Josh Hug · You
Teaching Professor at UC Berkeley

(edited) 3y

(Neil deGrasse Tyson voice): Well... actually... if 1000 people randomly select a balloon, the chance that at least one of them finds their own balloon on the first try is 63%. In fact, for a group of more than 6 people, the chance remains around the same that someone will find their own balloon on the first try, approximately 63%, no matter how large the group. This is because the number of different permutations where nobody gets their own balloon (called a derangement by mathematicians) is only a bit more than a third of the number of possible permutations.

(https://blogs.sas.com/content/iml/2013/12/04/probability-of-pulling-your-own-name-from-a-hat.html)

Like · 14 | Reply | 410 impressions

164,534

Blake Green, MBA · 3rd+
Operations and Business Development Professional
3yr ·

+ Follow

A professor gave a balloon to every student, who had to inflate it, write their name on it and throw it in the hallway. The professor then mixed all the balloons. The students were then given 5 minutes to find their own balloon. Despite a hectic search, no one found their balloon.

At that point, the professor told the students to take the first balloon that they found and hand it to the person whose name was written on it. Within 5 minutes, everyone had their own balloon.

The professor said to the students: "These balloons are like happiness. We will never find it if everyone is looking for their own. But if we care about other people's happiness, we'll find ours too."



4,186 comments · 18,351 reposts

Example Variance Calculations

Lecture 22, CS70 Fall 2026

Coupon Collector and Expectation

Variance

- Law of the Unconscious Statistician
- Derivation and Definition
- Properties of Variance
- **Example Variance Calculations**
- Sums of Random Variables

CC BY NC SA

Computing the Variance of a Random Variable

When computing the variance of a random variable X, we need two quantities:

- The expectation $\mathbb{E}[X]$.
- The expectation of X^2, i.e. $\mathbb{E}[X^2]$

Often, but not always, $\mathbb{E}[X^2]$ requires some cleverness to compute.

CC BY NC SA

Example 1: Bernoulli RV

Suppose $X \sim \text{Bernoulli}(p)$. Recall that $\mathbb{E}[X] = p$.

CC BY NC SA

Example 1: Bernoulli RV

Suppose $X \sim \mathbf{Bernoulli}(p)$. Recall that $\mathbb{E}[X] = p$.

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

- Since X is always 0 or 1, we have that $X^2 = X$, therefore $\mathbb{E}[X^2] = \mathbb{E}[X] = p$.
- The variance is therefore $\mathbf{Var}(X) = p - \mathbb{E}[X]^2 = p - p^2$

CC BY NC SA

Example 1: Bernoulli RV

Suppose $X \sim \mathbf{Bernoulli}(p)$. Recall that $\mathbb{E}[X] = p$.

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

- Since X is always 0 or 1, we have that $X^2 = X$, therefore $\mathbb{E}[X^2] = \mathbb{E}[X] = p$.
- The variance is therefore $\mathbf{Var}(X) = p - \mathbb{E}[X]^2 = p - p^2$

Often this is written in the form $p(1-p)$

Observation, if we pick $p = 1$ (always heads), then the variance is 0.

- This makes sense! If the RV isn't actually random, there should be no variance.
- Similarly, if p is close to 0 or 1, the variance is low, usually get same flip.

CC BY NC SA

Example 2: Rolling a Fair Six Sided Die

Recall that if X is the result of rolling a fair six-sided die, we have that $\mathbb{E}[X] = 7/2$.

To compute variance, we just need $\mathbb{E}[X^2]$. This is straightforward:

CC BY NC SA

Example 2: Rolling a Fair Six Sided Die

Recall that if X is the result of rolling a fair six-sided die, we have that $\mathbb{E}[X] = 7/2$.

To compute variance, we just need $\mathbb{E}[X^2]$. This is straightforward:

$$\begin{aligned}\mathbb{E}[X^2] &= \frac{1}{6}\times 1^2 + \frac{1}{6}\times 2^2 + \frac{1}{6}\times 3^2 + \frac{1}{6}\times 4^2 + \frac{1}{6}\times 5^2 + \frac{1}{6}\times 6^2 \\ &= \frac{1}{6}\times (1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) \\ &= \frac{91}{6}\end{aligned}$$

CC BY NC SA

Example 2: Rolling a Fair Six Sided Die

Recall that if X is the result of rolling a fair six-sided die, we have that $\mathbb{E}[X] = 7/2$. We also just computed $\mathbb{E}[X^2] = 91/6$

The variance of a six sided die roll is given by $\mathbf{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

$$= \frac{91}{6} - \left(\frac{7}{2}\right)^2$$

$$= \frac{35}{12} = 2.92$$

2.92 will come up again later today.

CC BY NC SA

Example 3: Fixed Points of Permutation

Let X_n be the number of students who receive their own homework if we take homeworks from every student and hand them back randomly.

- Last time we showed that $\mathbb{E}[X_n] = 1$.
- How dispersed is this? Do we expect to sometimes get $X_n = 10$, $X_n = 20$?

As before let $X_n = I_1 + I_2 + \cdots + I_n$, where I_i means student i got their homework back.

To compute variance, we already have $\mathbb{E}[X_n]$. We just need $\mathbb{E}[X_n^2]$.

cc BY NC SA

Example 3: Fixed Points of Permutation

$X_n = I_1 + I_2 + \cdots + I_n$, where I_i means student i got their homework back.

We need $\mathbb{E}[X_n^2]$. To get this, we compute $\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2]$. Using the same formula from before, we have that:

$$\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2] = \sum_{i=1}^{n} \mathbb{E}[I_i^2] + 2\sum_{i<j} \mathbb{E}[I_i I_j]$$

What is $\mathbb{E}[I_i^2]$?

- joshh.ug/F



CC BY NC SA

Example 3: Fixed Points of Permutation

$X_n = I_1 + I_2 + \cdots + I_n$, where I_i means student i got their homework back.

We need $\mathbb{E}[X_n^2]$. To get this, we compute $\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2]$. Using the same formula from before, we have that:

$$\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2] = \sum_{i=1}^{n} \mathbb{E}[I_i^2] + 2\sum_{i<j} \mathbb{E}[I_i I_j]$$

Since I_i is just a Bernoulli indicator variable $I_i^2 = I_i$. Since the chance a student gets back their homework is $1/n$, $I_i \sim \text{Bernoulli}(1/n)$. Together these facts give us:

$$\mathbb{E}[I_i^2] = \mathbb{E}[I_i] = 1/n$$

CC BY NC SA

Example 3: Fixed Points of Permutation

$X_n = I_1 + I_2 + \cdots + I_n$, where I_i means student i got their homework back.

We need $\mathbb{E}[X_n^2]$. To get this, we compute $\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2]$. Using the same formula from before, we have that:

$$\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2] = \sum_{i=1}^{n} \frac{1}{n} + 2\sum_{i<j} \mathbb{E}[I_i I_j]$$

Since $\sum_{i=1}^{n} \frac{1}{n} = 1$, we can simplify by replacing $\sum_{i=1}^{n} \frac{1}{n}$ by 1.

CC BY NC SA

Example 3: Fixed Points of Permutation

$X_n = I_1 + I_2 + \cdots + I_n$, where I_i means student i got their homework back.

We need $\mathbb{E}[X_n^2]$. To get this, we compute $\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2]$. Using the same formula from before, we have that:

$$\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2] = 1 + 2\sum_{i<j} \mathbb{E}[I_i I_j]$$

Since $\sum_{i=1}^{n} \frac{1}{n} = 1$, we can simplify by replacing $\sum_{i=1}^{n} \frac{1}{n}$ by 1.

- Now we just need to know $2\sum_{i<j} \mathbb{E}[I_i I_j]$

CC BY NC SA

Example 3: Fixed Points of Permutation

$X_n = I_1 + I_2 + \cdots + I_n$, where I_i means student i got their homework back.

We need $\mathbb{E}[X_n^2]$. To get this, we compute $\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2]$. Using the same formula from before, we have that:

$$\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2] = 1 + 2\sum_{i<j} \mathbb{E}[I_i I_j]$$

What is $\mathbb{E}[I_i I_j]$?

- joshh.ug/G

We'll take a 4 minute break. Feel free to answer.



CC BY NC SA

Example 3: Fixed Points of Permutation

$X_n = I_1 + I_2 + \cdots + I_n$, where I_i means student i got their homework back.

We need $\mathbb{E}[X_n^2]$. To get this, we compute $\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2]$. Using the same formula from before, we have that:

$$\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2] = 1 + 2\sum_{i<j} \mathbb{E}[I_i I_j]$$

What is $\mathbb{E}[I_i I_j]$?

CC BY NC SA

Example 3: Fixed Points of Permutation

$X_n = I_1 + I_2 + \cdots + I_n$, where I_i means student i got their homework back.

We need $\mathbb{E}[X_n^2]$. To get this, we compute $\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2]$. Using the same formula from before, we have that:

$$\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2] = 1 + 2\sum_{i<j} \mathbb{E}[I_i I_j]$$

$$\mathbb{E}[I_i I_j] = 0\times 0\times P(I_i = 0,\ I_j = 0) + 0\times 1\times P(I_i = 0,\ I_j = 1)$$
$$+1\times 0\times P(I_i = 1,\ I_j = 0) + 1\times 1\times P(I_i = 1,\ I_j = 1)$$

But only one of these terms is not trivially zero: $1\times 1\times P(I_i = 1,\ I_j = 1)$

CC BY NC SA

Example 3: Fixed Points of Permutation

$X_n = I_1 + I_2 + \cdots + I_n$, where I_i means student i got their homework back.

We need $\mathbb{E}[X_n^2]$. To get this, we compute $\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2]$. Using the same formula from before, we have that:

$$\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2] = 1 + 2\sum_{i<j} \mathbb{E}[I_i I_j]$$

Since both I_i and I_j are indicators, only one term is not trivially zero:

$$\begin{aligned}
\mathbb{E}[I_i I_j] &= P(I_i I_j = 1) \\
&= P(i \text{ and } j \text{ get their homework back}) \\
&= \frac{1}{n(n-1)}
\end{aligned}$$

CC BY NC SA

Example 3: Fixed Points of Permutation

$X_n = I_1 + I_2 + \cdots + I_n$, where I_i means student i got their homework back.

We need $\mathbb{E}[X_n^2]$. To get this, we compute $\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2]$. Using the same formula from before, we have that:

$$\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2] = 1 + 2\sum_{i<j} \frac{1}{n(n-1)}$$

Since both I_i and I_j are indicators, only one term is not trivially zero:

$$\begin{aligned}
\mathbb{E}[I_i I_j] &= P(I_i I_j = 1) \\
&= P(i \text{ and } j \text{ get their homework back}) \\
&= \frac{1}{n(n-1)}
\end{aligned}$$

Example 3: Fixed Points of Permutation

$X_n = I_1 + I_2 + \cdots + I_n$, where I_i means student i got their homework back.

We need $\mathbb{E}[X_n^2]$. To get this, we compute $\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2]$. Using the same formula from before, we have that:

$$\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2] = 1 + 2\sum_{i<j} \frac{1}{n(n-1)}$$

What is $2\sum_{i<j} \frac{1}{n(n-1)}$?

CC BY NC SA

Example 3: Fixed Points of Permutation

$X_n = I_1 + I_2 + \cdots + I_n$, where I_i means student i got their homework back.

We need $\mathbb{E}[X_n^2]$. To get this, we compute $\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2]$. Using the same formula from before, we have that:

$$\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2] = 1 + 2\sum_{i<j} \frac{1}{n(n-1)}$$

What is $2\sum_{i<j} \frac{1}{n(n-1)}$? There are $\binom{n}{2}$ ways that i is less than j, so

$$2\sum_{i<j} \frac{1}{n(n-1)} = 2\binom{n}{2} \times \frac{1}{n(n-1)}$$

$$= 2\frac{n(n-1)}{2} \times \frac{1}{n(n-1)} \quad = 1$$

CC BY NC SA

Example 3: Fixed Points of Permutation

$X_n = I_1 + I_2 + \cdots + I_n$, where I_i means student i got their homework back.

We need $\mathbb{E}[X_n^2]$. To get this, we compute $\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2]$. Using the same formula from before, we have that:

$$\mathbb{E}[X_n^2] = \mathbb{E}[(I_1 + I_2 + \cdots + I_n)^2] = 1 + 1 = 2$$

After all that work, we've learned that $\mathbb{E}[X_n^2] = 2$

CC BY NC SA

Example 3: Fixed Points of Permutation

Let X_n be the number of students who receive their own homework if we take homeworks from every student and hand them back randomly.

- Last time we showed that $\mathbb{E}[X_n] = 1$.
- How dispersed is this? Do we expect to sometimes get $X_n = 10$, $X_n = 20$?

To compute variance, we know $\mathbb{E}[X_n] = 1$. We also know $\mathbb{E}[X_n^2] = 2$.

What is $\mathbf{Var}(X_n)$?

- joshh.ug/H

Example 3: Fixed Points of Permutation

Let X_n be the number of students who receive their own homework if we take homeworks from every student and hand them back randomly.

- Last time we showed that $\mathbb{E}[X_n] = 1$.
- How dispersed is this? Do we expect to sometimes get $X_n = 10$, $X_n = 20$?

To compute variance, we know $\mathbb{E}[X_n] = 1$. We just need $\mathbb{E}[X_n^2] = 2$.

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

$$= 2 - 1$$

$$= 1$$

Note: If you weren't sure it was 1 or -1, keep in mind variance must be non-negative.

CC BY NC SA

Example 3: Fixed Points of Permutation

Let X_n be the number of students who receive their own homework if we take homeworks from every student and hand them back randomly.

- Last time we showed that $\mathbb{E}[X_n] = 1$.
- And we also know that $\mathbf{Var}(X_n) = 1$.

How dispersed is this? Do we expect to sometimes get $X_n = 10$, $X_n = 20$?

- We don't expect to get a value near 10, no matter how big n gets.
- On average, the number of students who get their homework back is 1.
- And the variance (and standard deviation) is 1.
- If we were to run a bunch of experiments, the usual deviation between our observation and the mean (1) would be around 1.

cc BY NC SA

Example 3: Fixed Points of Permutation

Let's simulate handing the homework back at random:

Students, n **30**

Simulate | **+10 trials** | **+100 trials** | **+1000 trials** | **Reset**

1 → 17	2 → 23	3 → 27	4 → 1	5 → 28	6 → 8
7 → 24	8 → 25	9 → 21	**10 → 10**	11 → 12	12 → 16
13 → 3	14 → 4	15 → 9	16 → 19	17 → 26	18 → 14
19 → 7	20 → 15	21 → 30	22 → 2	23 → 22	24 → 20
25 → 11	26 → 29	27 → 18	28 → 5	29 → 6	30 → 13

Small number: the student. Large number: the homework they got back. their own homework

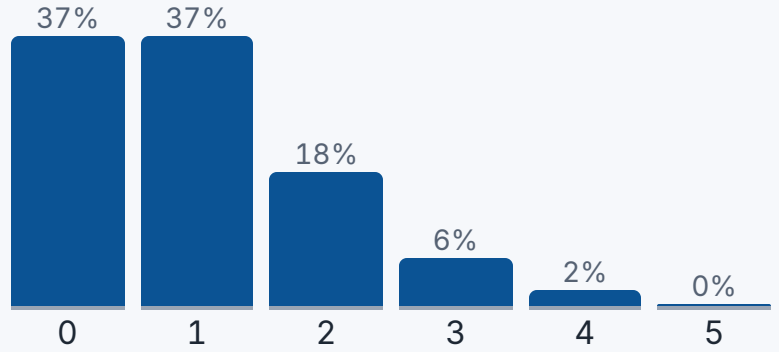
Got their own (this trial)
1
of 30 students

Trials
1000

Average
0.99
theory: 1

Variance
1.00
theory: 1

How many got their own — share of trials



CC BY NC SA

Sums of Independent Random Variables

Lecture 22, CS70 Fall 2026

Coupon Collector and Expectation

Variance

- Law of the Unconscious Statistician
- Derivation and Definition
- Properties of Variance
- Example Variance Calculations
- **Sums of Random Variables**

CC BY NC SA

Sums of Independent Random Variables

In a previous lecture we already saw that if we have independent random variables $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$, and $\text{S} = X + Y$, then $S \sim \text{Poisson}(\lambda + \mu)$.

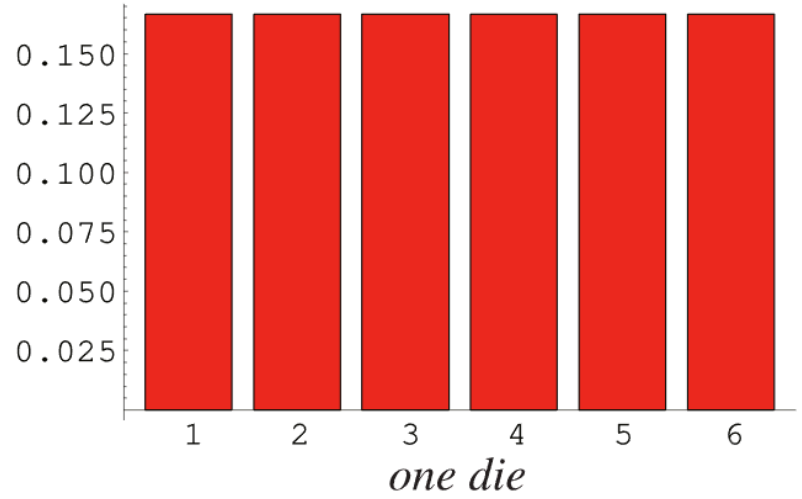
In general, summing random variables is a complicated business.

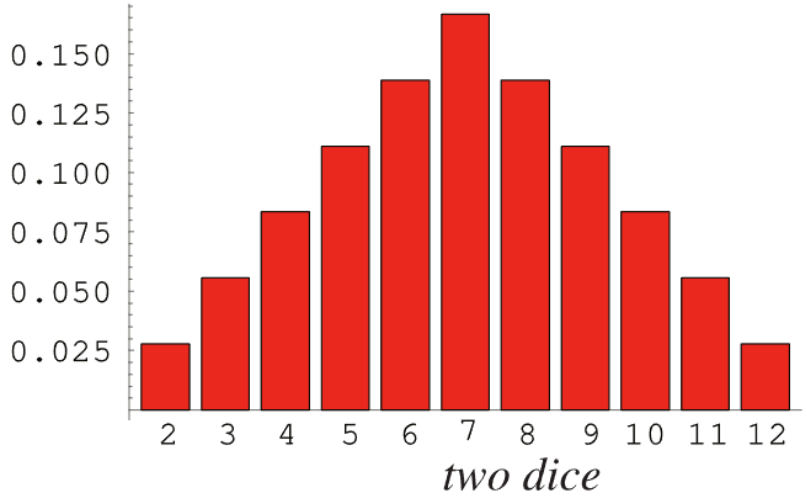
- It's a notable property that two Poisson RVs sum to be Poisson as well.
- Not (always) true for the sum of Bernoulli, Binomial, Geometric.

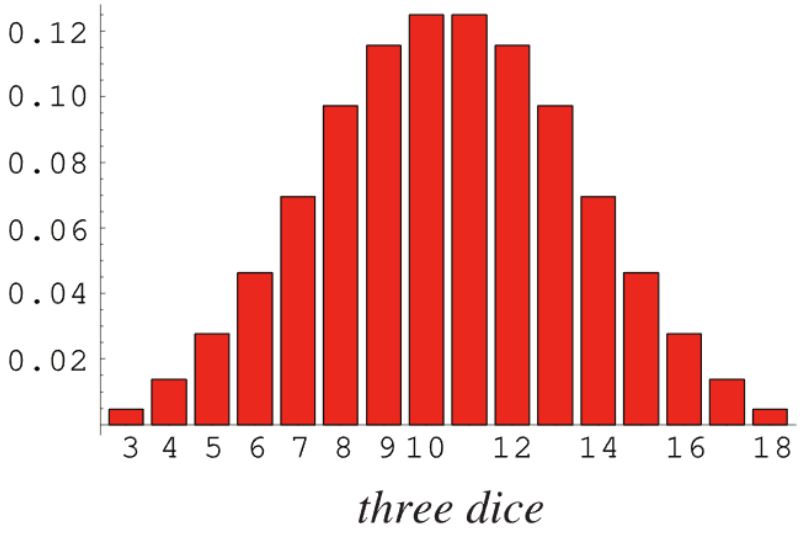
As an example, consider summing n rolls of fair six-sided dice (next slide).

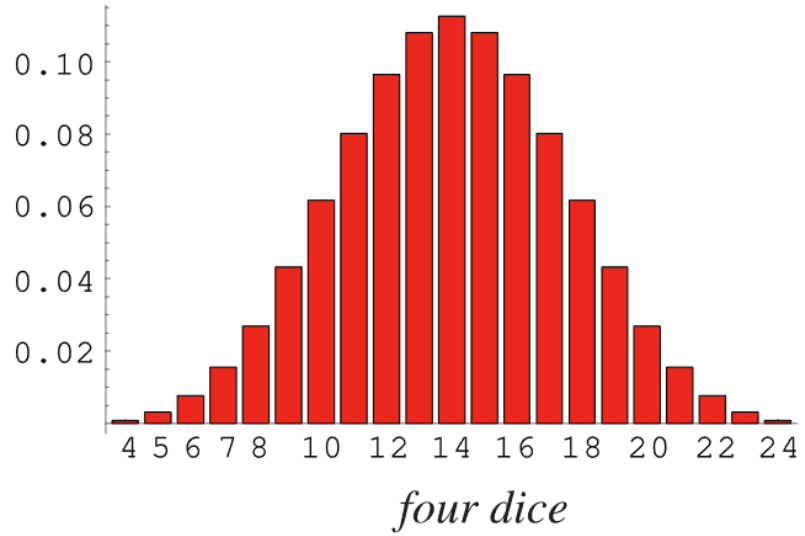
CC BY NC SA

From https://mathworld.wolfram.com/Dice.html



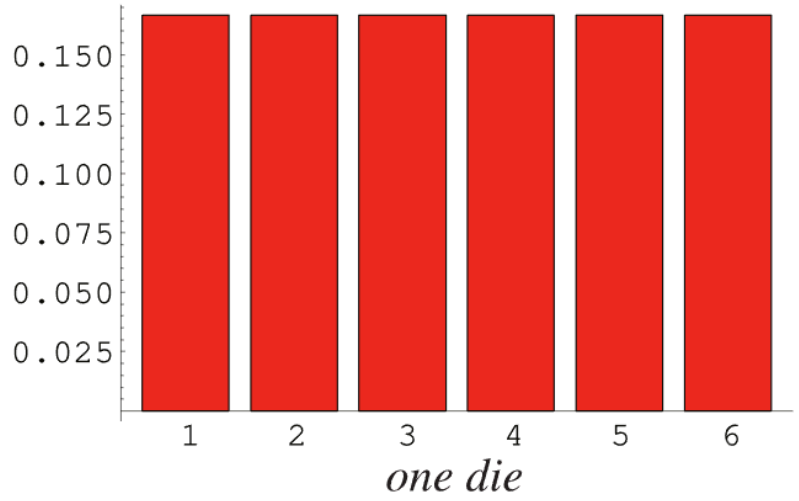


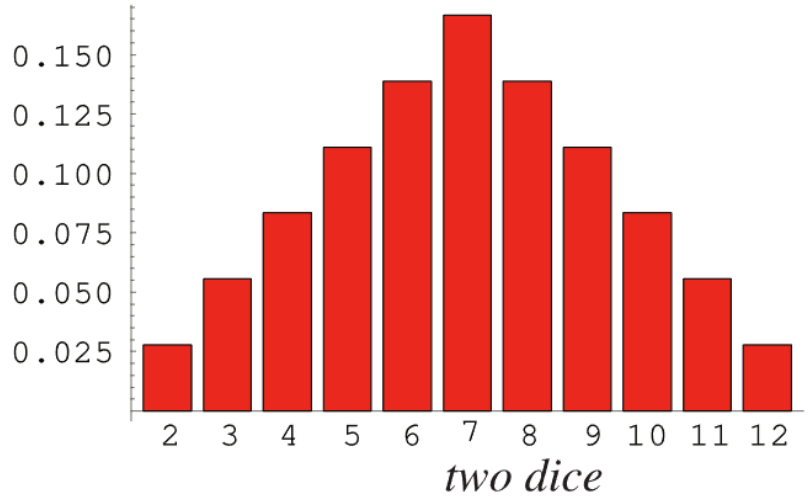


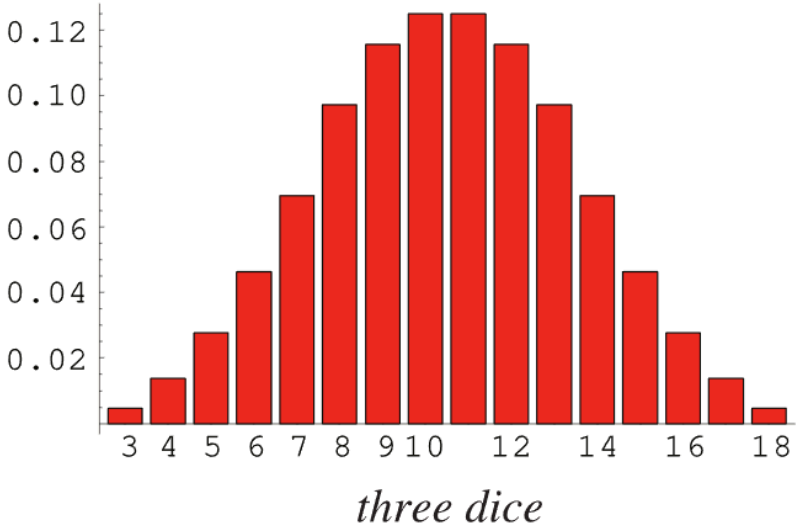


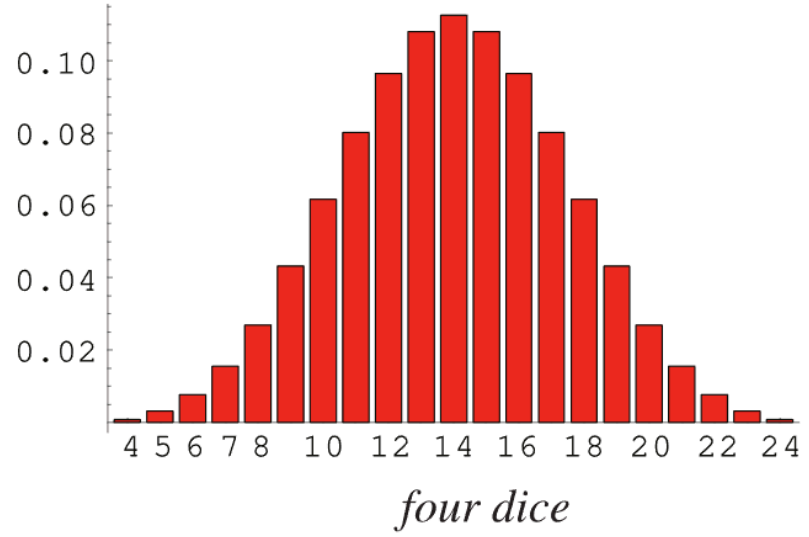
Note: The operation that is happening here is called “convolution”. We won’t discuss today.

Expected Value and Standard Deviation vs. N



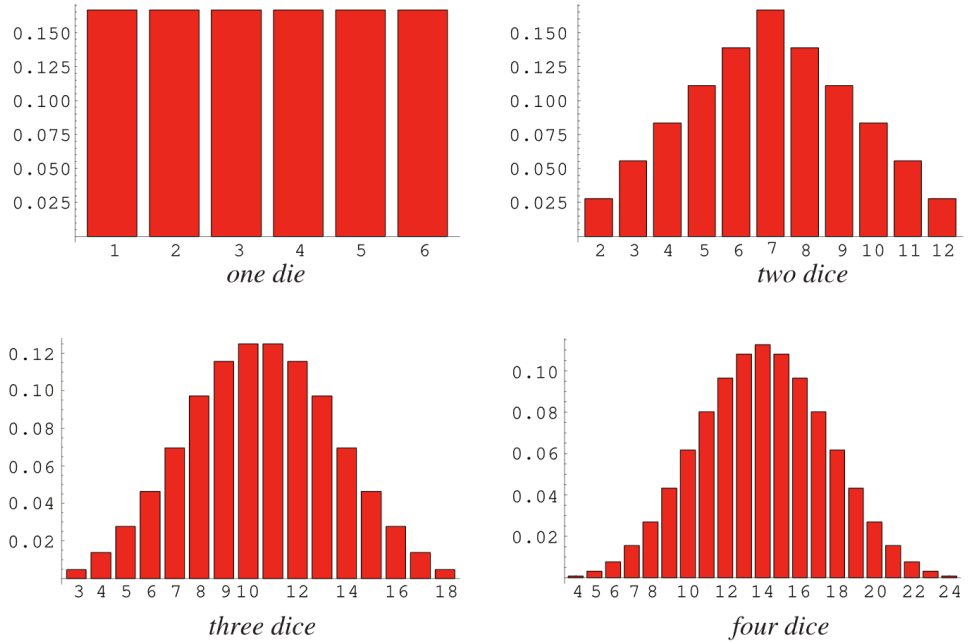






n	$E[X]$	$\sigma(X)$
1	3.5	1.71
2	7	2.42
3	10.5	2.96
4	14	3.42
5	17.5	3.82

Expected Value and Standard Deviation vs. N



n	$E[X]$	$\sigma(X)$
1	3.5	1.71
2	7	2.42
3	10.5	2.96
4	14	3.42
5	17.5	3.82

Two interesting observations as n, the number of random variables summed together, increases:

- The mean increases (linearly).
- The distribution concentrates around the mean, i.e. the standard deviation grows relatively slowly.
- In fact, $\sigma(n) \approx 1.7\sqrt{n}$, or equivalently $\mathrm{Var}(n) \approx 2.9n$

CC BY NC SA

Expectation and Sums of Independent Random Variables

While we can't concisely give the distribution of the sum of two independent random variables, we can reason about the expectation and variance of the sum.

We already know that $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

- Example: Flipping 100 fair coins yields an average of 50 heads.

But what about their variance?

- Example: We saw empirically with dice that rolling 4 dice gives 4 times the variance (or 2 times the standard deviation).
- Let's investigate further.

CC BY NC SA

Expectation and Sums of Independent Random Variables

Claim: If X and Y are independent, then $\mathbf{Var}(X+Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)$.

- We need to consider $\mathbf{Var}(X+Y) = \mathbb{E}[(X+Y)^2] - (\mathbb{E}[X+Y])^2$

CC BY NC SA

Expectation and Sums of Independent Random Variables

Claim: If X and Y are independent, then $\mathbf{Var}(X+Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)$.

- We need to consider $\mathbf{Var}(X+Y) = \mathbb{E}[(X+Y)^2] - (\mathbb{E}[X+Y])^2$

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

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

With these two quantities expanded, we can write:

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

To deal with this, let's take a detour.

CC BY NC SA

Expectation and Sums of Independent Random Variables

Lemma: If X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Proof:

CC BY NC SA

Expectation and Sums of Independent Random Variables

Lemma: If X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Proof:

$$\mathbb{E}[XY] = \sum_a \sum_b ab \times P(X = a, Y = b)$$

$$= \sum_a \sum_b ab \times P(X = a) \times P(Y = b)$$

$$= \left(\sum_a a \times P(X = a) \right) \times \left(\sum_b b \times P(Y = b) \right)$$

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

Example term for rolling a four sided and six sided die:

$$2 \times 5 \times \mathrm{P}(X = 2, Y = 5)$$

$$= 2 \times 5 \times \mathrm{P}(X = 2) \times \mathrm{P}(Y = 5)$$

$$= 2 \times 5 \times \frac{1}{4} \times \frac{1}{6}$$

$$= \left(2 \times \frac{1}{4}\right) \times \left(5 \times \frac{1}{6}\right)$$

CC BY NC SA

Expectation and Sums of Independent Random Variables

Claim: If X and Y are independent, then $\mathbf{Var}(X+Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)$.

- We need to consider $\mathbf{Var}(X+Y) = \mathbb{E}[(X+Y)^2] - (\mathbb{E}[X+Y])^2$

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

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

$$= \mathbf{Var}(X) + \underbrace{2\mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y]}_{\text{We just showed that if } X \text{ and } Y \text{ are independent, then } \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]} + \mathbf{Var}(Y)$$

$$= \mathbf{Var}(X) + \mathbf{Var}(Y)$$

CC BY NC SA

Summing i.i.d. Random Variables

If we have n random variables X_i that are independent and identically distributed, and $S = X_1 + \cdots + X_n$, then we have that:

- $\mathbb{E}[S] = n\mathbb{E}[X_1]$
- $\mathbf{Var}(S) = n\mathbf{Var}(X_1)$

Example: For rolling n six-sided dice, we have that:

- $\mathbb{E}[S] = 3.5n$
- $\mathbf{Var}(S) = 2.92n$
- $\sigma(S) = 1.7\sqrt{n}$

In the notes, they use the example of flipping n coins (rather than dice).

CC BY NC SA

Summary

Key Discoveries of Slides/Notes 21:

- For the coupon collector problem $X = X_1 + X_2 + \cdots + X_n$ where $X_i \sim \text{Geometric}$
  - $\mathbb{E}[X] \approx n(\ln n + \gamma_E)$
- $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$
- $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- $\sigma(X) = \sqrt{\text{Var}(X)}$
- $\text{Var}(X + Y) = \text{Var}(X) + 2\mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y] + \text{Var}(Y)$
- If X and Y independent:
  - $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$
  - $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

In the next set of slides/notes, we'll talk about the variance of two random variables that are not independent.

CC BY NC SA

Covariance

Lecture 22, CS70 Fall 2026

Covariance

- Definition
- Relationship with Independence
- Detour: Standardizing Random Variables
- Property of Covariance, Correlation

CC BY NC SA

Covariance

Earlier, we showed that:

$$\begin{aligned}\mathbf{Var}(X+Y) &= \mathbb{E}[(X+Y)^2] - (\mathbb{E}[X+Y])^2 \\ &= \mathbf{Var}(X) + 2\mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbf{Var}(Y)\end{aligned}$$

If X and Y are independent, then these blue terms cancel out.

If X and Y are not independent, the variance of the sum will depend on this term.

- This term (excluding the factor of 2) is called the **covariance**.

CC BY NC SA

Covariance Definition

The **covariance** of two random variables X and Y is defined as:

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

We already know a couple of facts about the covariance:

- $\mathbf{Var}(X+Y) = \mathbf{Var}(X) + \mathbf{Var}(Y) + 2\mathbf{cov}(X,Y)$
- If X and Y are independent, then $\mathbf{cov}(X,Y) = 0$. The converse is not true: $\mathbf{cov}(X,Y) = 0$ does not mean that X and Y are independent.

CC BY NC SA

Covariance Intuitively

Suppose we have a normal playing card deck. Let R_1 be 1 if the first draw is a red card (♥ or ♦), let B_2 be 1 if the second draw is a black card (♠ or ♣).

The covariance is given by $\mathbf{cov}(R_1,\ B_2) = \mathbb{E}[R_1 B_2] - \mathbb{E}[R_1]\mathbb{E}[B_2]$

cc BY NC SA

Covariance Intuitively

Suppose we have a normal playing card deck. Let R_1 be 1 if the first draw is a red card (♥ or ♦), let B_2 be 1 if the second draw is a black card (♠ or ♣).

The covariance is given by $\mathbf{cov}(R_1,\ B_2) = \mathbb{E}[R_1 B_2] - \mathbb{E}[R_1]\mathbb{E}[B_2]$

- $\mathbb{E}[R_1] = 1/2$
- $\mathbb{E}[B_2] = 1/2$

Then we have:

$$\begin{aligned}\mathbb{E}[R_1 B_2] = {} & 0\times 0\times P(R_1 = 0, B_2 = 0)+ \\ & 0\times 1\times P(R_1 = 0, B_2 = 1)+ \\ & 1\times 0\times P(R_1 = 1, B_2 = 0)+ \\ & 1\times 1\times P(R_1 = 1, B_2 = 1)\end{aligned}$$

Note: As we've seen before, the $\mathbb{E}$ of the product of two indicator variables is just the probability they are both true.

CC BY NC SA

Covariance Intuitively

Suppose we have a normal playing card deck. Let R_1 be 1 if the first draw is a red card (♥ or ♦), let B_2 be 1 if the second draw is a black card (♠ or ♣).

The covariance is given by $\mathbf{cov}(R_1,\ B_2) = \mathbb{E}[R_1 B_2] - \mathbb{E}[R_1]\mathbb{E}[B_2]$

- $\mathbb{E}[R_1] = 1/2$
- $\mathbb{E}[B_2] = 1/2$

Then we have: $\mathbb{E}[R_1 B_2] = P(R_1 = 1, B_2 = 1)$

$$\begin{aligned} P(R_1 = 1, B_2 = 1) &= P(B_2 = 1 | R_1 = 1) \times P(R_1 = 1) \\ &= 26/51 \times 26/52 \\ &= 13/51 \end{aligned}$$

cc BY NC SA

Covariance Intuitively

Suppose we have a normal playing card deck. Let R_1 be 1 if the first draw is a red card (♥ or ♦), let B_2 be 1 if the second draw is a black card (♠ or ♣).

The covariance is given by $\mathbf{cov}(R_1,\ B_2) = \mathbb{E}[R_1 B_2] - \mathbb{E}[R_1]\mathbb{E}[B_2]$

- $\mathbb{E}[R_1] = 1/2$
- $\mathbb{E}[B_2] = 1/2$
- $\mathbb{E}[R_1 B_2] = 13/51$

Thus we have $\mathbf{cov}(R_1,\ B_2) = 13/51 - 1/2 \times 1/2 \approx 0.0049$

The positive covariance means that the random variables tend to move together, e.g. if R_1 is 1, B_2 tends to be 1 as well.

cc BY NC SA

Covariance Intuitively (Example 2)

Suppose we have a normal playing card deck. Let R_1 be 1 if the first draw is a red card (♥ or ♦), let R_2 be 1 if the second draw is also a red card.

The covariance is given by $\mathbf{cov}(R_1,\ R_2) = \mathbb{E}[R_1 R_2] - \mathbb{E}[R_1]\mathbb{E}[R_2]$

Do we expect this covariance to be positive or negative?

- In other words, if the first draw is red, is it more likely or less likely the second card will be red?

CC BY NC SA

Covariance Intuitively (Example 2)

Suppose we have a normal playing card deck. Let R_1 be 1 if the first draw is a red card (♥ or ♦), let R_2 be 1 if the second draw is also a red card.

The covariance is given by $\mathbf{cov}(R_1,\ R_2) = \mathbb{E}[R_1 R_2] - \mathbb{E}[R_1]\mathbb{E}[R_2]$

- $\mathbb{E}[R_1] = 1/2$
- $\mathbb{E}[R_2] = 1/2$
- $\mathbb{E}[R_1 R_2] = 25/102$ (not shown)

Thus we have $\mathbf{cov}(R_1, R_2) = 25/102 - 1/2 \times 1/2 \approx -0.0049$

The negative covariance means that the random variables tend to move apart, e.g. if R_1 is 1, R_2 tends to avoid being 1 as well.

cc BY NC SA

Random Variables and Sampled Data

Let M be the Midterm 2 score of a student in CS61B, Fall 2026.

- Observation: M is a random variable. It's just as real as the Bernoulli, Binomial, Geometric, etc. distributions.
- What would M look like? What are the values for a? $P(M = a)$?

a	$P(X = a)$
0	0.016
1	0.094
2	0.234
3	0.313
4	0.234
5	0.094
6	0.016

Distribution for $\mathbf{X} \sim \mathbf{Binomial}(6,\ 0.5)$

a	$P(M = a)$
?	?
?	?
?	?
?	?
?	?
?	?
?	?

CC BY NC SA

Random Variables and Sampled Data

Let M be the Midterm 2 score of a student in CS61B, Fall 2026.

- Observation: M is a random variable. It's just as real as the Bernoulli, Binomial, Geometric, etc. distributions.
- What would M look like? What are the values for a? $P(M = a)$?
  - range(M): Set of scores on the midterm.
  - $P(M = a)$: Fraction of students with that score.

a	$P(X = a)$
0	0.016
1	0.094
2	0.234
3	0.313
4	0.234
5	0.094
6	0.016

Distribution for $\mathbf{X} \sim \mathbf{Binomial}(6,\ 0.5)$

a	$P(M = a)$
0	1/342
28	1/342
83.67	1/342
88.83	1/342
...	
1082	1/342
1088	1/342
1095	1/342
1098	1/342

CC BY NC SA

Random Variables and Sampled Data

Let M be the Midterm 2 score of a student in CS61B, Fall 2026.

- What would M look like? What are the values for a? $P(M = a)$?
  - range(M): Set of scores on the midterm.
  - $P(M = a)$: Fraction of students with that score.
- What is the sample space Ω?

a	$P(X = a)$
0	0.016
1	0.094
2	0.234
3	0.313
4	0.234
5	0.094
6	0.016

Distribution for $\mathbf{X} \sim \mathbf{Binomial}(6,\ 0.5)$

https://joshh.ug/A

a	$P(M = a)$
0	1/342
28	1/342
83.67	1/342
88.83	1/342
...	
1082	1/342
1088	1/342
1095	1/342
1098	1/342

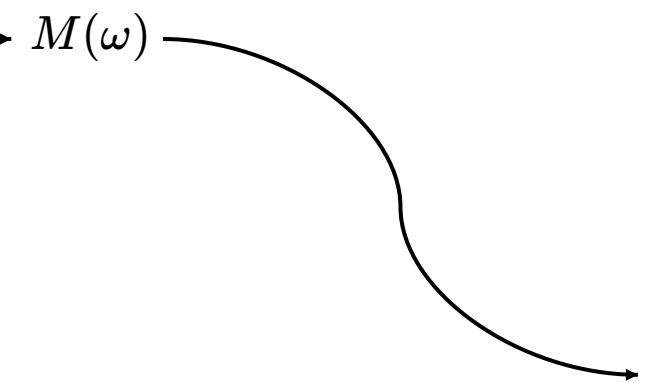
Random Variables and Sampled Data

Let M be the Midterm 2 score of a student in CS61B, Fall 2026.

- What would M look like? What are the values for a? $P(M = a)$?
  - range(M): Set of scores on the midterm.
  - $P(M = a)$: Fraction of students with that score.
- What is the sample space Ω?
  - The set of all 342 students in our CS61B data.

$\Omega =$

Aleš Balqis
Eskender Yann
Yawen Jin
...
Terence Tao
Roshan Sunitha
Francisco Kali



a	$P(M = a)$
0	1/342
28	1/342
83.67	1/342
88.83	1/342
...	
1082	1/342
1088	1/342
1095	1/342
1098	1/342

For privacy reasons, I've made up 342 new names.

CC BY NC SA

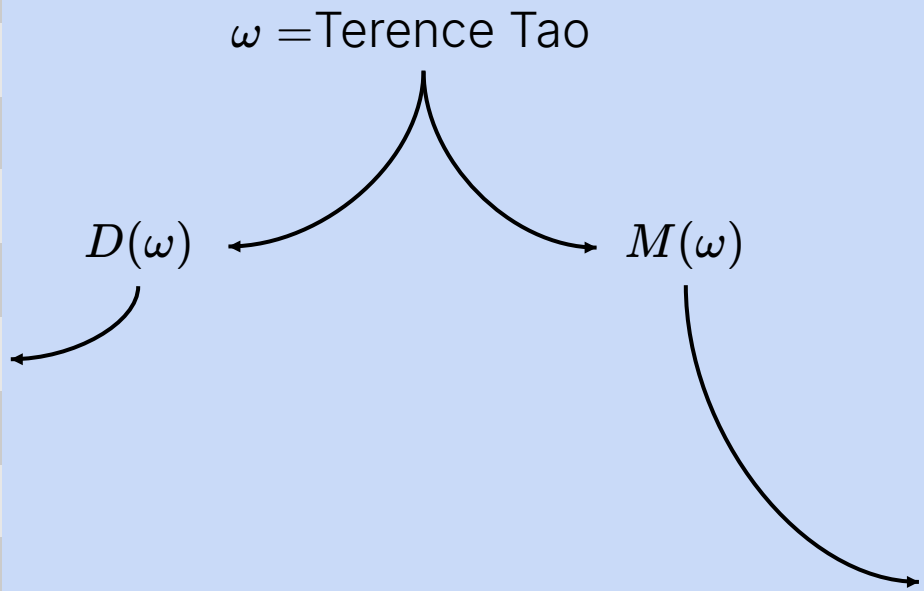
Random Variables and Sampled Data

Let D be the discussion attendance score of a CS61B, Fall 2026 student.

Suppose we want to compute the covariance of D and M.

- What does it mean if these RVs have positive covariance?

a	$P(D = a)$
15	1/342
30	1/342
45	5/342
60	12/342
75	16/342
90	35/342
105	64/342
120	115/342
135	93/342



a	$P(M = a)$
0	1/342
28	1/342
83.67	1/342
88.83	1/342
...	
1082	1/342
1088	1/342
1095	1/342
1098	1/342

CC BY NC SA

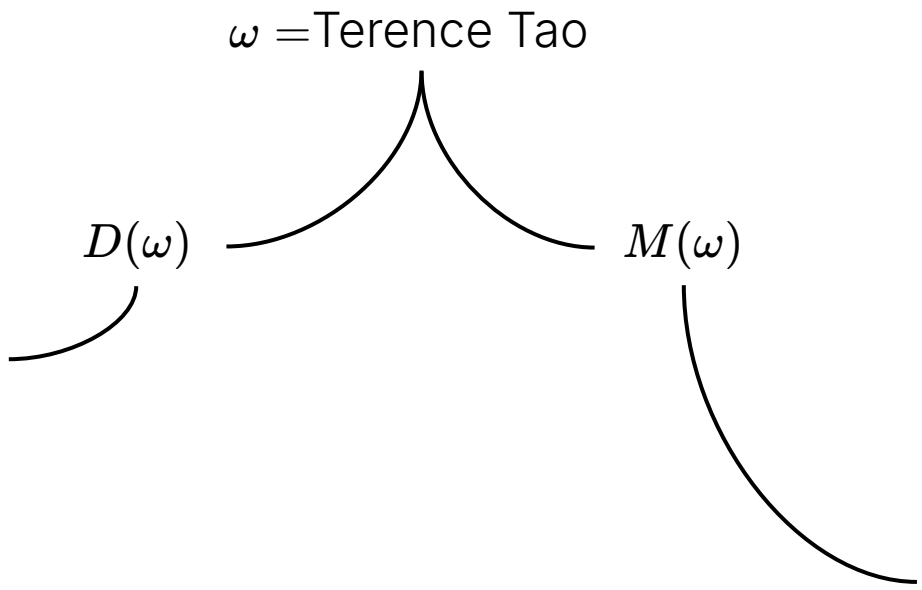
Random Variables and Sampled Data

Let D be the discussion attendance score of a CS61B, Fall 2026 student.

Suppose we want to compute the covariance of D and M.

- What does it mean if these RVs have positive covariance? Students who go to discussion tend to also do better on exams. Note: No causality established!

a	$P(D = a)$
15	1/342
30	1/342
45	5/342
60	12/342
75	16/342
90	35/342
105	64/342
120	115/342
135	93/342



a	$P(M = a)$
0	1/342
28	1/342
83.67	1/342
88.83	1/342
...	
1082	1/342
1088	1/342
1095	1/342
1098	1/342

CC BY NC SA

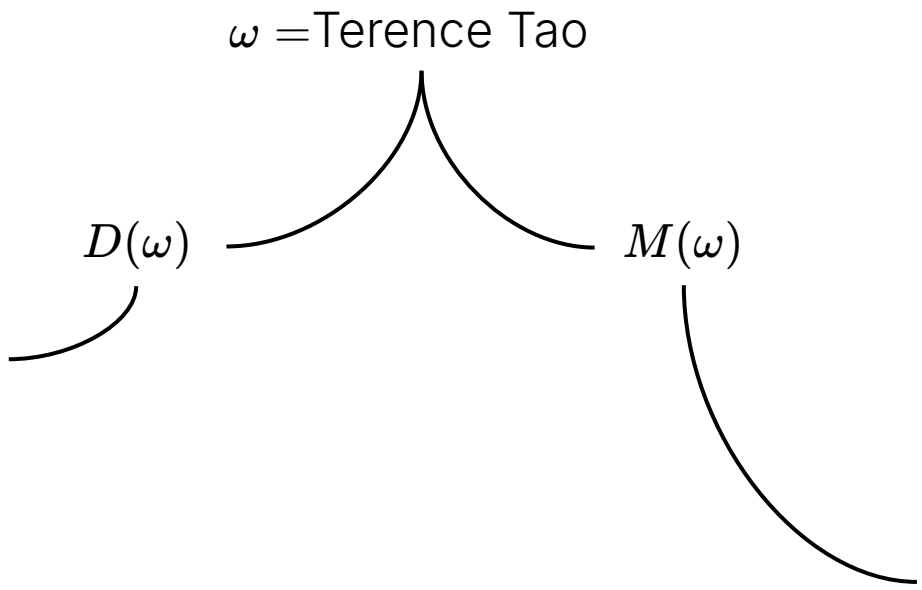
Random Variables and Sampled Data

Let D be the discussion attendance score of a CS61B, Fall 2026 student.

Suppose we want to compute the covariance of D and M.

- Requires computing $\mathbb{E}[DM]$.

a	$P(D = a)$
15	1/342
30	1/342
45	5/342
60	12/342
75	16/342
90	35/342
105	64/342
120	115/342
135	93/342



a	$P(M = a)$
0	1/342
28	1/342
83.67	1/342
88.83	1/342
...	
1082	1/342
1088	1/342
1095	1/342
1098	1/342

CC BY NC SA

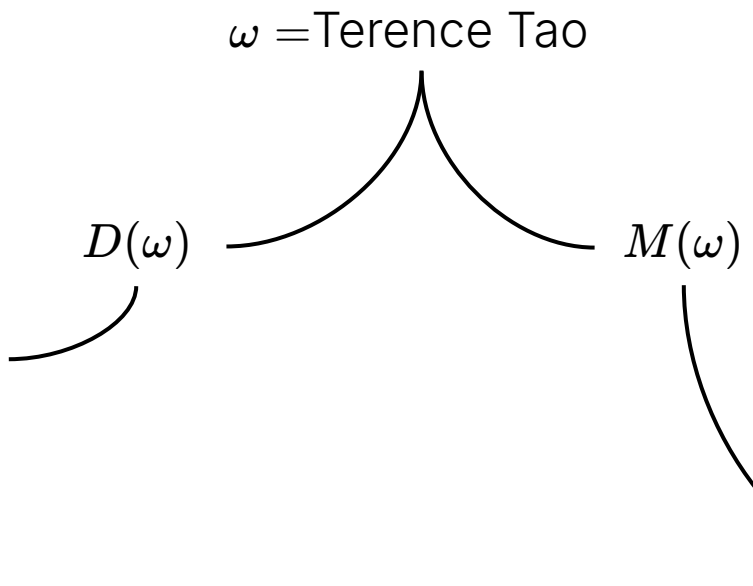
Random Variables and Sampled Data

Suppose we want to compute the covariance of D and M.

- $\mathbb{E}[M] = 660.462$
- $\mathbb{E}[D] = 112.325$
- $\mathbb{E}[DM] = 75631.367$

$$\begin{aligned}\text{cov}(D, M) &= \mathbb{E}[DM] - \mathbb{E}[D]\mathbb{E}[M] \\ &= 75631.367 - 74186.048 \\ &= 1445.319\end{aligned}$$

a	$P(D = a)$
15	1/342
30	1/342
45	5/342
60	12/342
75	16/342
90	35/342
105	64/342
120	115/342
135	93/342



a	$P(M = a)$
0	1/342
28	1/342
83.67	1/342
88.83	1/342
...	
1082	1/342
1088	1/342
1095	1/342
1098	1/342

$\mathbb{E}[DM]$ = `np.mean(df["Midterm"]*df["Discussion"])`

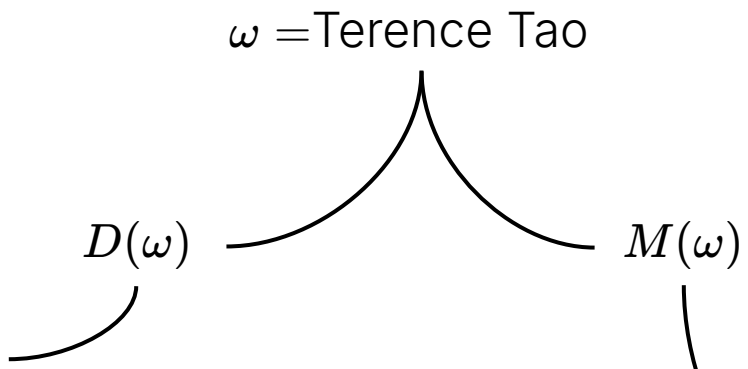
CC BY NC SA

Random Variables and Sampled Data

Suppose we want to compute the covariance of D and M.

- The covariance is ~1445.3. This implies that D and M tend to increase together.
- In other words, attending discussion is positively correlated with midterm score.
- The 1445.3 is hard to interpret directly, we'll come back to it later.

a	$P(D = a)$
15	1/342
30	1/342
45	5/342
60	12/342
75	16/342
90	35/342
105	64/342
120	115/342
135	93/342



a	$P(M = a)$
0	1/342
28	1/342
83.67	1/342
88.83	1/342
...	
1082	1/342
1088	1/342
1095	1/342
1098	1/342

$\mathbb{E}[DM]$ `= np.mean(df["Midterm"]*df["Discussion"])`

CC BY NC SA

Quick Minor Technical Notes

550 CS61B students took Midterm 2 (out of 1100 points).

- Discussion attendance is opt-in: 482 students have attendance records, with 15 points per discussion attended (9 discussions before the midterm).
- 407 students are in both groups, but 65 of them have 0 attendance points (most have just one or two "absent" records).

In our case, I took Ω to be the 342 students who took Midterm 2 and attended at least one discussion, so that students who aren't really doing discussion attendance don't distort the covariance.

CC BY NC SA

Covariance and Independence

Lecture 22, CS70 Fall 2026

Covariance

- Definition
- Relationship with Independence
- Detour: Standardizing Random Variables
- Property of Covariance, Correlation

CC BY NC SA

Example

Suppose we have two random variables with joint distribution given by:

Y \ X	-1	0	1
1	0	1/5	0
0	1/5	1/5	1/5
-1	0	1/5	0

Question 1: Are X and Y independent?

https://joshh.ug/B



CC BY NC SA

Example

Suppose we have two random variables with joint distribution given by:

Y \ X	-1	0	1
1	0	1/5	0
0	1/5	1/5	1/5
-1	0	1/5	0

Question 1: Are X and Y independent?

- No, knowing something about one value tells you about the other.
- For example, if we know $X = 0$, then Y could be anything, whereas if $X = 1$, we know Y is 0.
- Or in terms of specific probabilities: $P(X = 1,\ Y = 1) \neq P(X = 1) \times P(Y = 1)$

CC BY NC SA

Example

Suppose we have two random variables with joint distribution given by:

Y \ X	-1	0	1
1	0	1/5	0
0	1/5	1/5	1/5
-1	0	1/5	0

Question 2: What is $\mathbf{cov}(X,\ Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$?

- joshh.ug/C

https://joshh.ug/C

Example

Suppose we have two random variables with joint distribution given by:

Y \ X	-1	0	1
1	0	1/5	0
0	1/5	1/5	1/5
-1	0	1/5	0

Question 2: What is $\mathbf{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$?

- $\mathbb{E}[XY] = 0$, at least one variable is always zero.
- $\mathbb{E}[X] = -1 \times 1/5 + 1 \times 1/5 = 0$
- $\mathbb{E}[Y] = -1 \times 1/5 + 1 \times 1/5 = 0$

Zero covariance does not imply independence!

CC BY NC SA

Covariance and Independence

Y \ X	-1	0	1
1	0	1/5	0
0	1/5	1/5	1/5
-1	0	1/5	0

If two random variables are independent, their covariance is zero.

- However, zero covariance can occur even with dependency.

CC BY NC SA

Standardizing Random Variables

Lecture 22, CS70 Fall 2026

Covariance

- Definition
- Relationship with Independence
- Detour: Standardizing Random Variables
- Property of Covariance, Correlation

CC BY NC SA

Standardizing a Random Variable

If X is a random variable with mean μ and stdev σ, then the standardized version of this random variable is given by $\tilde{X}$:

$$\tilde{X} = \frac{X - \mu}{\sigma}$$

The standardized RV is a shifted and rescaled version of X such that $\mathbb{E}[\tilde{X}] = 0$ and $\mathbf{Var}(\tilde{X}) = 1$.

CC BY NC SA

Example: Standardizing Binomial(4, 0.25)

Suppose $X \sim \text{Binomial}(4,\ 0.25)$. We know from earlier that:

CC BY NC SA

Example: Standardizing Binomial(4, 0.25)

Suppose $X \sim \text{Binomial}(4,\ 0.25)$. We know from earlier that:

$$\mathbb{E}[X] = np = 1$$
$$\text{Var}(X) = np(1-p) = 4 \times 0.25 \times 0.75 = 0.75$$

Then we have that $\tilde{X}$ is given by:

$$\tilde{X} = \frac{X - 1}{\sqrt{0.75}}$$

CC BY NC SA

Standardized Binomial

The probabilities for a standardized random variable are exactly the same as the non-standardized version. The only difference is that the range of the RV is shifted and rescaled. Example for $X \sim \text{Binomial}(4,\ 0.25)$ and $\tilde{X}$ below.

- Note: The values in the range of the new RV are what you may have heard called "z-scores".

a	$P(X = a)$	ω
0	31.6%	TTTT
1	42.2%	HTTT, THTT, TTHT, TTTH
2	21.1%	HHTT, HTHT, HTTH, THHT, THTH, TTHH
3	4.7%	HHHT, HHTH, HTHH, THHH
4	0.4%	HHHH

a	$P(\tilde{X} = a)$	ω
-1.155	31.6%	TTTT
0	42.2%	HTTT, THTT, TTHT, TTTH
1.155	21.1%	HHTT, HTHT, HTTH, THHT, THTH, TTHH
2.309	4.7%	HHHT, HHTH, HTHH, THHH
3.464	0.4%	HHHH

CC BY NC SA

Standardized RV Expectation and Variance

If X is a random variable with mean μ and stdev σ, and $\tilde{X} = \frac{X-\mu}{\sigma}$, then:

$$\mathbb{E}[\tilde{X}] = \mathbb{E}\left[\frac{X-\mu}{\sigma}\right]$$

CC BY NC SA

Standardized RV Expectation and Variance

If X is a random variable with mean μ and stdev σ, and $\tilde{X} = \frac{X-\mu}{\sigma}$, then:

$$\mathbb{E}[\tilde{X}] = \mathbb{E}\left[\frac{X-\mu}{\sigma}\right] = \frac{\mu-\mu}{\sigma} = 0$$

$$\begin{aligned}\mathrm{Var}(\tilde{X}) = \mathrm{Var}\left(\frac{X-\mu}{\sigma}\right) &= \frac{1}{\sigma^2}\mathrm{Var}(X-\mu) \\ &= \frac{1}{\sigma^2}\mathrm{Var}(X) \\ &= \frac{\sigma^2}{\sigma^2} = 1\end{aligned}$$

CC BY NC SA

Another Useful Property of Standardized Random Variables

If X is a random variable with mean μ and stdev σ, and $\tilde{X} = \frac{X-\mu}{\sigma}$, what is $\mathbb{E}[\tilde{X}^2]$?

Note: We just showed that:

- $\mathbb{E}[\tilde{X}] = 0$
- $\mathbf{Var}(\tilde{X}) = 1$

Answer at joshh.ug/D

https://joshh.ug/D

Another Useful Property of Standardized Random Variables

If X is a random variable with mean μ and stdev σ, and $\tilde{X} = \frac{X-\mu}{\sigma}$, what is $\mathbb{E}[\tilde{X}^2]$?

Observation: Since $\mathbb{E}[\tilde{X}] = 0$, $\mathbf{Var}(\tilde{X}) = \mathbb{E}[\tilde{X}^2]$. This is because:

$$\begin{aligned}\mathbf{Var}(\tilde{X}) &= \mathbb{E}[\tilde{X}^2] - \mathbb{E}[\tilde{X}]^2 \\ &= \mathbb{E}[\tilde{X}^2]\end{aligned}$$

Since $\mathbf{Var}(\tilde{X}) = \mathbb{E}[\tilde{X}^2]$, this means $\mathbb{E}[\tilde{X}^2] = 1$.

CC BY NC SA

Use of Standardized Random Variables

Standardizing random variables can be useful for many purposes. For example:

- Comparing exam distributions between different semesters or different courses.
  - Maximum z-score on Fa24 CS70 midterm: 2.82
  - Maximum z-score on Fa25 CS70 midterm: 1.77
- Putting different independent variables on to the same scale when training machine learning models.
- Writing proofs: Standardized random variables have useful mathematical properties, like the ones we've just derived.
  - $\mathbb{E}[\tilde{X}] = 0$
  - $\mathbb{E}[\tilde{X}^2] = \mathbf{Var}(\tilde{X}) = 1$

CC BY NC SA

Properties of Covariance, Correlation

Lecture 22, CS70 Fall 2026

Covariance

- Definition
- Relationship with Independence
- Detour: Standardizing Random Variables
- Property of Covariance, Correlation

CC BY NC SA

Variance

In previous lecture, we showed three properties of variance $\mathbb{E}[(X - \mathbb{E}[X])^2]$:

- Can rewrite as $\mathbf{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- $\mathbf{Var}(cX) = c^2\mathbf{Var}(X)$
- $\mathbf{Var}(c + X) = \mathbf{Var}(X)$

The covariance $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ has similar properties:

- Can rewrite as $\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$
- $\mathbf{cov}(X + a,\ Y + b) = \mathbf{cov}(\mathrm{X}, \mathrm{Y})$
- $\mathbf{cov}(cX,\ dY) = \mathrm{c} \cdot d \cdot \mathbf{cov}(\mathrm{X}, \mathrm{Y})$

Proofs are similar to variance property proofs.

- Proofs of all three are in the extra slides at the end.

CC BY NC SA

Two More Properties of Covariance

The notes give two additional properties:

- $\mathbf{cov}(X, X) = \mathbf{Var}(X)$
- Bilinearity: for random variables $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ and constants $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$,

$$\mathbf{cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \,\mathbf{cov}(X_i, Y_j)$$

As with the notes, we will not prove these properties.

CC BY NC SA

Bilinearity of Covariance

Let's see an example using these two properties:

- $\mathbf{cov}(X, X) = \mathbf{Var}(X)$
- Bilinearity: for random variables $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ and constants $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$,

$$\mathbf{cov}\left(\sum_{i=1}^{n} a_i X_i,\ \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \,\mathbf{cov}(X_i, Y_j)$$

Example: Write $\mathbf{cov}(X + 2Y,\ 3X - Y)$ in terms of $\mathbf{Var}(X)$, $\mathbf{Var}(Y)$, and $\mathbf{cov}(X, Y)$.

Bilinearity of Covariance

Let's see an example using these two properties:

- $\operatorname{cov}(X, X) = \operatorname{Var}(X)$
- Bilinearity: for random variables $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ and constants $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$,

$$\operatorname{cov}\left(\sum_{i=1}^{n} a_i X_i,\ \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \operatorname{cov}(X_i, Y_j)$$

Example: Write $\operatorname{cov}(X + 2Y,\ 3X - Y)$ in terms of $\operatorname{Var}(X)$, $\operatorname{Var}(Y)$, and $\operatorname{cov}(X, Y)$.

$$\begin{aligned}\operatorname{cov}(X + 2Y,\ 3X - Y) &= 1 \cdot 3 \operatorname{cov}(X, X) + 1 \cdot (-1) \operatorname{cov}(X, Y) \\ &\quad + 2 \cdot 3 \operatorname{cov}(Y, X) + 2 \cdot (-1) \operatorname{cov}(Y, Y) \\ &= 3 \operatorname{Var}(X) - \operatorname{cov}(X, Y) + 6 \operatorname{cov}(Y, X) - 2 \operatorname{Var}(Y) \\ &= 3 \operatorname{Var}(X) + 5 \operatorname{cov}(X, Y) - 2 \operatorname{Var}(Y)\end{aligned}$$

CC BY NC SA

Bilinearity of Covariance

Let's see an example using these two properties:

- $\operatorname{cov}(X, X) = \operatorname{Var}(X)$
- Bilinearity: for random variables $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ and constants $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$,

$$\operatorname{cov}\left(\sum_{i=1}^{n} a_i X_i,\ \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \operatorname{cov}(X_i, Y_j)$$

Example: Write $\operatorname{cov}(X + 2Y,\ 3X - Y)$ in terms of $\operatorname{Var}(X)$, $\operatorname{Var}(Y)$, and $\operatorname{cov}(X, Y)$.

$$\operatorname{cov}(X + 2Y,\ 3X - Y) = 3\operatorname{Var}(X) + 5\operatorname{cov}(X, Y) - 2\operatorname{Var}(Y)$$

CC BY NC SA

Correlation

As we saw earlier, the covariance of random variables can take on arbitrary values. Example: Covariance of discussion and midterm scores: $\mathbf{cov}(D, M) \approx 1445.3$

- Units of covariance are the product of the units of X times the units of Y.

The **correlation** is a rescaled version of the covariance. If X and Y are random variables with standard deviations $\sigma_X > 0$ and $\sigma_Y > 0$ respectively, then the **correlation** of X and Y is given by:

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}$$

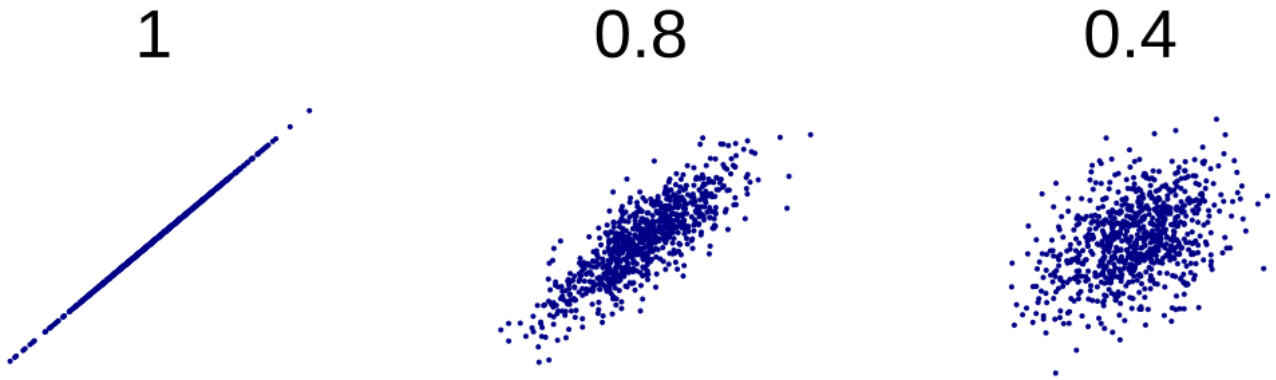
Why do we care?

- Always lies between -1 and +1. More intuitive to interpret.

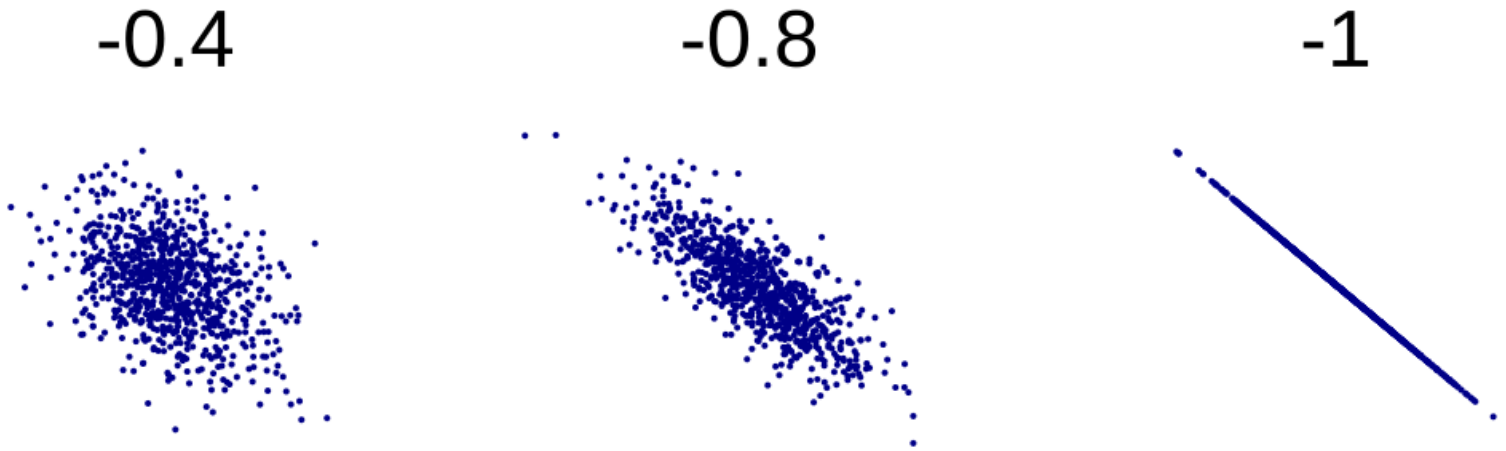
CC BY NC SA

Correlation Examples

Examples of positively correlated random variables, $\mathbf{Corr}(X, Y) > 0$:



Examples of negatively correlated random variables, $\mathbf{Corr}(X, Y) < 0$:

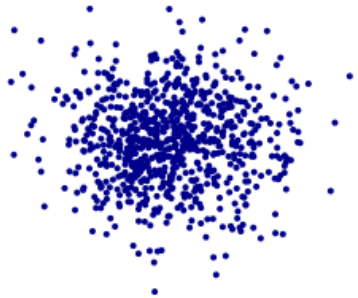


CC BY NC SA

Correlation Of Zero

Examples of independent RV with $\mathbf{Corr}(X,\ Y) = 0$:

0



Example of dependent RV with $\mathbf{Corr}(X,\ Y) = 0$. Why are these not independent?

0



CC BY NC SA

Correlation Of Discussion and Midterm Scores

We can compute the correlation of midterm and discussion scores as:

$$\text{corr}(D,\ M) = \frac{\mathbb{E}[DM] - \mathbb{E}[D]\mathbb{E}[M]}{\sigma_D \sigma_M}$$

All 407 students: −0.01

On actual 61B data this semester, this correlation is 0.26.

a	$P(D = a)$
15	1/342
30	1/342
45	5/342
60	12/342
75	16/342
90	35/342
105	64/342
120	115/342
135	93/342

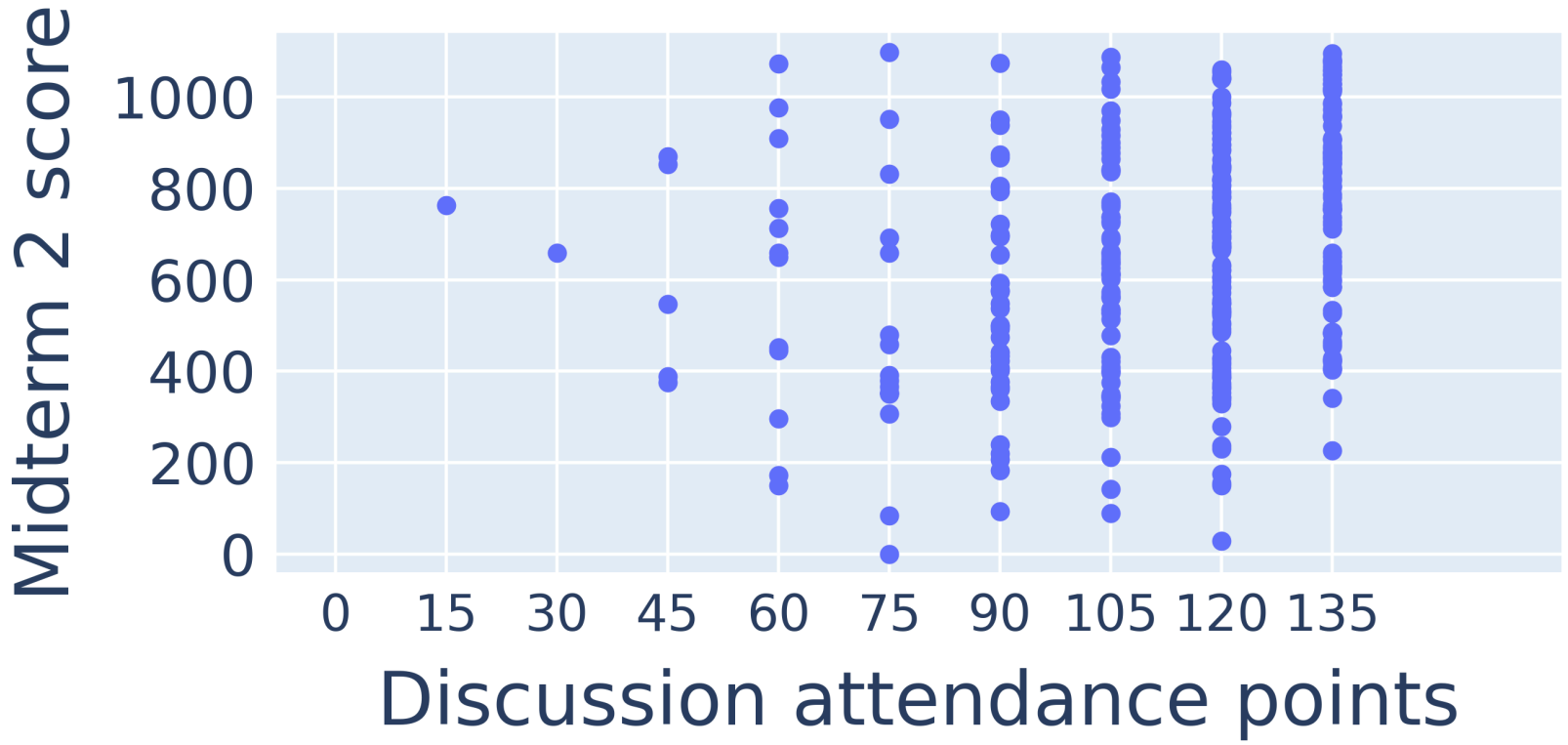
a	$P(M = a)$
0	1/342
28	1/342
83.67	1/342
88.83	1/342
...	
1082	1/342
1088	1/342
1095	1/342
1098	1/342

```
((mt*dis).mean() - mt.mean() * dis.mean()) / (mt.std()*dis.std())
```

Scatter Plot (Beyond Scope of Course)

If we scatterplot midterm vs. discussion score, we get the plot below.

- Can kind of see an upward relationship, though the “overplotting” makes this tricky since overlapping points obscure each other.

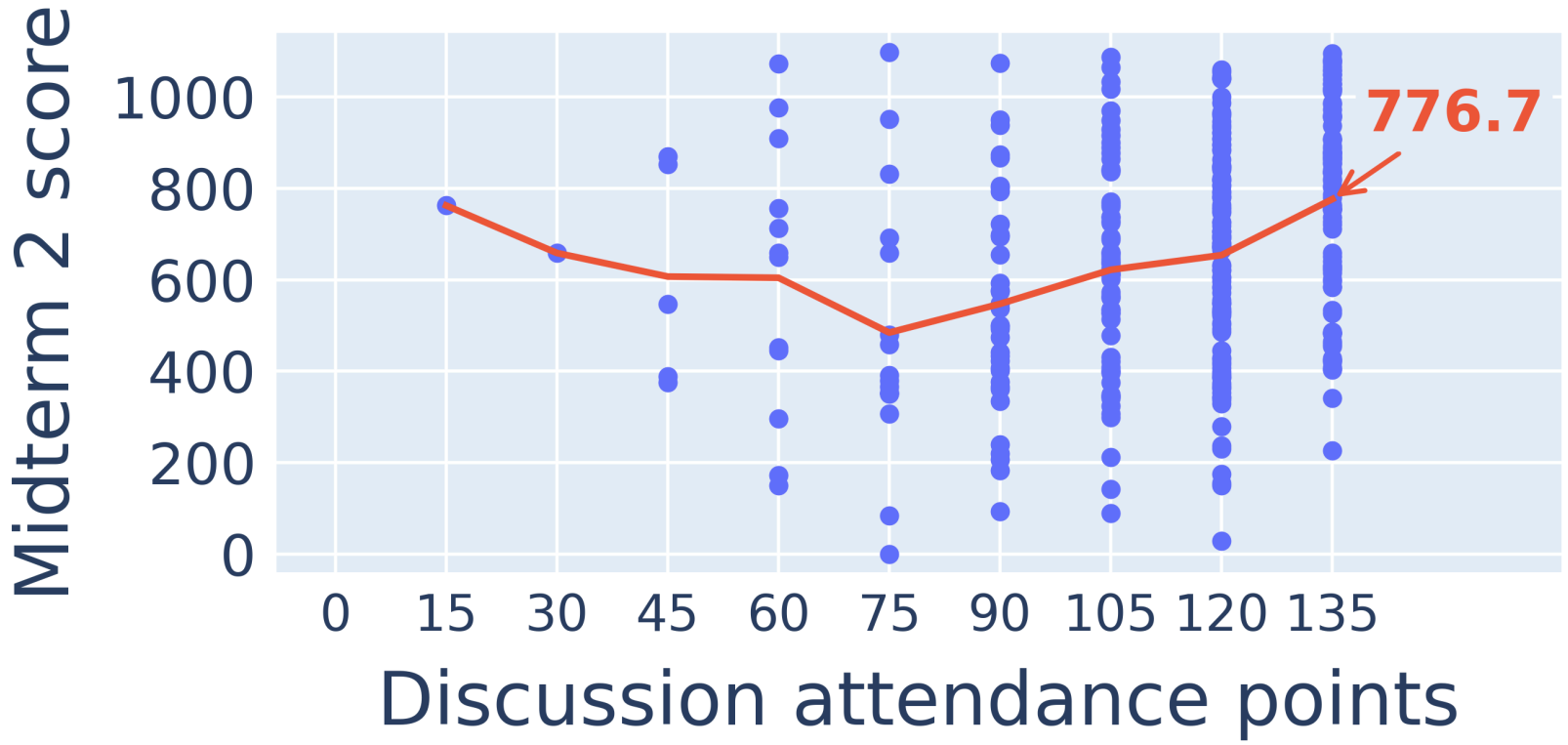


cc BY NC SA

Scatter Plot (Beyond Scope of Course)

Can see the relationship a bit better if we overlay the average midterm score per possible discussion value.

- Example: Average midterm score of the 93 students with a discussion score of 135 is 776.7.



CC BY NC SA

Summary

- $\mathbf{cov}(X,\ Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
  - Can rewrite as $\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$
  - $\mathbf{cov}(X + a,\ Y + b) = \mathbf{cov}(X, Y)$
  - $\mathbf{cov}(cX,\ dY) = c \cdot d \cdot \mathbf{cov}(X, Y)$
  - Bilinearity (given in Note 22):
    $\mathbf{cov}\left(\sum_{i=1}^{n} a_i X_i,\ \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \,\mathbf{cov}(X_i, Y_j)$
- $\mathbf{Var}(X + Y) = \mathbf{Var}(X) + \mathbf{Var}(Y) + 2\mathbf{cov}(X, Y)$
- If X and Y are independent, then $\mathbf{cov}(X, Y) = 0$. The converse is not true.
- Covariance is in units of X times units of Y.
- Standardized random variables have mean 0 and variance 1.
- Correlation is always between -1 and 1.

CC BY NC SA

Extra Slide: Proof of First Useful Property

$\mathbf{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$:

- Can also be written as $\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$

Proof: For conciseness of notation, let $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$.

$$\begin{aligned}
\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] &= \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \\
&= \mathbb{E}[XY - \mu_Y X - \mu_X Y + \mu_X \mu_Y] \\
&= \mathbb{E}[XY] - \mu_Y \mathbb{E}[X] - \mu_X \mathbb{E}[Y] + \mu_X \mu_Y \\
&= \mathbb{E}[XY] - \mu_X \mu_Y - \mu_X \mu_Y + \mu_X \mu_Y \\
&= \mathbb{E}[XY] - \mu_X \mu_Y
\end{aligned}$$

CC BY NC SA

Extra Slide: Proof of Second Useful Property

$$\text{cov}(X + a,\ Y + b) = \text{cov}(X, Y)$$

Proof: Use the first property, along with $\mathbb{E}[X + a] = \mathbb{E}[X] + a$ and $\mathbb{E}[Y + b] = \mathbb{E}[Y] + b$.

$$\begin{aligned}
\text{cov}(X + a,\ Y + b) &= \mathbb{E}[(X + a - \mathbb{E}[X + a])(Y + b - \mathbb{E}[Y + b])] \\
&= \mathbb{E}[(X + a - \mathbb{E}[X] - a)(Y + b - \mathbb{E}[Y] - b)] \\
&= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\
&= \text{cov}(X, Y)
\end{aligned}$$

CC BY NC SA

Extra Slide: Proof of Third Useful Property

$$\text{cov}(cX,\ dY) = c \cdot d \cdot \text{cov}(X, Y)$$

Proof: Use linearity of expectation, i.e. $\mathbb{E}[cX] = c\mathbb{E}[X]$, $\mathbb{E}[dY] = d\mathbb{E}[Y]$ and $\mathbb{E}[cXdY] = cd\mathbb{E}[XY]$.

$$\begin{aligned}
\text{cov}(cX,\ dY) &= \mathbb{E}[cX \cdot dY] - \mathbb{E}[cX]\mathbb{E}[dY] \\
&= cd\,\mathbb{E}[XY] - c\mathbb{E}[X] \cdot d\mathbb{E}[Y] \\
&= cd\,(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) \\
&= cd \cdot \text{cov}(X, Y)
\end{aligned}$$

Extra Slide: Proof of Correlation Range

Claim: The correlation is always between -1 and 1.

$$-1 \leq \mathrm{Corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} \leq +1$$

Lemma: Let $\mathbb{E}[X] = \mu_X$, $\mathbb{E}[Y] = \mu_Y$. Let $\tilde{X}$ and $\tilde{Y}$ be the standardized versions of X and Y. Then $\mathrm{Corr}(X,Y) = \mathbb{E}[\tilde{X}\tilde{Y}]$

Proof of Lemma:

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} = \mathbb{E}\left[\frac{X-\mu_X}{\sigma_X}\frac{Y-\mu_Y}{\sigma_Y}\right] = \mathbb{E}[\tilde{X}\tilde{Y}]$$

CC BY NC SA

Extra Slide: Proof of Correlation Range

Claim: The correlation is always between -1 and 1.

$$-1 \leq \text{Corr}(X,Y) = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \leq +1$$ ✔

Lemma: Let $\mathbb{E}[X] = \mu_X$, $\mathbb{E}[Y] = \mu_Y$. Let $\tilde{X}$ and $\tilde{Y}$ be the standardized versions of X and Y. Then $\text{Corr}(X,Y) = \mathbb{E}[\tilde{X}\tilde{Y}]$

Proof that $1 \geq \text{Corr}(X,Y)$. Consider the quantity $\mathbb{E}[(\tilde{X} - \tilde{Y})^2]$, which is ≥ 0.

$$\begin{aligned}
\mathbb{E}[(\tilde{X} - \tilde{Y})^2] &= \mathbb{E}[\tilde{X}^2] - 2\mathbb{E}[\tilde{X}\tilde{Y}] + \mathbb{E}[\tilde{Y}^2] \\
&= 1 - 2\mathbb{E}[\tilde{X}\tilde{Y}] + 1 \\
&= 2 - 2\text{Corr}(X,\ Y)
\end{aligned}$$

Thus, $2 - 2\text{Corr}(X,\ Y) \geq 0$ So: $1 \geq \text{Corr}(X,\ Y)$

CC BY NC SA

Extra Slide: Proof of Correlation Range

Claim: The correlation is always between -1 and 1.

$$\checkmark \; -1 \leq \text{Corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \leq +1 \; \checkmark$$

Lemma: Let $\mathbb{E}[X] = \mu_X$, $\mathbb{E}[Y] = \mu_Y$. Let $\tilde{X}$ and $\tilde{Y}$ be the standardized versions of X and Y. Then $\text{Corr}(X, Y) = \mathbb{E}[\tilde{X}\tilde{Y}]$

Proof that $\text{Corr}(X, Y) \geq -1$. Consider the quantity $\mathbb{E}[(\tilde{X} + \tilde{Y})^2]$, which is ≥ 0.

$$\begin{aligned}
\mathbb{E}[(\tilde{X} + \tilde{Y})^2] &= \mathbb{E}[\tilde{X}^2] + 2\mathbb{E}[\tilde{X}\tilde{Y}] + \mathbb{E}[\tilde{Y}^2] \\
&= 1 + 2\mathbb{E}[\tilde{X}\tilde{Y}] + 1 \\
&= 2 + 2\text{Corr}(X,\ Y)
\end{aligned}$$

Thus, $2 + 2\text{Corr}(X,\ Y) \geq 0$ So: $\text{Corr}(X,\ Y) \geq -1$

cc BY NC SA

Extra Slide: Note on Unity Correlation

Suppose $\text{Corr}(X, Y) = 1$. In that case, we have that $\tilde{X} = \tilde{Y}$. Why?

- Because $\mathbb{E}[(\tilde{X} - \tilde{Y})^2] = 2 - 2\text{Corr}(X, Y)$
- If $\text{Corr}(X, Y) = 1$, then $\mathbb{E}[(\tilde{X} - \tilde{Y})^2] = 0$, so $\tilde{X} = \tilde{Y}$

Implication: if $\text{Corr}(X, Y) = 1$, then random variables X and Y are just rescaled and shifted versions of each other, i.e. there exist $a > 0$ and b such that:

$$Y = aX + b$$

Can show a similar result if $\text{Corr}(X, Y) = -1$, in which case $a < 0$.

CC BY NC SA