$$



Variance

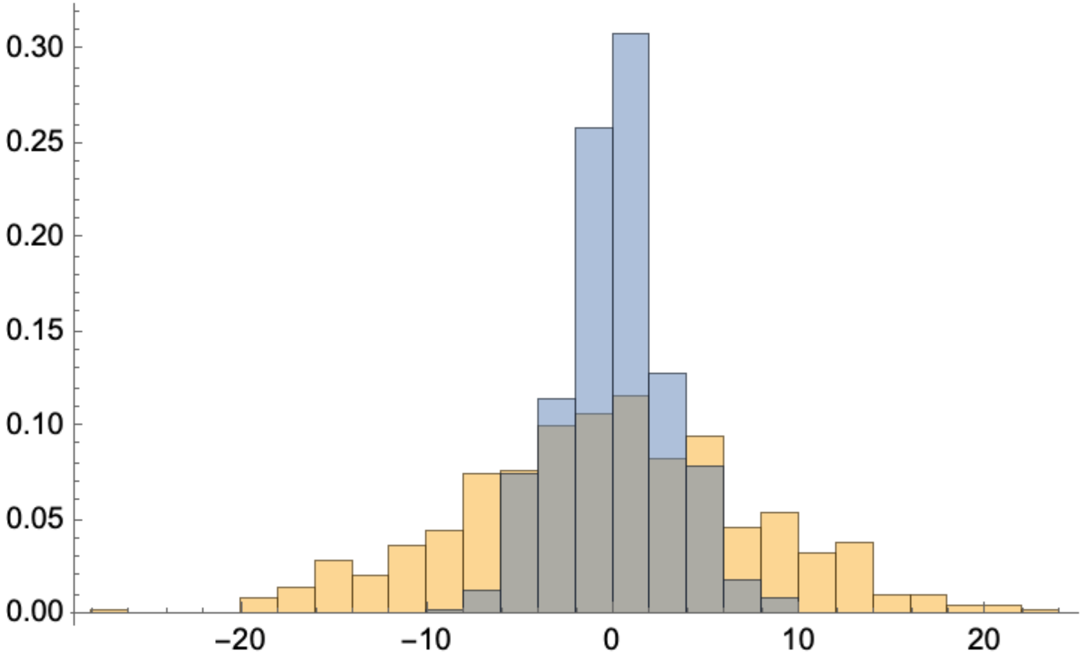


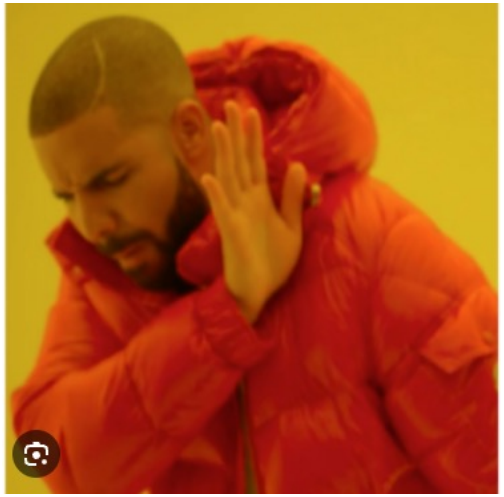
Intuition.

How do we characterize this?

deviation from expectation: $X - \mathbb{E}[X]$

$\mathbb{E}[|X - \mathbb{E}[X]|]$? Hard to take derivative..





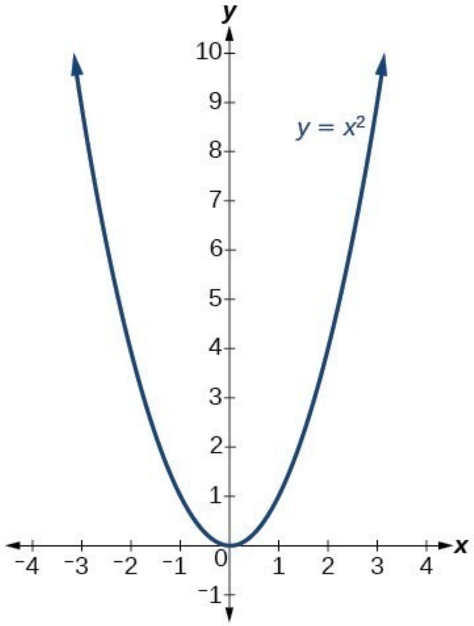
Variance

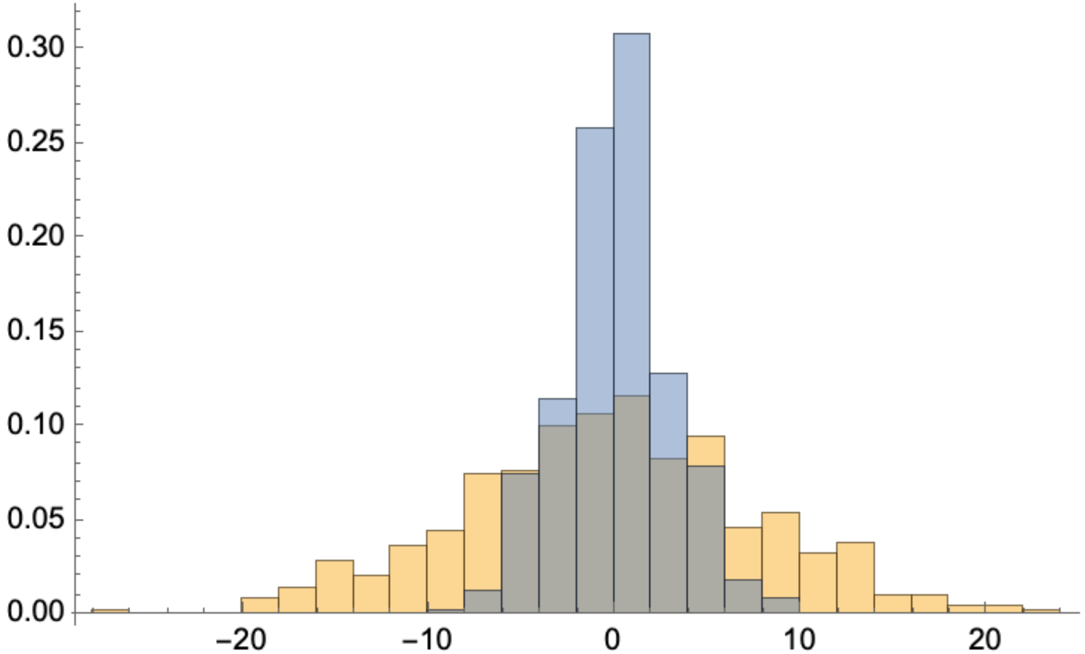
Intuition.

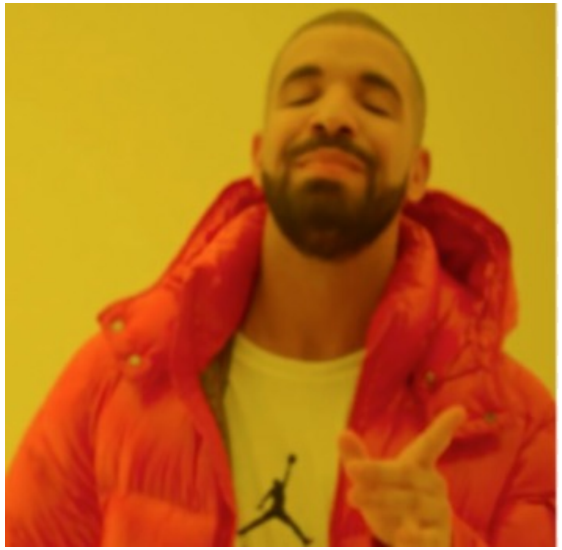
How do we characterize this?

deviation from expectation: $X - \mathbb{E}[X]$

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







Variance

Definition.

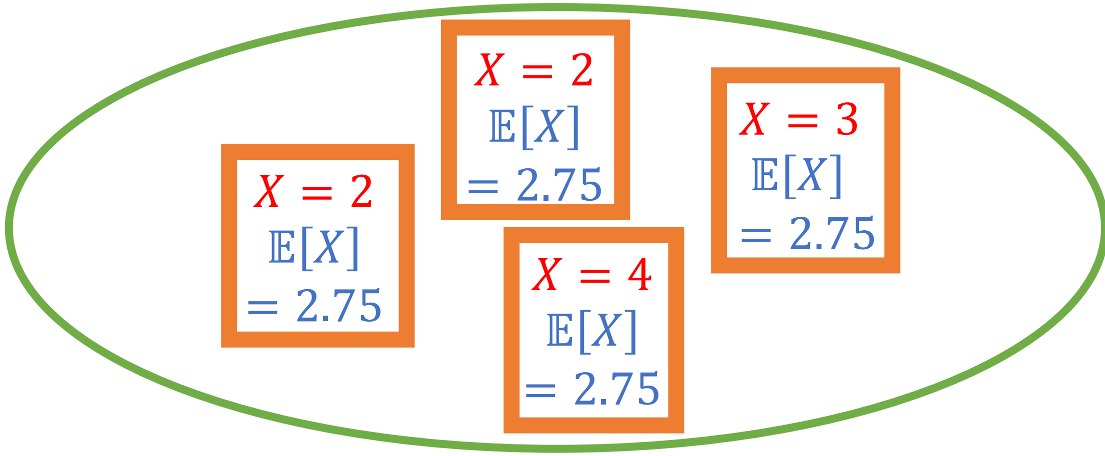
If $X : \Omega \to \mathbb{R}$ is a random variable, its variance is:

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

Example.

Let's slow down and truly understand it!

$\mathbb{E}[X]$ is a number that doesn't depend on probability space of outer X.



Variance

Definition.

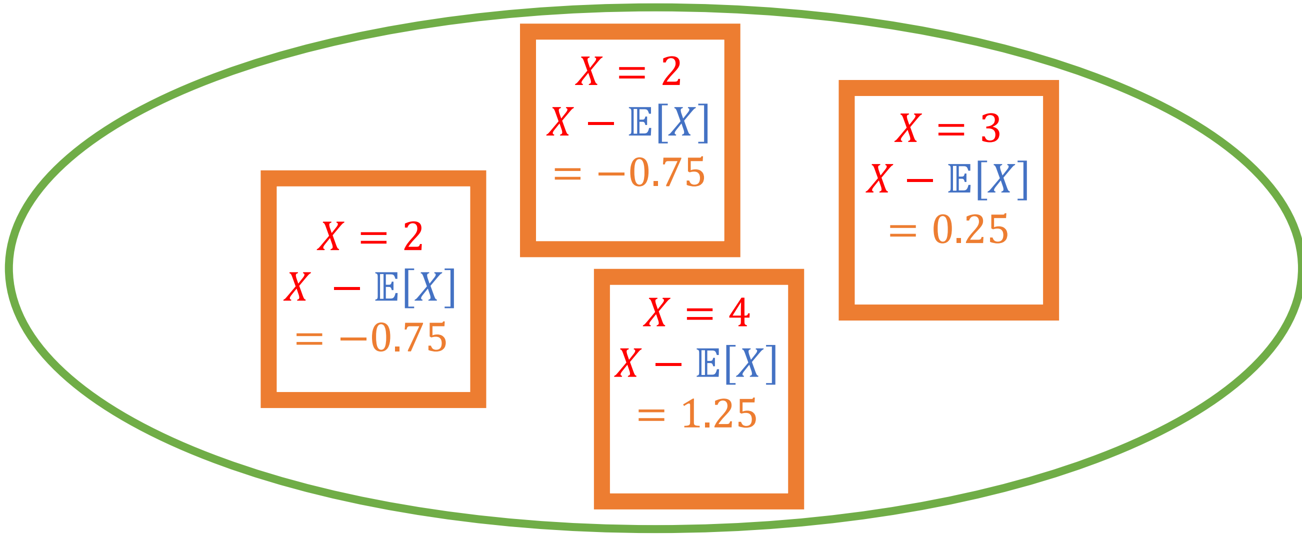
If $X: \Omega \to \mathbb{R}$ is a random variable, its variance is:

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

Example.

Let's slow down and truly understand it!

$X - \mathbb{E}[X]$ is random variable.



Variance

Definition.

If $X : \Omega \to \mathbb{R}$ is a random variable, its variance is:

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

Example.

Let's slow down and truly understand it!

$(X - \mathbb{E}[X])$^2 is also a random variable and $\mathbb{E}_X[(X - \mathbb{E}[X])^2]$ is its expectation.

Alternative Formula

Alternative Formula.

If $X: \Omega \to \mathbb{R}$ is a random variable, its variance is:

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

"Expectation of the square minus the square of expectation"

Proof.

$$\begin{aligned}
\text{Var}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] && \text{(Definition)} \\
&= \mathbb{E}[X^2 - 2 \cdot \mathbb{E}[X] \cdot X + \mathbb{E}[X]^2] && ((a-b)^2 = a^2 - 2ab + b^2) \\
&= \mathbb{E}[X^2] - 2 \cdot \mathbb{E}[X] \cdot \mathbb{E}[X] + \mathbb{E}[X]^2 && \text{(Linearity of expectation)} \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2
\end{aligned}$$

Example: The matching problem, revisited.

Example (The matching problem).

n cats each has a bowl with their name on it. However, when it comes time for dinner, each cat i goes to a random bowl p_i such that no two cats select the same bowl.

Suppose p is uniformly random over all permutations. Let X be the number of cats who got their own bowl.

We have seen $\mathbb{E}[X] = 1$.

What is $\mathrm{Var}[X]$?

Example: The matching problem, revisited.

Sol. Recall the method of indicators,

$$X_i = \mathbf{1}[\text{the } i - \text{th cat got its own bowl}]$$
$$X = X_1 + X_2 + \cdots + X_n$$

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

$$X^2 = (X_1 + X_2 + \cdots + X_n)^2$$
$$= X_1^2 + X_2^2 + \cdots + X_n^2 + 2X_1X_2 + 2X_1X_3 + \cdots + 2X_{n-1}X_n$$

We then need to calculate $\mathbb{E}[X_i^2]$ and $\mathbb{E}[X_iX_j]$ $(i < j)$.

Example: The matching problem, revisited.

Sol.

$$X_i = \mathbf{1}[\text{the } i - \text{th cat got its own bowl}]$$

We then need to calculate $\mathbb{E}\left[X_i^2\right]$ and $\mathbb{E}\left[X_i X_j\right]$ $(i < j)$.

Since X_i is indicator, $X_i^2 = X_i$. So $\mathbb{E}\left[X_i^2\right] = \mathbb{E}[X_i] = \frac{1}{n}$.

$\mathbb{E}\left[X_i X_j\right] = \mathbb{P}[i - \text{th and } j - \text{th cats got their bowls}] = \frac{1}{n} \cdot \frac{1}{n-1}$.

$$\mathbb{E}\,[X^2] = \mathbb{E}\left[X_1^2 + X_2^2 + \cdots + X_n^2 + 2X_1X_2 + 2X_1X_3 + \cdots + 2X_{n-1}X_n\right]$$

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

Multiply by constant

Definition.

$$\mathrm{Var}[cX] = c^2 \cdot \mathrm{Var}[X]$$

Proof.

$$\begin{aligned}\mathrm{Var}[cX] &= \mathbb{E}[(cX)^2] - \mathbb{E}[cX]^2 \quad \text{(Alternative formula)} \\ &= c^2 \cdot \mathbb{E}[X^2] - c^2 \cdot \mathbb{E}[X]^2 \\ &= c^2 \cdot \mathrm{Var}[X]\end{aligned}$$

Standard Deviation

Definition.

$$\sigma[X] = \sqrt{\mathrm{Var}[X]}$$

Intuition

Say with ½ probability $X = -a$

with ½ probability $X = +a$

$\mathbb{E}[X] = 0$

$\mathrm{Var}[X] = a^2$

$\sigma[X] = a$

"≈ deviation from expectation"

Standard Deviation

Definition.

$$\sigma[X] = \sqrt{\mathrm{Var}[X]}$$

Lemma

$$\sigma[c \cdot X] = c \cdot \sigma[X]$$

Variance of the Sum of independent R.V.s

Lemma.

If X, Y are two independent random variable,

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

Proof.

$$\mathrm{Var}[X+Y] = \mathbb{E}[(X+Y)^2] - \mathbb{E}[X+Y]^2$$ (alternative formula)

$$= \mathbb{E}[X^2 + 2XY + Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2$$

$$= \mathbb{E}[X^2] + 2\,\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X]^2 + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y]^2)$$

(linearity of expectation)

$$= \mathbb{E}[X^2] + 2\,\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y^2] - (\mathbb{E}[X]^2 + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y]^2)$$

(independence)

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

Variance of the Sum of independent R.V.s

Why it is true? What does the proof tell us?

For independent RV,

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

In other words,

$$\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y] = 0$$

Variance of the Sum of independent R.V.s

Lemma.

If X, Y are two independent random variable,

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

Comment.

What if they are not independent?

Eg. $X = Y$ always. $\mathrm{Var}[X+Y] = \mathrm{Var}[2X] = 4 \cdot \mathrm{Var}[X] \gg \mathrm{Var}[X] + \mathrm{Var}[Y]$

(positively correlated)

$X = -Y$ always. $\mathrm{Var}[X+Y] = \mathrm{Var}[0] = 0$

(negatively correlated)

Covariance and Correlation

Definition.

If $X, Y: \Omega \to \mathbb{R}$ are two jointly distributed random variable, their covariance

$$\text{Cov}[X, Y] = \mathbb{E}_{X,Y}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Alternative Formula.

$$\text{Cov}[X, Y] = \mathbb{E}_{X,Y}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$$

Proof.

$$\mathbb{E}_{X,Y}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$
$$= \mathbb{E}_{X,Y}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\,\mathbb{E}[Y]$$
$$= \mathbb{E}_{X,Y}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$$

Covariance and Correlation

Lemma.

If X, Y are two random variables,

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\ \mathrm{Cov}[X, Y]$$

Proof.

Assuming X, Y are mean-zero ($\mathbb{E}[X]$=$\mathbb{E}[Y] = 0$).

$$\mathrm{Var}[X + Y] = \mathbb{E}[(X + Y)^2]$$

$$= \mathbb{E}[X^2] + \mathbb{E}[Y^2] - 2 \cdot \mathbb{E}[XY]$$

What if X, Y are not mean-zero ?

X'= X- $\mathbb{E}[X]$. Y'= Y- $\mathbb{E}[Y]$.

Covariance and Correlation

Definition.

If $X, Y : \Omega \to \mathbb{R}$ are two jointly distributed random variable, their correlation

$$\text{Corr}[X, Y] = \frac{\text{Cov}[X,Y]}{\sigma(X)\sigma(Y)} = \frac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}{\sigma(X)\sigma(Y)}$$

Correlation is normalized.

$$\text{Corr}[c \cdot X, Y] = \text{Corr}[X, c \cdot Y] = \text{Corr}[X, Y].$$

$$-1 \leq \text{Corr}[X, Y] = \text{Corr}[X, c \cdot Y] = \text{Corr}[X, Y] \leq 1.$$

(Proof by Cauchy–Schwarz inequality)

Covariance and Correlation

Example.

Say X is the temperature in Berkeley tmr.

Y is the temperature in Palo alto tmr.

As toy model, let's there is a 1/2 chance of raining in bay area tmr.

If it is sunny, the temperature in Berkeley will be 70F.

That in Palo alto will be either 70F or 80F each with prob 1/2.

If it rains, the temperature in Berkeley will be 60F.

That in Palo alto will be 50F or 80F each with prob 1/2.

Covariance and Correlation

Example.

If it is sunny, the temperature in Berkeley will be 70F.

That in Palo alto will be either 70F or 80F each with prob 1/2.

If it rains, the temperature in Berkeley will be 60F.

That in Palo alto will be 50F or 80F each with prob 1/2.

For Berkeley, $\mathbb{E}[X] = 65$. For Palo alto, $\mathbb{E}[Y] = 70.$

$$\mathrm{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \frac{5*0+5*10+(-5)*(-20)+(-5)*10}{4} = 25$$

Covariance and Correlation

Example.

If it is sunny, the temperature in Berkeley will be 70F.

That in Palo alto will be either 70F or 80F each with prob 1/2.

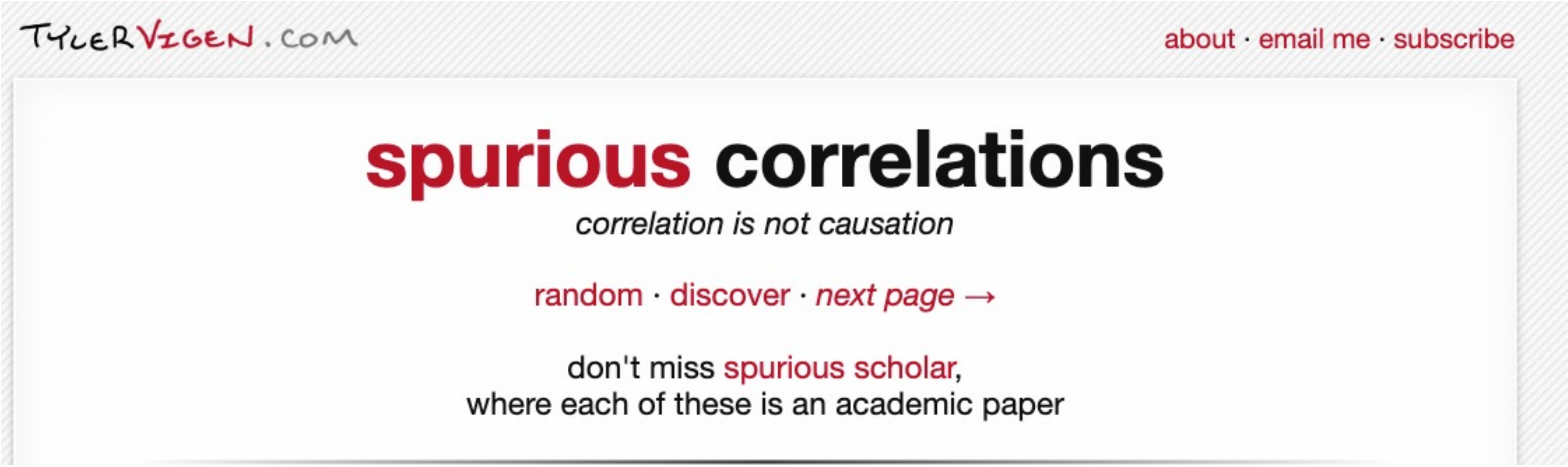
If it rains, the temperature in Berkeley will be 60F.

That in Palo alto will be 50F or 80F each with prob 1/2.

$$\mathrm{Cov}[X, Y] = 25,\ \sigma[X] = \sqrt{\mathrm{Var}[X]} = 5,\ \sigma[Y] = \sqrt{\mathrm{Var}[Y]} = 12.24 \ldots.$$

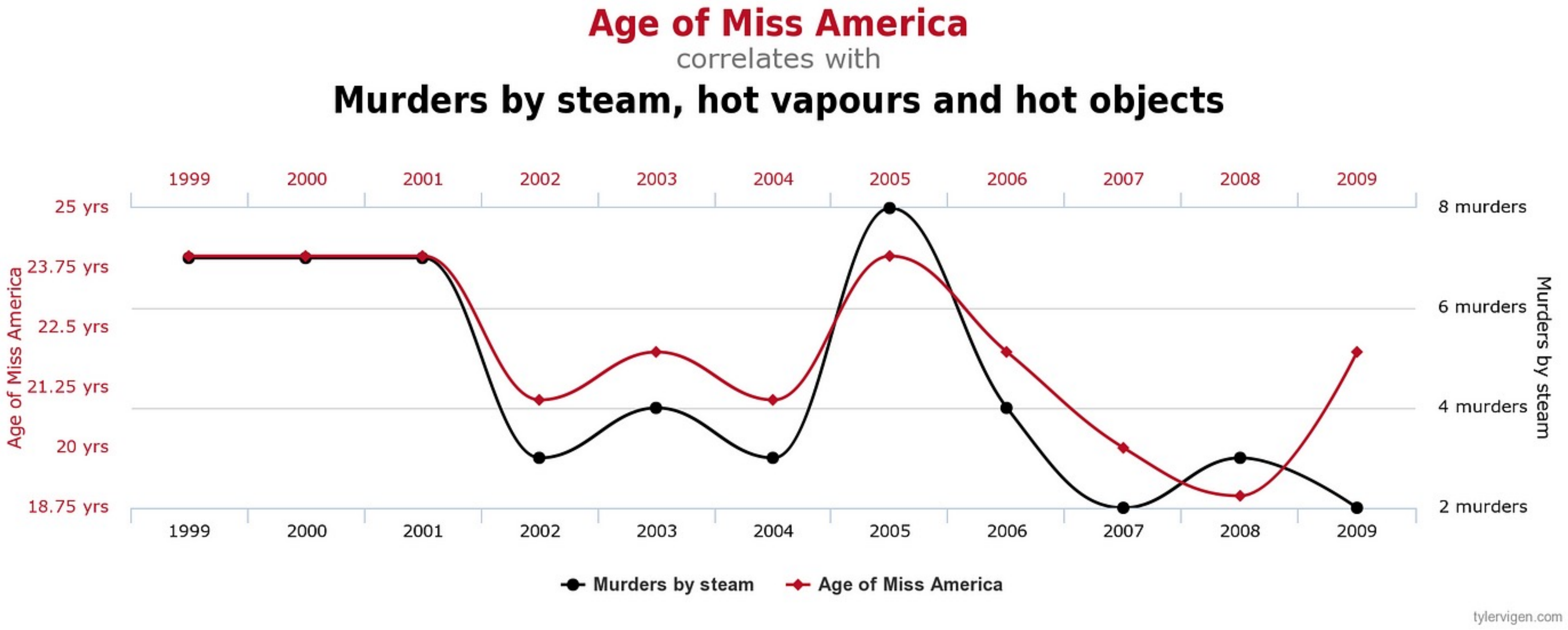
$$\mathrm{Corr}[X, Y] = \frac{25}{5 * 12.24} = 0.408 \ldots..$$

Correlation does not imply Causality

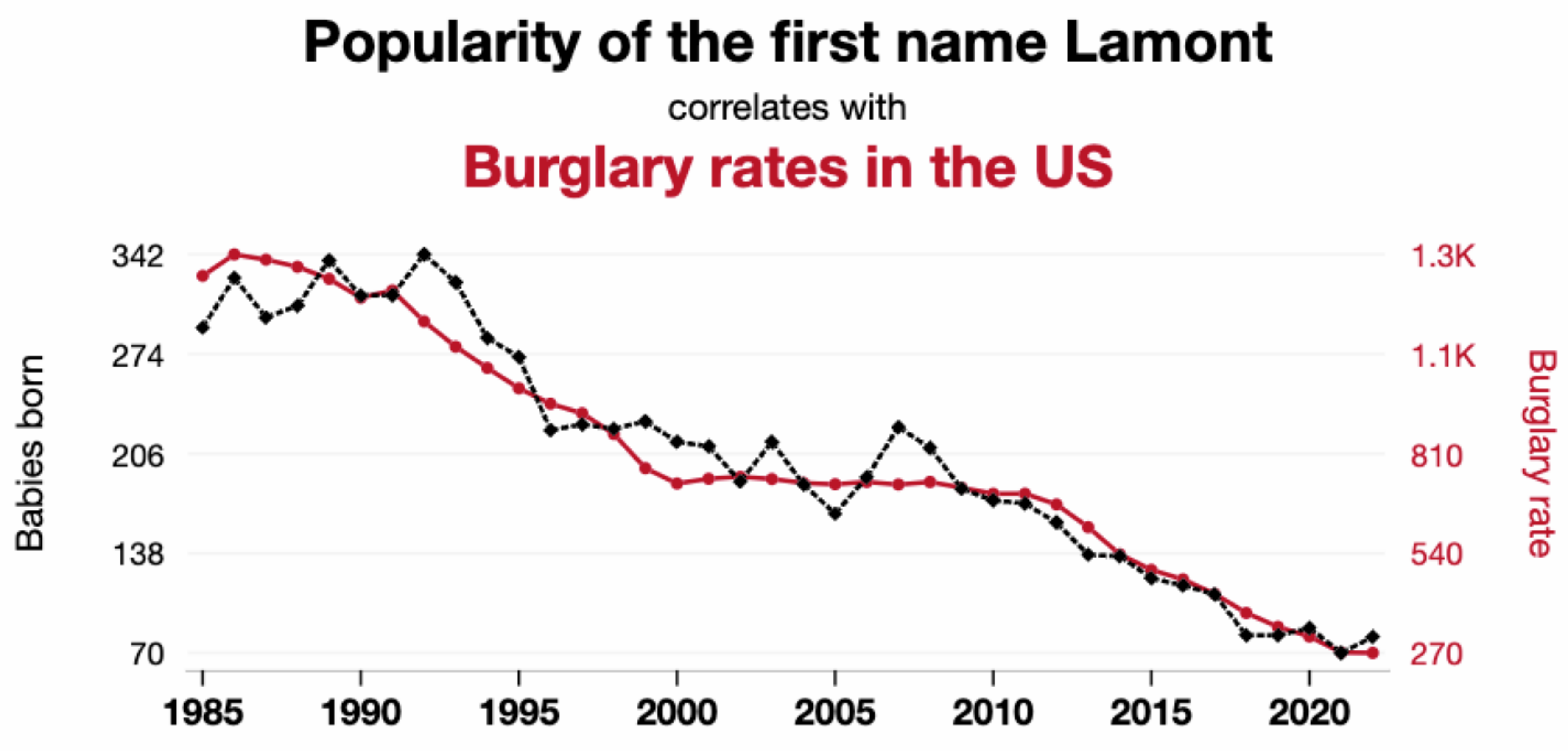


https://www.tylervigen.com/spurious-correlations

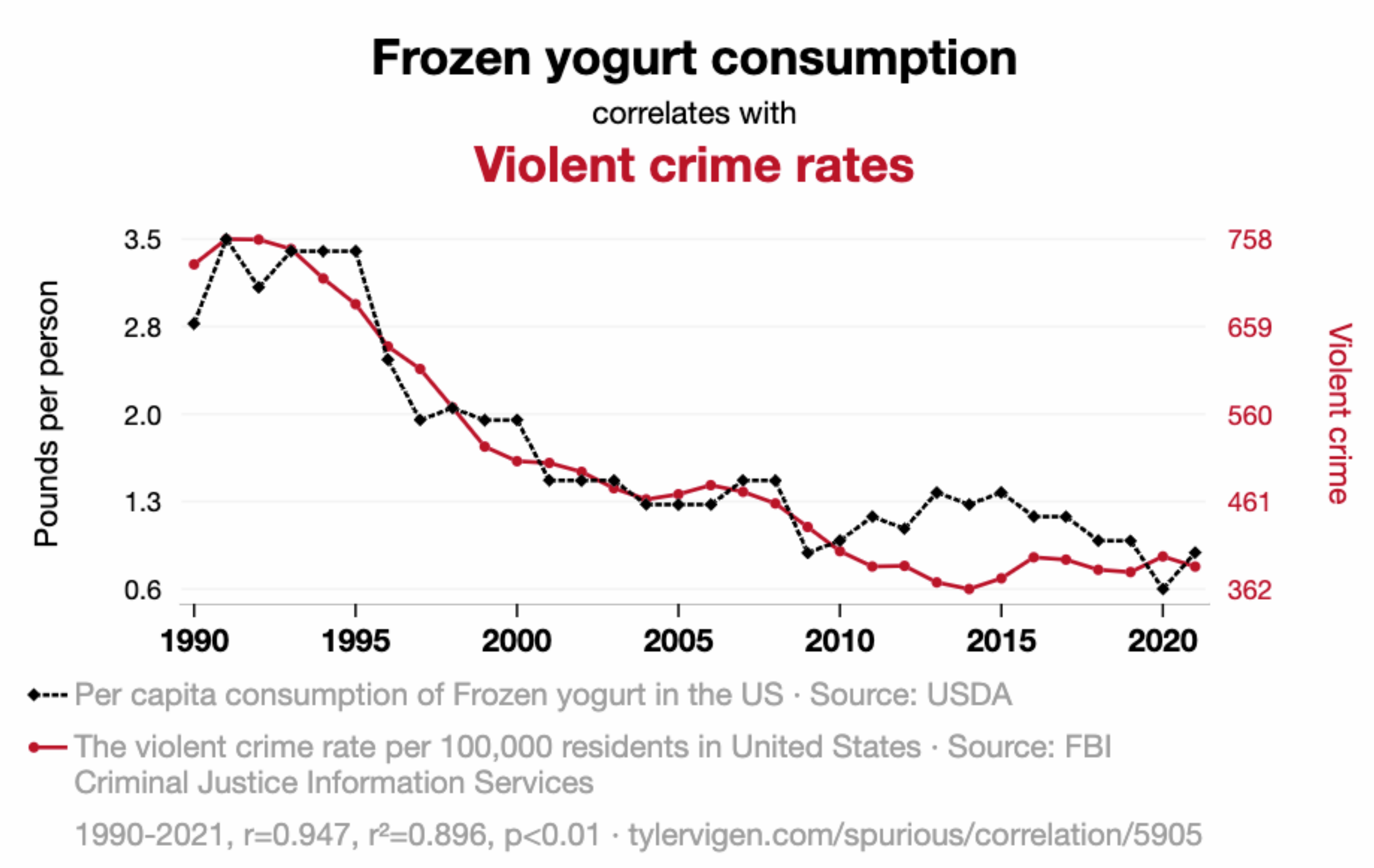
Correlation does not imply Causality



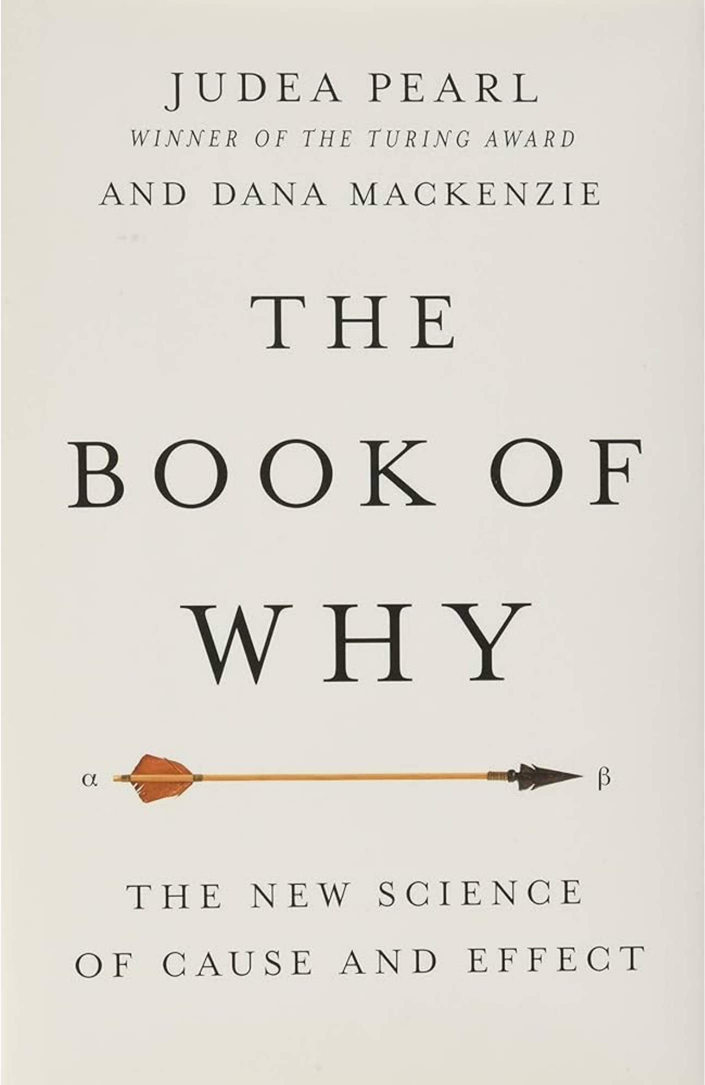
Correlation does not imply Causality

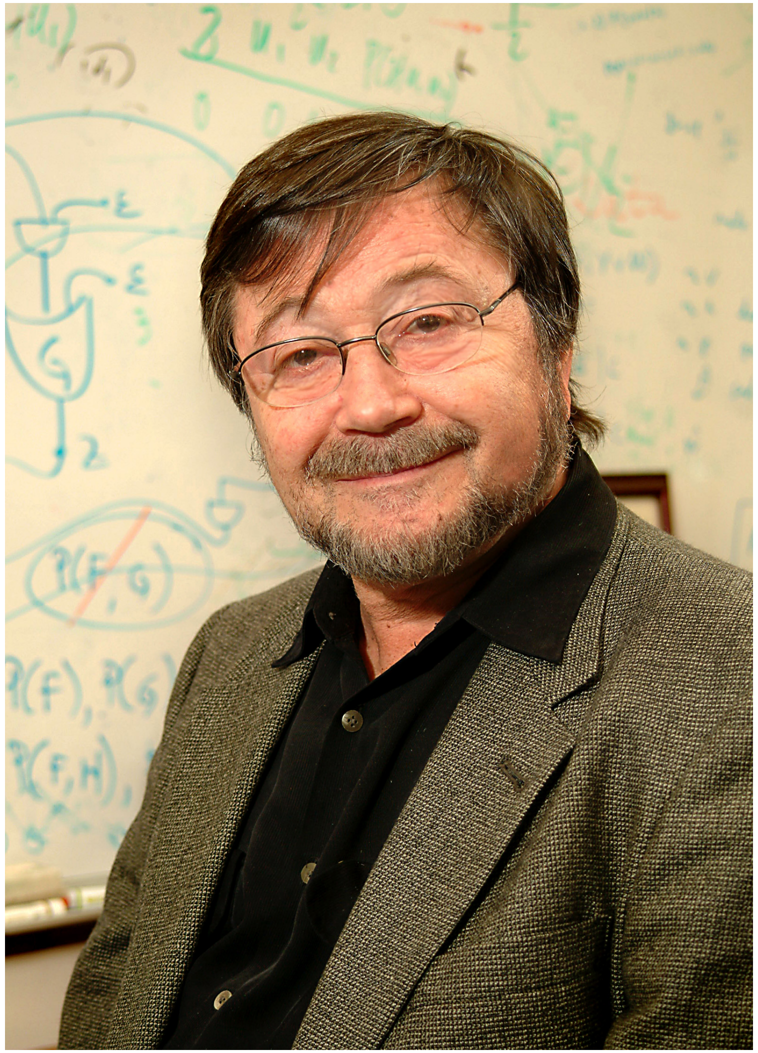


Correlation does not imply Causality



The Book of Why by Judea Pearl





Example: Variance of Bernoulli distribution

Bernoulli distribution

With probability p, we have $X = 1$.
With probability $1 - p$, we have $X = 0$.

Variance

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

$$\mathbb{E}[X] = p$$
$$\mathbb{E}[X^2] = p$$

$$\text{Var}[X] = p - p^2 = p(1 - p)$$

Example: Variance of Geometric distribution

Geometric distribution

With probability p, we have $X = 1$.

With probability $p \cdot (1-p)$, we have $X = 2$.

With probability $p \cdot (1-p)^2$, we have $X = 3$.

....

Variance

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

$$\mathbb{E}[X^2] = p \cdot 1 + (1-p)\mathbb{E}[X^2 \mid X \geq 2]$$

Example: Variance of Geometric distribution

Geometric distribution

With probability p, we have $X = 1$.

With probability $p \cdot (1 - p)$, we have $X = 2$.

With probability $p \cdot (1 - p)^2$, we have $X = 3$.

....

Distribution of $X \mid X \geq 2$ = Distribution of $X + 1$!

With probability p, we have $X = 2$.

With probability $p \cdot (1 - p)$, we have $X = 3$.

With probability $p \cdot (1 - p)^2$, we have $X = 4$.

....

Example: Variance of Geometric distribution

Geometric distribution

With probability p, we have $X = 1$.

With probability $p \cdot (1-p)$, we have $X = 2$.

With probability $p \cdot (1-p)^2$, we have $X = 3$.

....

Variance

We know $\mathbb{E}[X] = \frac{1}{p}$. $\mathbb{E}[X^2] = p + (1-p)\mathbb{E}[(1+X)^2]$ (self-ref trick)

Example: Variance of Geometric distribution

Geometric distribution

With probability p, we have $X = 1$.

With probability $p \cdot (1-p)$, we have $X = 2$.

With probability $p \cdot (1-p)^2$, we have $X = 3$.

....

Variance

We know $\mathbb{E}[X] = \frac{1}{p}$. $\mathbb{E}[X^2] = p + (1-p)\mathbb{E}[(1+X)^2]$ (self-ref trick)

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

Solve the equation => $\mathbb{E}[X^2] = \frac{1+2(1-p)/p}{p} = \frac{2-p}{p^2}$

Example: Variance of Geometric distribution

Geometric distribution

With probability p, we have $X = 1$.

With probability $p \cdot (1-p)$, we have $X = 2$.

With probability $p \cdot (1-p)^2$, we have $X = 3$.

…………

Variance

We know $\mathbb{E}[X] = \frac{1}{p}$. $\mathbb{E}[X^2] = \frac{1+2(1-p)/p}{p} = \frac{2-p}{p^2}$

$$\mathrm{Var}[X] = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}.$$