Coupon Collector and Expectation

Lecture 21, CS70 Summer 2025

Coupon Collector and Expectation

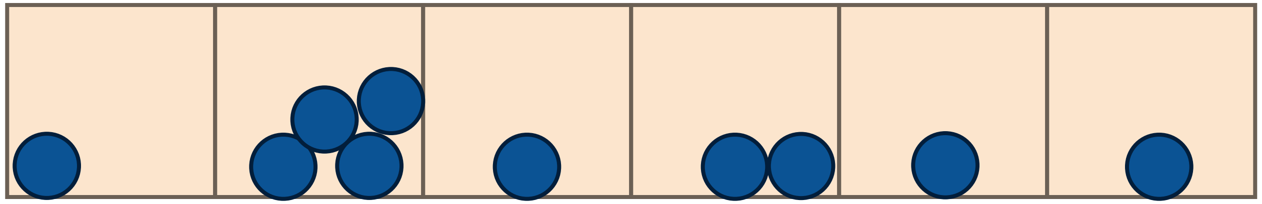
Variance

- Derivation and Definition
- Properties of Variance
- Example Variance Calculations
- Sums of Random Variables
- Covariance
- Covariance and Independence

Back to Balls and Bins

What if we throw balls into bins until every bin has at least one ball?

Or: randomly drawing items of n types to test and want to test every type.



Took 10 throws for 6 bins this time.

- In general: What's the distribution?
- How many throws before 50% probability of all bins having a ball?
- What's the expected value?

Problem Restatement – Coupon Collecting

In the coupon collecting problem, we imagine a contest where every time you buy a box of cereal, there is a coupon in the box.

- n different coupons.
- Once you collect all n coupons, you can redeem them all for a discount on your next cereal.

Let m be the number of boxes of cereal you buy. How many boxes of cereal must we buy, i.e., how big must m get before we have a 50% chance of getting all of the coupons?

Coupon Collector: Union Bound for Probability of Missing a Coupon

Define event A_i as the event where we are missing the ith coupon, and A as the event where we are missing at least one coupon. Then:

$$P(A) = P\left(\bigcup_{i=1}^{n} A_i\right)$$

A_i: Pick m boxes, coupon $\neq i$ with prob $1 - \frac{1}{n}$ each box: $P(A_i) = \left(1 - \frac{1}{n}\right)^m$

$$P(A) \leq n\left(1 - \frac{1}{n}\right)^m$$

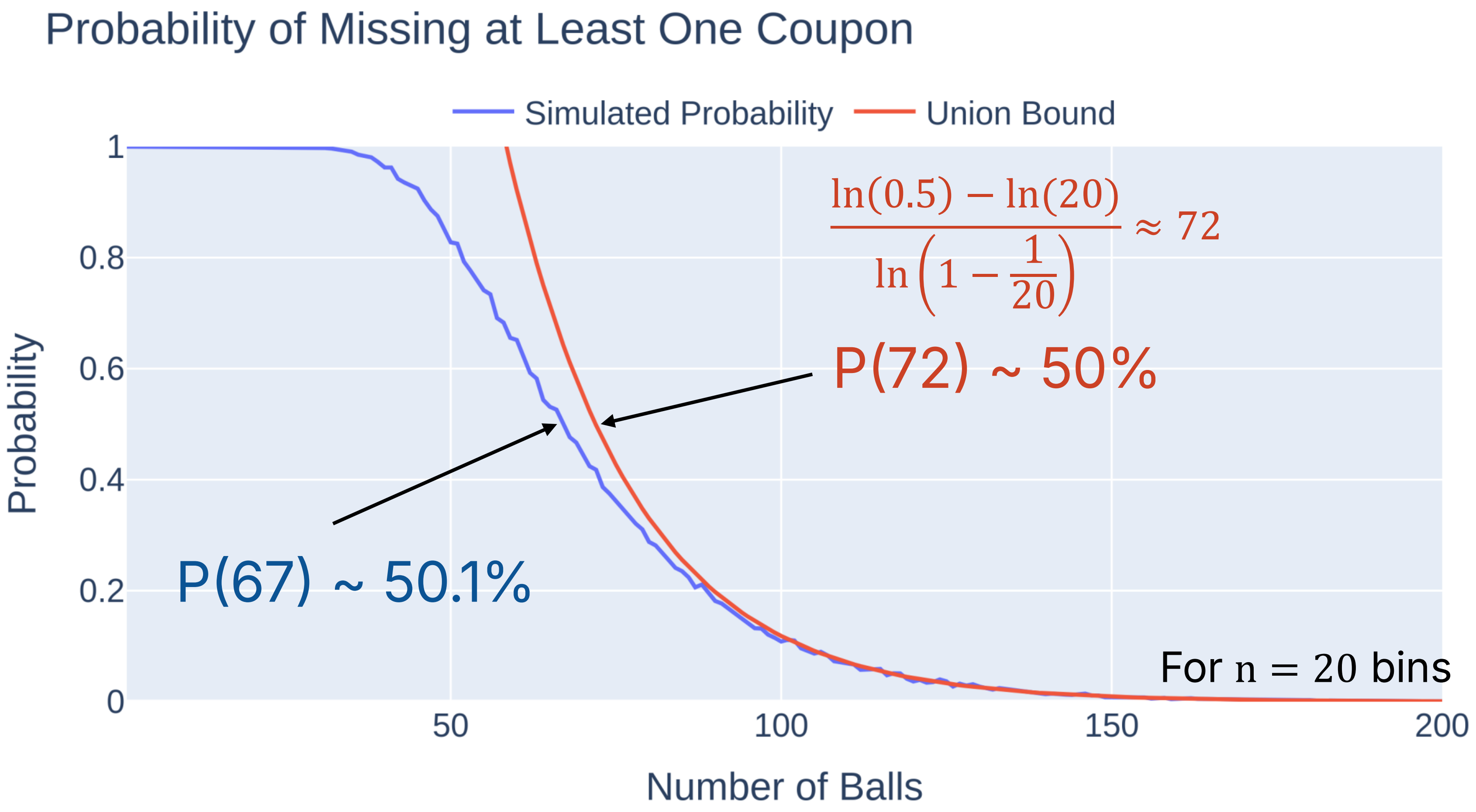
We want m_{50} such that this probability is at most 50%:

$$0.5 = n\left(1 - \frac{1}{n}\right)^{m_{50}} \longrightarrow \frac{\ln(0.5) - \ln(n)}{\ln\left(1 - \frac{1}{n}\right)} = m_{50}$$

Buying this many boxes ensures you have all coupons with probability $\geq 50\%$.

Coupon Collector: Union Bound vs. Empirical Result

We can see this upper bound overlaid on experiments with $n = 20$.



Can also show that with $n \ln n + n$ boxes, missing one with prob $\leq e^{-1} \approx 0.3679 \ldots$

Expected Value

What about the expected value of the number of boxes?

Let X be the number of cereal boxes we need to buy before we get all n coupons. Our goal: Compute $E[X]$.

The first key idea is to define $X = X_1 + X_2 + \cdots + X_n$, where X_i is the number of cereal boxes we need to buy before we get the ith new coupon (after $(i-1)$st).

- Example: X_1 is always 1 because we always get a new coupon on the first purchase (we have no other coupons).

Expected Value

We define $X = X_1 + X_2 + \cdots + X_n$, where X_i is the number of cereal boxes we buy while trying to get the ith new coupon, *after* getting the $(i-1)$st.

- Example: X_1 is always 1 because we always get a new coupon on the first purchase (we have no other coupons).

Then suppose we buy:

- 2 cereal boxes while trying to get the 2^{nd} new coupon.
- 3 cereal boxes while trying to get the 3^{rd} new coupon.
- 1 cereal box while trying to get the 4^{th} new coupon (lucky!).
- 9 cereal boxes while trying to get the 5^{th} new coupon (unlucky!).

Then $X = 1 + 2 + 3 + 1 + 9 + \cdots$

Expected Value

We define $X = X_1 + X_2 + \cdots + X_n$, where X_i is the number of cereal boxes we buy while trying to get the ith new coupon, *after* getting the $(i-1)$st.

- Example: X_1 is always 1 because we always get a new coupon on the first purchase (we have no other coupons).

What is the distribution for X_2?

- X_2 is how many boxes we need to buy if probability $\frac{n-1}{n}$ of a new coupon.
- That is, each box we buy has probability $\frac{n-1}{n}$ of giving us the new coupon. Each box selection is a Bernoulli trial, with probability of success $\frac{n-1}{n}$.
- This is precisely the geometric distribution, i.e., $X_2 \sim \text{Geometric}\left(\frac{n-1}{n}\right)$

Expected Value

We define $X = X_1 + X_2 + \cdots + X_n$, where X_i is the number of cereal boxes we buy while trying to get the ith new coupon, *after* getting the $(i-1)$st.

- Example: X_1 is always 1 because we always get a new coupon on the first purchase (we have no other coupons).

What is $\mathrm{E}[X_i]$?

- $X_1 = 1$
- $X_2 \sim \mathrm{Geometric}\left(\frac{n-1}{n}\right) \Rightarrow E[X_2] = \dfrac{n}{n-1}$

What is the distribution for X_3? In other words, how many boxes do we expect to buy after second coupon to get the third coupon?

Expected Value

We define $X = X_1 + X_2 + \cdots + X_n$, where X_i is the number of cereal boxes we buy while trying to get the ith new coupon, *after* getting the $(i-1)$st.

- Example: X_1 is always 1 because we always get a new coupon on the first purchase (we have no other coupons).

What is $\mathrm{E}[X_i]$?

- $X_1 = 1$
- $X_2 \sim \text{Geometric}\left(\frac{n-1}{n}\right) \Rightarrow E[X_2] = \frac{n}{n-1}$
- $X_3 \sim \text{Geometric}\left(\frac{n-2}{n}\right) \Rightarrow E[X_3] = \frac{n}{n-2}$
- …

Expected Value

We define $X = X_1 + X_2 + \cdots + X_n$, where X_i is the number of cereal boxes we buy while trying to get the ith new coupon, *after* getting the $(i-1)$st.

$$\mathrm{E}[X] = E[X_1] + E[X_2] + E[X_3] + \cdots + E[X_n]$$

$$E[X_1] = 1 \qquad E[X_2] = \frac{n}{n-1} \qquad E[X_3] = \frac{n}{n-2} \qquad E[X_4] = \frac{n}{n-3} \qquad \cdots$$

$$\mathrm{E}[X] = \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{2} + \frac{n}{1}$$

$$= n\left(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \cdots + \frac{1}{2} + \frac{1}{1}\right)$$

Expected Value

We define $X = X_1 + X_2 + \cdots + X_n$, where X_i is the number of cereal boxes we buy while trying to get the ith new coupon, *after* getting the $(i-1)$st.

$$E[X] = E[X_1] + E[X_2] + E[X_3] + \cdots + E[X_n]$$

$$= n\underbrace{\left(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \cdots + \frac{1}{2} + \frac{1}{1}\right)}$$

This is the nth "Harmonic Number"

$$H_n = \sum_{i=1}^{n} \frac{1}{i} \approx \ln n + \gamma_E$$

where $\gamma_E \approx 0.5772 \ldots$

The Euler–Mascheroni constant

... Yes, that Euler. He got around...

$$E[X] = nH_n \approx n(\ln n + \gamma_E)$$

Expected Value

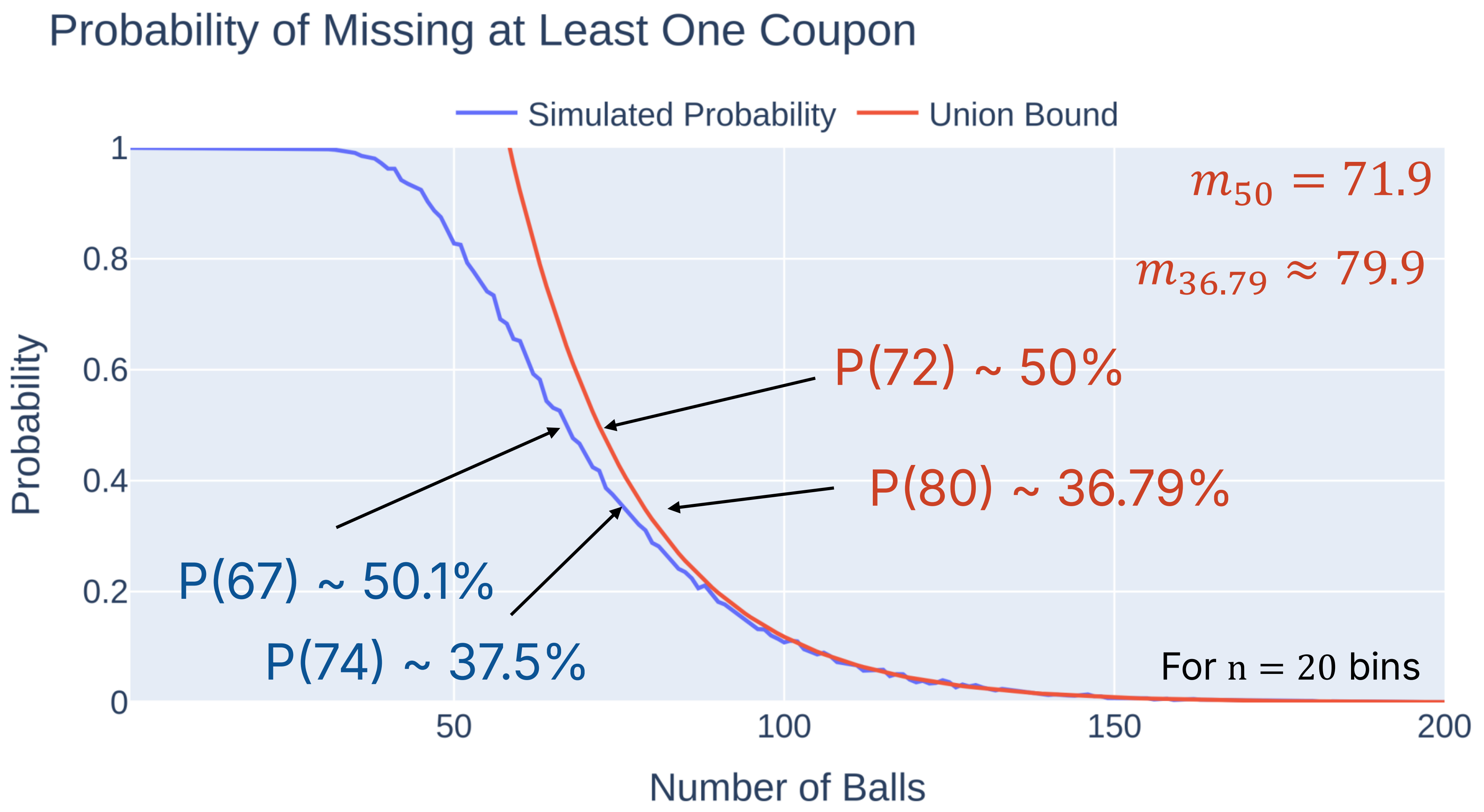
We define $X = X_1 + X_2 + \cdots + X_n$, where X_i is the number of cereal boxes we buy while trying to get the ith new coupon, *after* getting the $(i-1)$st.

$$\mathrm{E}[X] \approx n(\ln n + \gamma_E)$$

Example: $n = 20$ gives us $\approx 20(\ln 20 + 0.5772) \approx 71.5$

Union Bound Approximations and E[X] as Sum of Geometric RVs

Expected number of cereal boxes to get all coupons: $\mathrm{E}[X] \approx n(\ln n + \gamma_E) \approx 71.5$



Variance Derivation and Definition

Lecture 21, CS70 Summer 2025

Before We Get Going

There's a random math fact we'll need later, namely that

$$\left(\sum_{i=1}^{n} x_i\right)^2 = \sum_{i=1}^{n} x_i^2 + 2\sum_{i<j} x_i x_j$$

Example: $(500 + a + b + c)^2 = 500^2 + a^2 + b^2 + c^2$

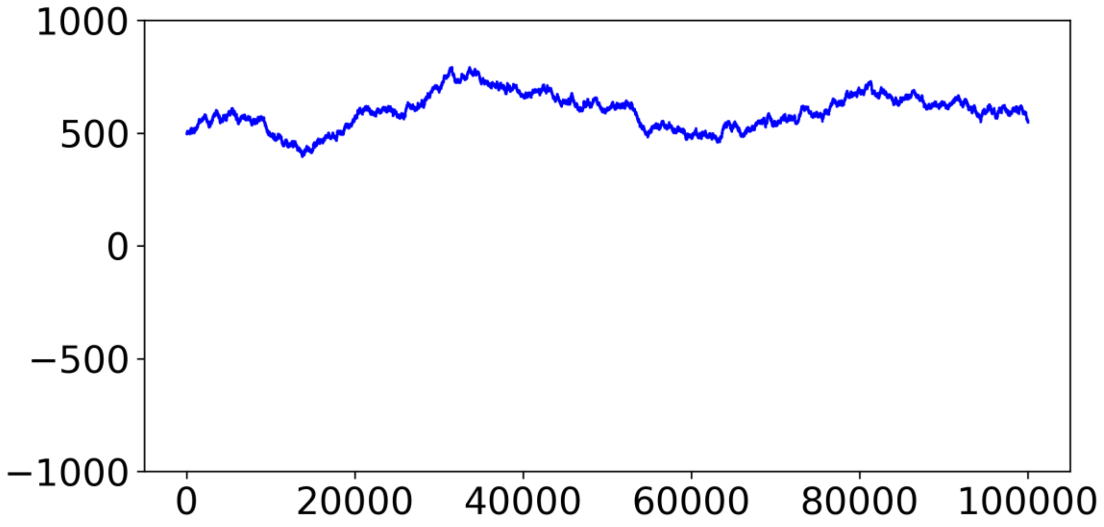
$$+1000a + 1000b + 1000c + 2ab + 2ac + 2bc$$

Motivation: Random Walks

Suppose we a have a particle that starts at $y = 500$. Every time step, it either:

- Moves up by 1 with probability ½.
- Moves down by 1 with probability ½.

For example, the plot below shows a random walk where the x-axis is time, and the y-axis is the position of the particle.



Motivation: Random Walks

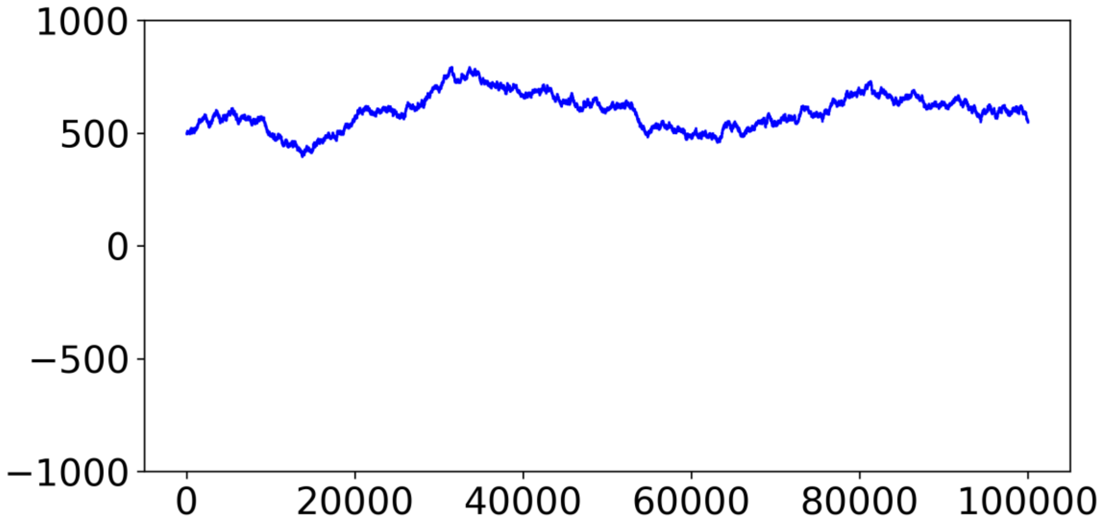
Suppose we a have a particle that starts at $y = 500$. Every time step, it either:

- Moves up by 1 with probability ½.
- Moves down by 1 with probability ½.

$$P[M_i = a] = \begin{cases} 1/2, & \text{if } a = +1 \\ 1/2, & \text{if } a = -1 \end{cases}$$

Let Y_n be the position of the particle after n moves.

- (Note: Y_n is called a "random process", though we won't use this term in 70)
- $Y_n = 500 + M_1 + M_2 + \cdots + M_n$



We can easily show that $E[Y_n] = 500$:

- $E[Y_n] = E[500 + M_1 + M_2 + \cdots + M_n]$

$$= 500 + nE[M_1]$$

$$= 500 + n \times 0 = 500$$

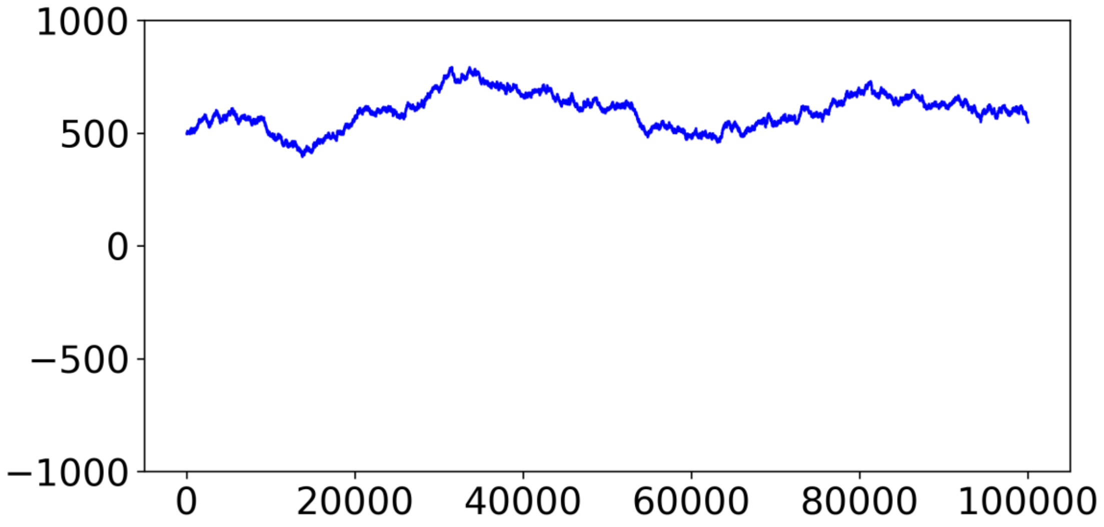
Motivation: Random Walks

Suppose we a have a particle that starts at $y = 500$. Every time step, it either:

- Moves up by 1 with probability ½.
- Moves down by 1 with probability ½.

$$P[M_i = a] = \begin{cases} 1/2, & \text{if } a = +1 \\ 1/2, & \text{if } a = -1 \end{cases}$$

Let Y_n be the position of the particle after n moves. $E[Y_n] = 500$.



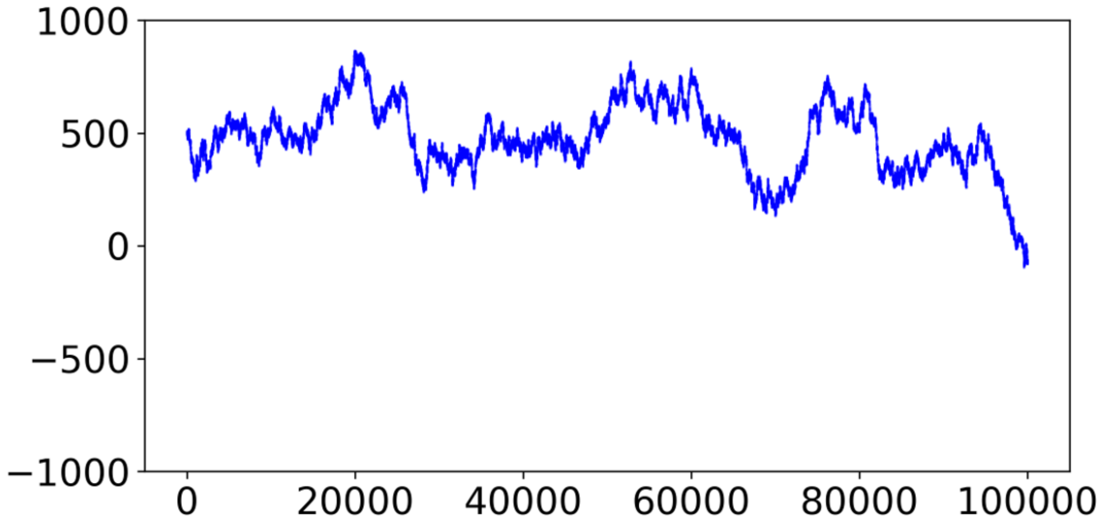
Motivation: Random Walks

Suppose we a have a particle that starts at $y = 500$. Every time step, it either:

- Moves up by 3 with probability ½.
- Moves down by 3 with probability ½.

$$P[M_i = a] = \begin{cases} 1/2, & \text{if } a = +3 \\ 1/2, & \text{if } a = -3 \end{cases}$$

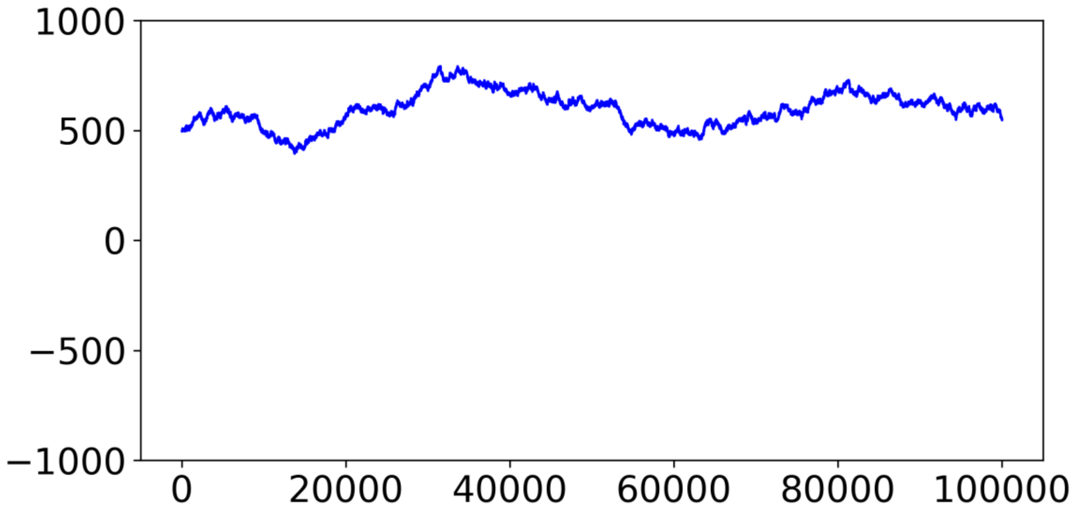
Let Y_n be the position of the particle after n moves. $E[Y_n] = 500$.

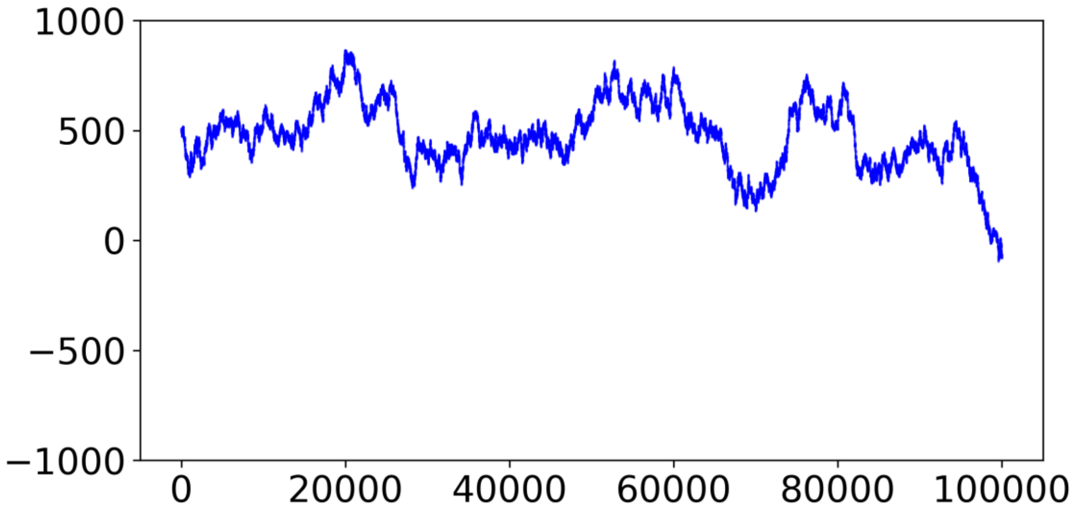


Motivation: Random Walks

Both of our random walks have an expected value of 500.

- ... but our random walk with a larger step size swings more widely.
- If we only use the expectation to summarize the random variable, we completely miss this important difference.

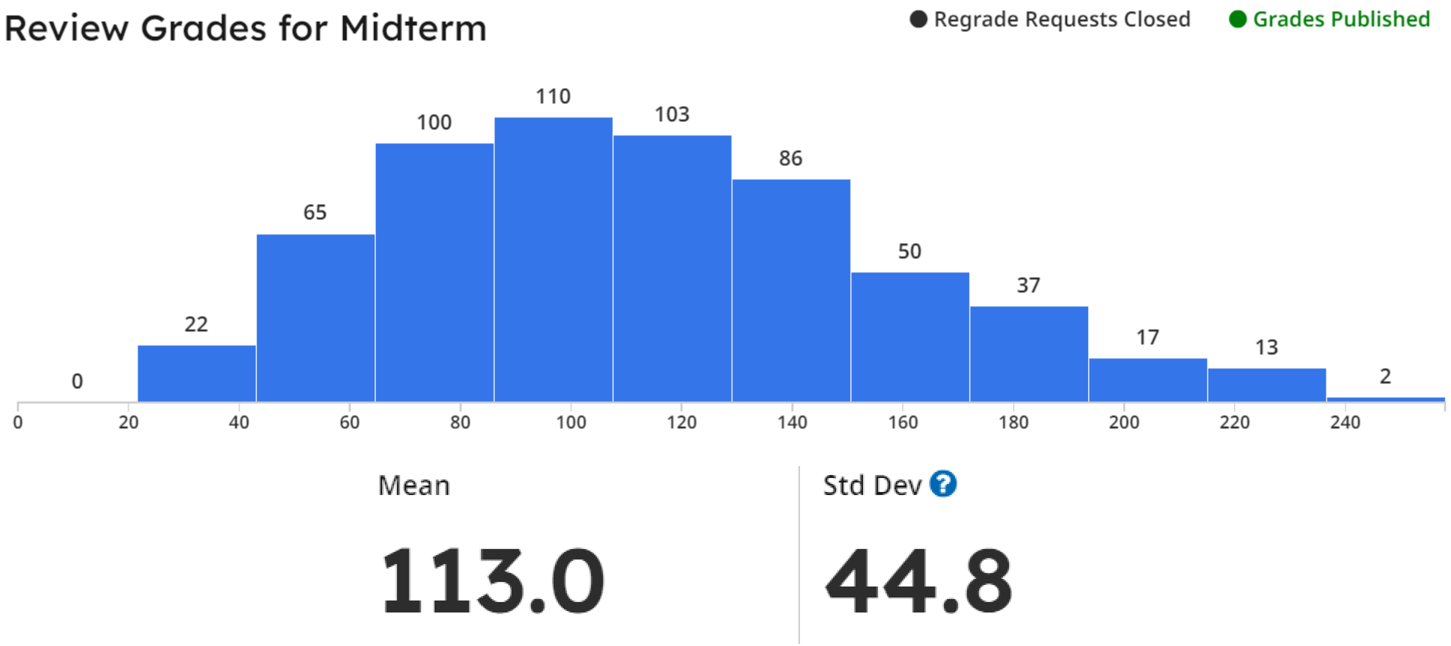




Measuring Dispersion

We'd like to measure the dispersion (or spread) of a distribution.

One summary statistic for dispersion is the standard deviation.



Warning: Easy to misinterpret this – a student taking an exam is not a r.v. or outcome

But selecting a random student's grade out of the class is

Probability theory vs statistics...

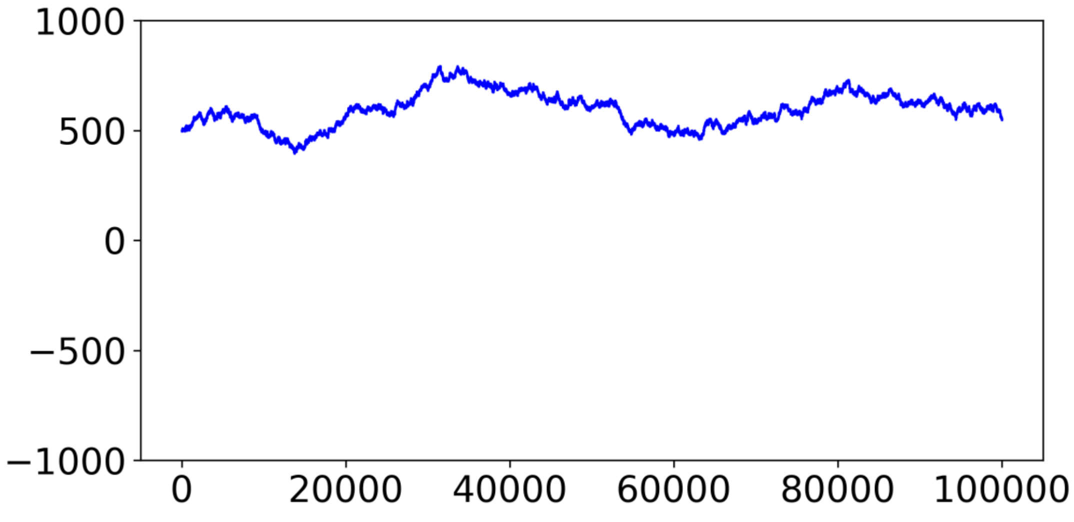
Attempt One: $E\left[Y_n - E[Y_n]\right]$

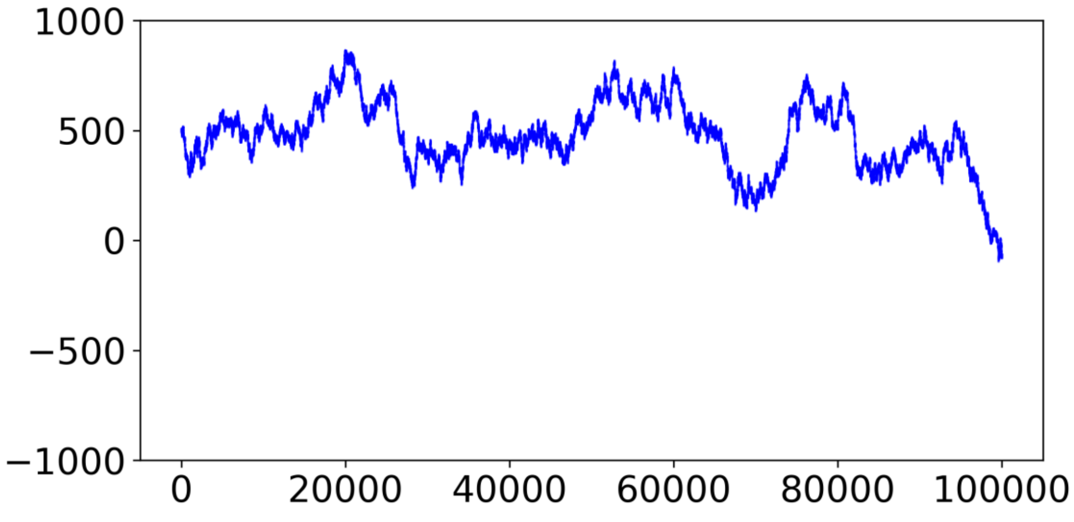
As a first attempt, imagine we attempt to define the "difference from the mean".

- Difference from the mean: $Y_n - E[Y_n]$

The difference from the mean is also random variable.

- A natural summary statistic would be the "expected difference from the mean" or $E\left[Y_n - E[Y_n]\right]$.
- Note: A summary statistic is a number which summarizes a distribution.





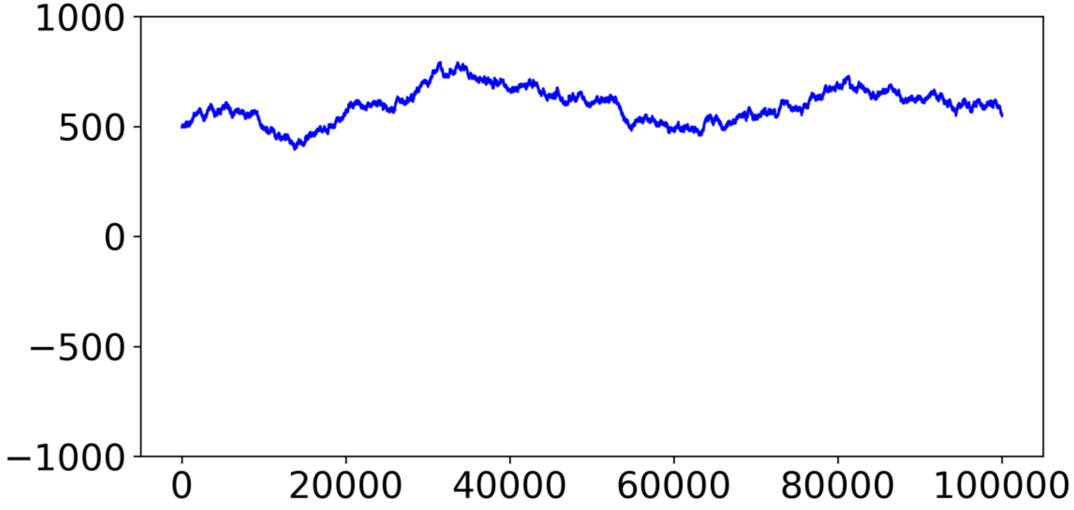
Attempt One: $E[Y_n - E[Y_n]]$

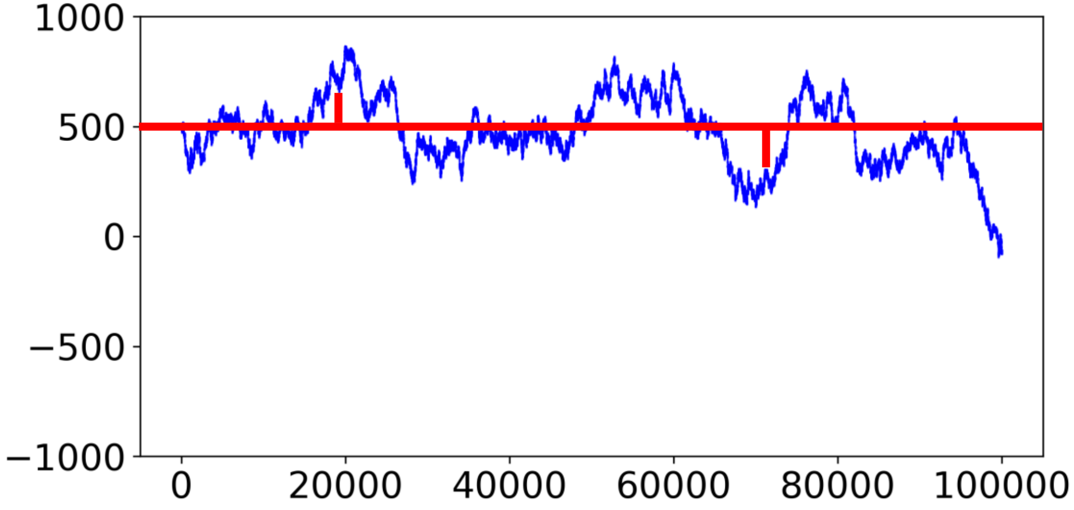
As a first attempt, imagine we attempt to define the "difference from the mean".

- Expected difference from the mean: $E[Y_n - E[Y_n]]$

Using linearity of expectation, we get:

- $E[Y_n - 500] = E[Y_n] - E[500] = E[Y_n] - 500 = 500 - 500 = 0.$
- Unfortunately, this quantity is on average zero.
- Positive and negative deviations cancel out.





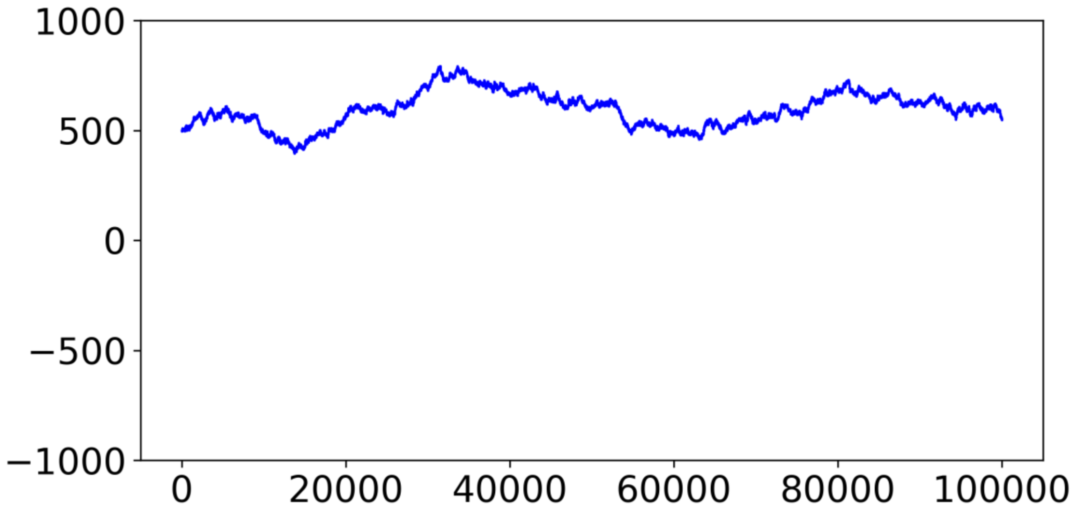
Attempt Two: $E[\,|Y_n - E[Y_n]|\,]$

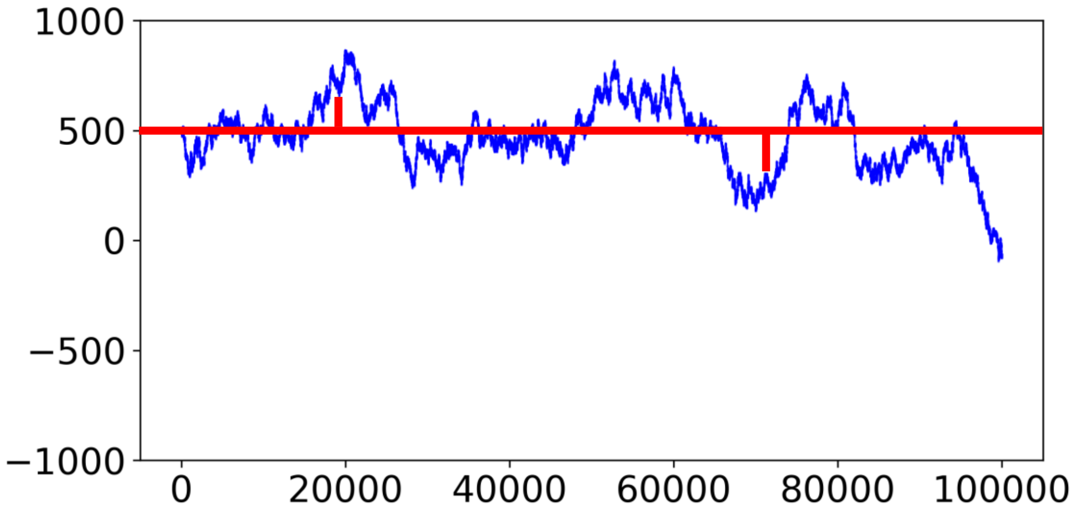
Our next idea: Compute the expected absolute difference from the mean.

- $E[\,|Y_n - E[Y_n]|\,] = E[\,|Y_n - 500|\,]$

Now we're a bit stuck. The absolute value stops us from using linearity of expectation to distribute the operator.

- We could start trying to break this into cases, e.g., when $Y_n \leq 500$ vs. $Y_n > 500$, but for many reasons we won't pursue this approach.





Attempt Three: $E[(Y_n - E[Y_n])^2]$

An alternate way to convert the distance into a positive value is to square it.

- That is, let's compute the expected squared difference between Y_n and $E[Y_n]$.

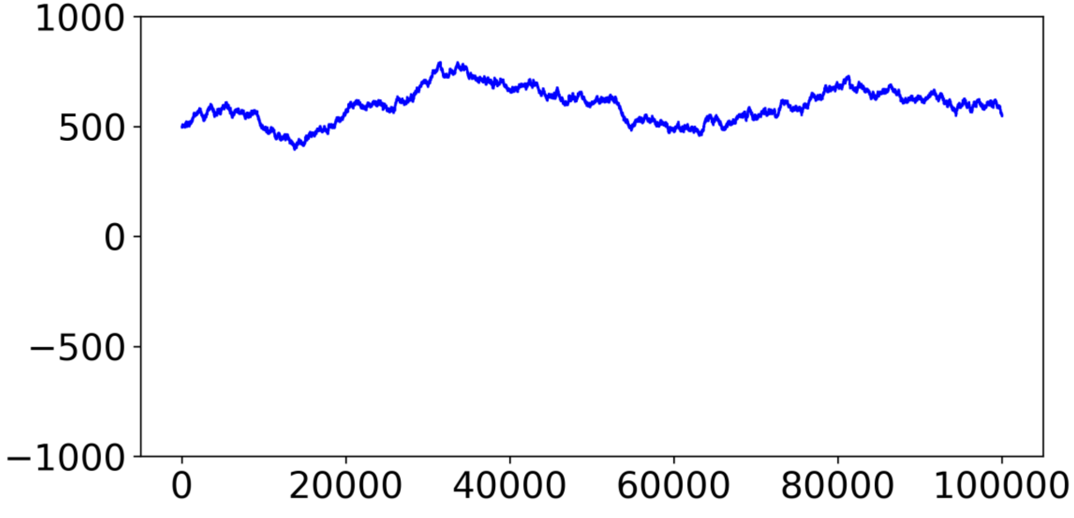
Using linearity of expectation, we can try to compute this quantity.

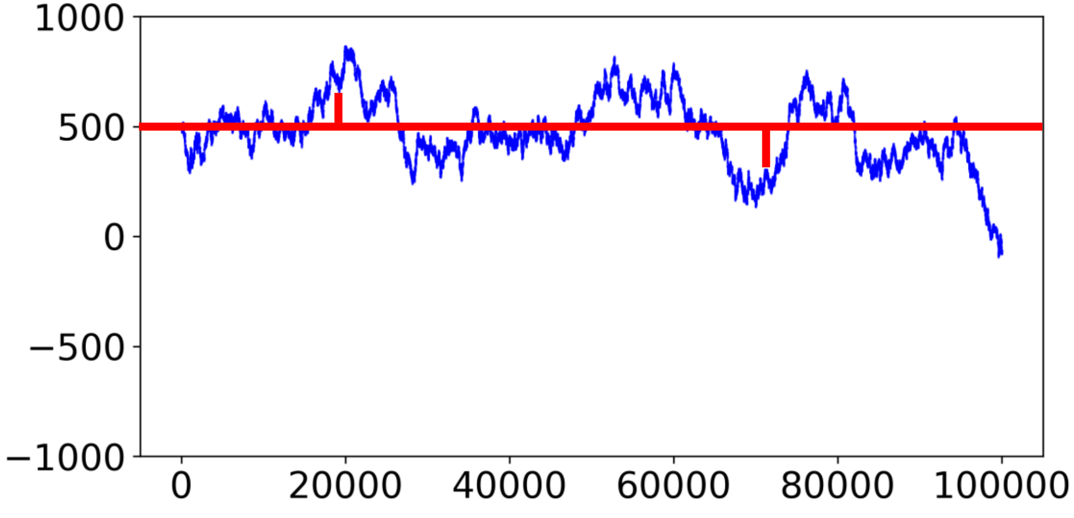
- $E[(Y_n - E[Y_n])^2] = E[(Y_n - 500)^2] = E[Y_n^2 - 1000Y_n + 250{,}000]$

$$= E[Y_n^2] - 1000E[Y_n] + 250{,}000$$

$$= E[Y_n^2] - 500{,}000 + 250{,}000 = E[Y_n^2] - 250{,}000$$

Need to compute this somehow.





Attempt Three: $E[(Y_n - E[Y_n])^2]$

Using linearity of expectation, compute expected squared difference from mean.

- $E[(Y_n - E[Y_n])^2] = E[Y_n^2] - 250{,}000$

To compute $E[Y_n^2]$, we can write $E[Y_n^2] = E[(500 + M_1 + M_2 + \cdots + M_n)^2]$

$$= E[500^2 + M_1^2 + M_2^2 + \cdots M_n^2$$

$$+1000M_1 + 1000M_2 + \cdots + 1000M_n$$

$$+2M_1M_2 + \cdots + 2M_{n-1}M_n]$$

Need to compute these

Math Interlude

What are $E[M_1^2]$, $E[M_1]$, and $E[M_1 M_2]$?

$$P[M_i = a] = \begin{cases} 1/2, & \text{if } a = \mathbf{+3} \\ 1/2, & \text{if } a = \mathbf{-3} \end{cases}$$

Math Interlude

What are $E[M_1^2]$, $E[M_1]$, and $E[M_1 M_2]$?

- M_1 is always 3 or -3, so M_1^2 is always 9. Thus, $E[M_1^2] = 9$.
- M_1 is -3 or 3 with probability ½, so $E[M_1] = -3 \times \frac{1}{2} + 3 \times \frac{1}{2} = 0$.
- $M_1 M_2$ can be:

$3 \times 3 = 9$ with probability ¼

$3 \times -3 = -9$ with probability ¼

$-3 \times 3 = -9$ with probability ¼

$3 \times 3 = 9$ with probability ¼

Thus, $E[M_1 M_2] = 9 \times \frac{1}{4} - 9 \times \frac{1}{4} - 9 \times \frac{1}{4} + 9 \times \frac{1}{4} = 0$.

$$P[M_i = a] = \begin{cases} 1/2, & \text{if } a = \mathbf{+3} \\ 1/2, & \text{if } a = \mathbf{-3} \end{cases}$$

Attempt Three: $E[(Y_n - E[Y_n])^2]$

Using linearity of expectation, we can try to compute this quantity.

- $E[(Y_n - E[Y_n])^2] = E[Y_n^2] - 250{,}000$

To compute $E[Y_n^2]$, we can write $E[Y_n^2] = E[(500 + M_1 + M_2 + \cdots + M_n)^2]$

$$= E[500^2 + M_1^2 + M_2^2 + \cdots M_n^2 \quad \leftarrow \quad E\left[M_i^2\right] = 9$$

$$+1000M_1 + 1000M_2 + \cdots + 1000M_n \quad \leftarrow \quad E[M_i] = 0$$

$$+2M_1M_2 + \cdots + 2M_{n-1}M_n] \quad \leftarrow \quad E\left[M_iM_j\right] = 0$$

$$= 250{,}000 + 9n$$

Attempt Three: $E[(Y_n - E[Y_n])^2]$

An alternate way to convert the distance into a positive value is to square it.

- That is, let's compute the expected squared difference between Y_n and $E[Y_n]$.

Using linearity of expectation, we can try to compute this quantity.

- $E[(Y_n - E[Y_n])^2] = E[(Y_n - 500)^2] = E[Y_n^2 - 1000Y_n + 250{,}000]$

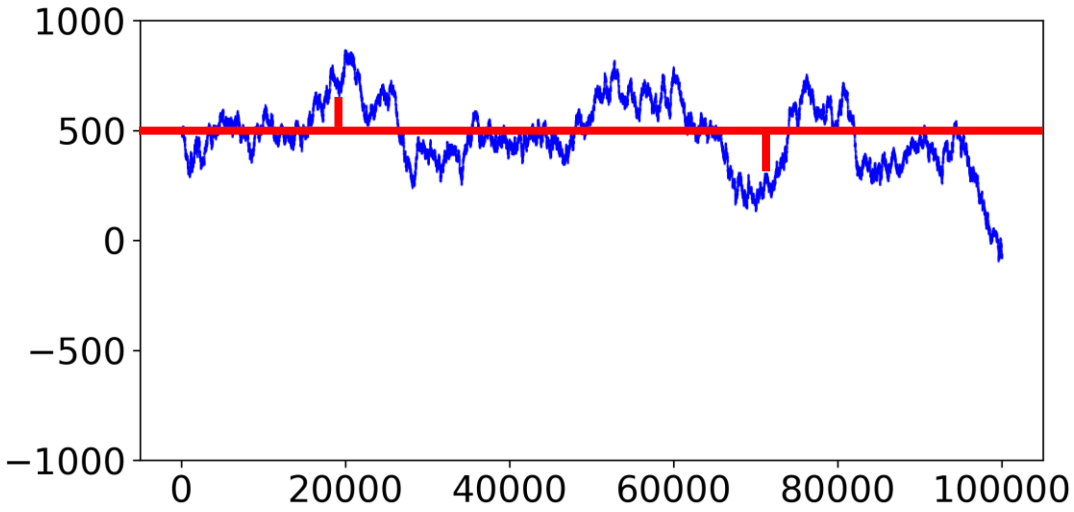
$$= E[Y_n^2] - 1000E[Y_n] + 250{,}000$$

$$= E[Y_n^2] - 500{,}000 + 250{,}000 = E[Y_n^2] - 250{,}000$$

Took some deeper thought. This is $250{,}000 + 9n$

$$= 250{,}000 + 9n - 250{,}000$$

$$= 9n$$



Attempt Three: $E[(Y_n - E[Y_n])^2]$

An alternate way to convert the distance into a positive value is to square it.

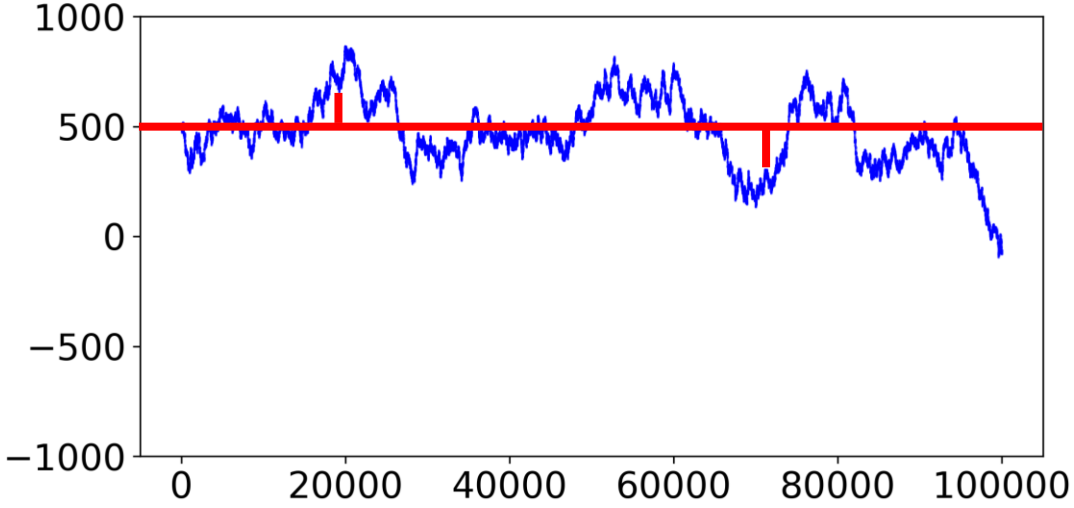
- That is, let's compute the expected squared difference between Y_n and $E[Y_n]$.

Using linearity of expectation, we can try to compute this quantity.

- $E[(Y_n - E[Y_n])^2] = 9n$

That is, on average, we can expect that by time step n, the squared distance of the particle from 500 will be around $9n$.

- We called this the "expected squared difference from the mean".
- This quantity is commonly known as the "variance" of the random variable.



Variance

The variance of a random variable X is defined as:

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

Note, we sometimes use $\mu = E[X]$ to avoid messy nested brackets, i.e.,

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

Properties of Variance

Lecture 21, CS70 Summer 2025

Variance Property 1

Variance is defined as $\text{Var}(X) = E[(X - E[X])^2]$

Useful Property 1: Can write variance as: $\text{Var}(X) = E[X^2] - E[X]^2$

Proof: Let $\mu = E[X]$, then $\text{Var}(X) = E[(X - \mu)^2]$

$$= E[X^2 - 2X\mu + \mu^2]$$

$$= E[X^2] - 2\mu E[X] + E[\mu^2]$$

$$= E[X^2] - 2\mu^2 + \mu^2$$

$$= E[X^2] - \mu^2$$

$$= E[X^2] - E[X]^2$$

Variance Property 2

Suppose we have a random variable X whose variance is $\mathrm{Var}(X)$. Then $\mathrm{Var}(cX) = c^2\mathrm{Var}(x)$.

Proof: $\mathrm{Var}(cX) = E[(cX)^2] - (E[cX])^2$ (by useful property 1)

$$= E[(cX)^2] - (cE[X])^2$$

$$= E[(cX)^2] - c^2E[X]^2$$

$$= E[c^2X^2] - c^2E[X]^2$$

$$= c^2E[X^2] - c^2E[X]^2$$

$$= c^2\mathrm{Var}(X)$$

Variance Property 3

Suppose we have a random variable X whose variance is $\mathrm{Var}(X)$. Then $\mathrm{Var}(c + X) = \mathrm{Var}(X)$.

Proof: $\mathrm{Var}(c + X) = E[(c + X - E[c + X])^2]$ (by basic definition)

$$= E[(c + X - E[c] - E[X])^2]$$

$$= E[(c + X - c - E[X])^2]$$

$$= E[(X - E[X])^2]$$

$$= \mathrm{Var}(X)$$

Recall basic definition: $\mathrm{Var}(X) = E[(X - E[X])^2]$

Variance and Standard Deviation

Variance is not in the same units as the original random variable.

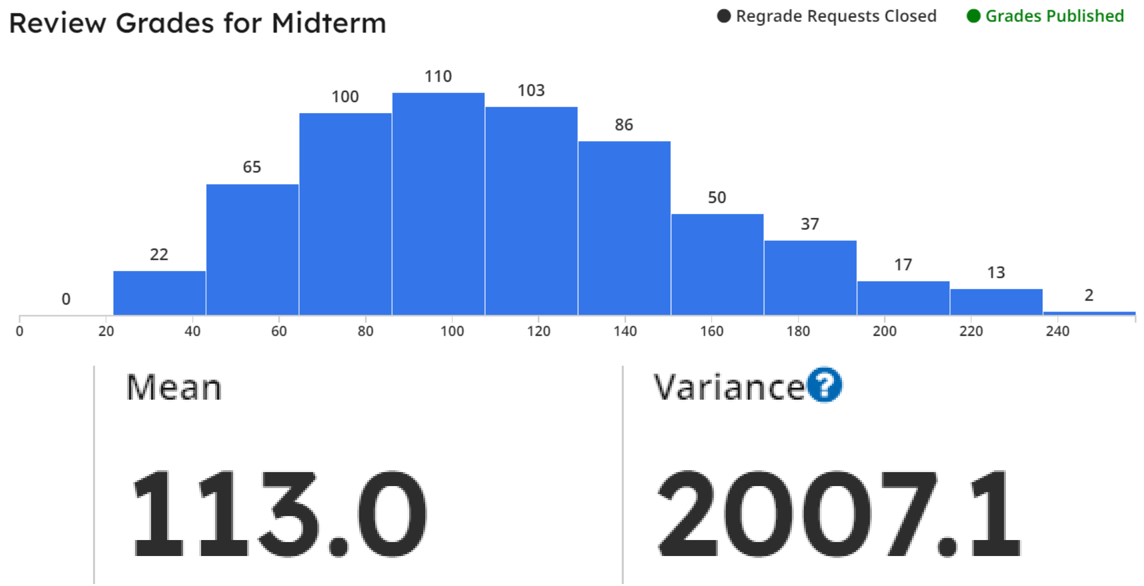
- Example: If X is a midterm score for a randomly selected student, then $\mathrm{Var}(X)$ is given in points squared.

The most natural correction is just to compute the square root of the variance. This is called the standard deviation (stdev for short).

- Variance of 2007.1 points squared: less intuitive than stdev of 44.8 points.

Often use σ as symbol for stdev

Warning again: data-driven (statistics) vs model-driven (probability theory) is different. Statistics has population variance, sample variance (biased and unbiased), ... not as “clean” as probability.



Useful Properties So Far

Basic definition of variance: $\mathrm{Var}(X) = E[(X - E[X])^2]$

Useful property 1: Can rewrite as $\mathrm{Var}(X) = E[X^2] - E[X]^2$

Useful property 2: $\mathrm{Var}(cX) = c^2\mathrm{Var}(X)$

Useful property 3: $\mathrm{Var}(c + X) = \mathrm{Var}(X)$

Units of variance are squared units of X. Can use square root of variance (standard deviation) instead.

Example Variance Calculations

Lecture 21, CS70 Summer 2025

Computing the Variance of a Random Variable

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

- The expectation $E[X]$.
- The expectation of X^2, i.e., $E[X^2]$

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

Example 1: Bernoulli RV

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

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

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

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.

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 $E[X] = 7/2$.

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

$$E[X^2] = \frac{1}{6} \times 1^2 + \frac{1}{6} \times 2^2 + \frac{1}{6} \times 3^2 + \frac{1}{6} \times 4^2 + \frac{1}{6} \times 5^2 + \frac{1}{6} \times 6^2$$

$$= \frac{1}{6} \times (1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2)$$

$$= \frac{91}{6}$$

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 $E[X] = 7/2$. We also just computed $E[X^2] = 91/6$

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

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

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

We'll use this result later today for multiple rolls.

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.

- Earlier we showed that $E[X_n] = 1$.
- 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 $E[X_n]$. We just need $E[X_n^2]$.

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

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

What is $E[I_i^2]$?

$$P(I_i^2 = 1) = P(I_i = 1) = \frac{1}{n}$$

$$P(I_i^2 = 0) = P(I_i = 0) = \frac{n-1}{n}$$

$$E[I_i^2] = \frac{1}{n} \times 1 = \frac{1}{n} \implies \sum_{i=1}^{n} E[I_i^2] = 1$$

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

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

What is $E[I_i I_j]$?

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

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

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

But only one of these terms is non-zero: $1 \times 1 \times P(I_i = 1, I_j = 1)$

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

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

I_i and I_j are indicators, so for the non-zero term:

$$E[I_i I_j] = 1 \times 1 \times P(I_i = 1, I_j = 1)$$
$$= P(i \text{ and } j \text{ get their own homework back})$$
$$= \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$$

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

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

I_i and I_j are indicators, so for the non-zero term :

$$E[I_i I_j] = 1 \times 1 \times P(I_i = 1, I_j = 1)$$

$$= P(i \text{ and } j \text{ get their own homework back})$$

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

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

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

What is $2\sum_{i<j} \frac{1}{n(n-1)}$? There are $\binom{n}{2}$ pairs $\{i, j\}$, each with one ordering $i < j$, so

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

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

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

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

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

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 $E[X_n] = 1$.
- How dispersed is this? Do we expect to sometimes get $X_n = 10$, $X_n = 20$?

To compute variance, we know $E[X_n] = 1$. We now know $E[X_n^2] = 2$.

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

$$\begin{aligned}\mathrm{Var}(X_n) &= E[X_n^2] - E[X_n]^2 \\ &= 2 - 1 \\ &= 1\end{aligned}$$

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 $E[X_n] = 1$.
- And we also know that $\mathrm{Var}(X_n) = 1$.

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 gets their homework back is 1.
- And the variance (and standard deviation) is 1.
- If we were to run a bunch of experiments, the average difference from our observation and the mean (1) would be 1.

Example(s) 4: Geometric and Poisson Random Variables

If $X \sim \text{Geometric}(p)$, then $\text{Var}(X)$ is given

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

If $X \sim \text{Poisson}(\lambda)$, then $\text{Var}(X) = \lambda$

The proofs are in the notes – study them there!

- The proof for geometric is especially worth going through. Uses a super-cool calculus trick...

Quick Note

The notes also discuss the uniform distribution.

$$X \sim \text{Uniform}(n)$$

Means that X takes on values from the set $\{1, 2, 3, \ldots, n\}$ with equal probability.

The notes give the mean and variance of the uniform distribution without proof. I won't state them here. This slide just exists to make sure you don't miss this in the notes (note 16, page 4).

Sums of Independent Random Variables

Lecture 21, CS70 Summer 2025

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 $S = X + Y$, then $S \sim \text{Poisson}(\lambda + \mu)$.

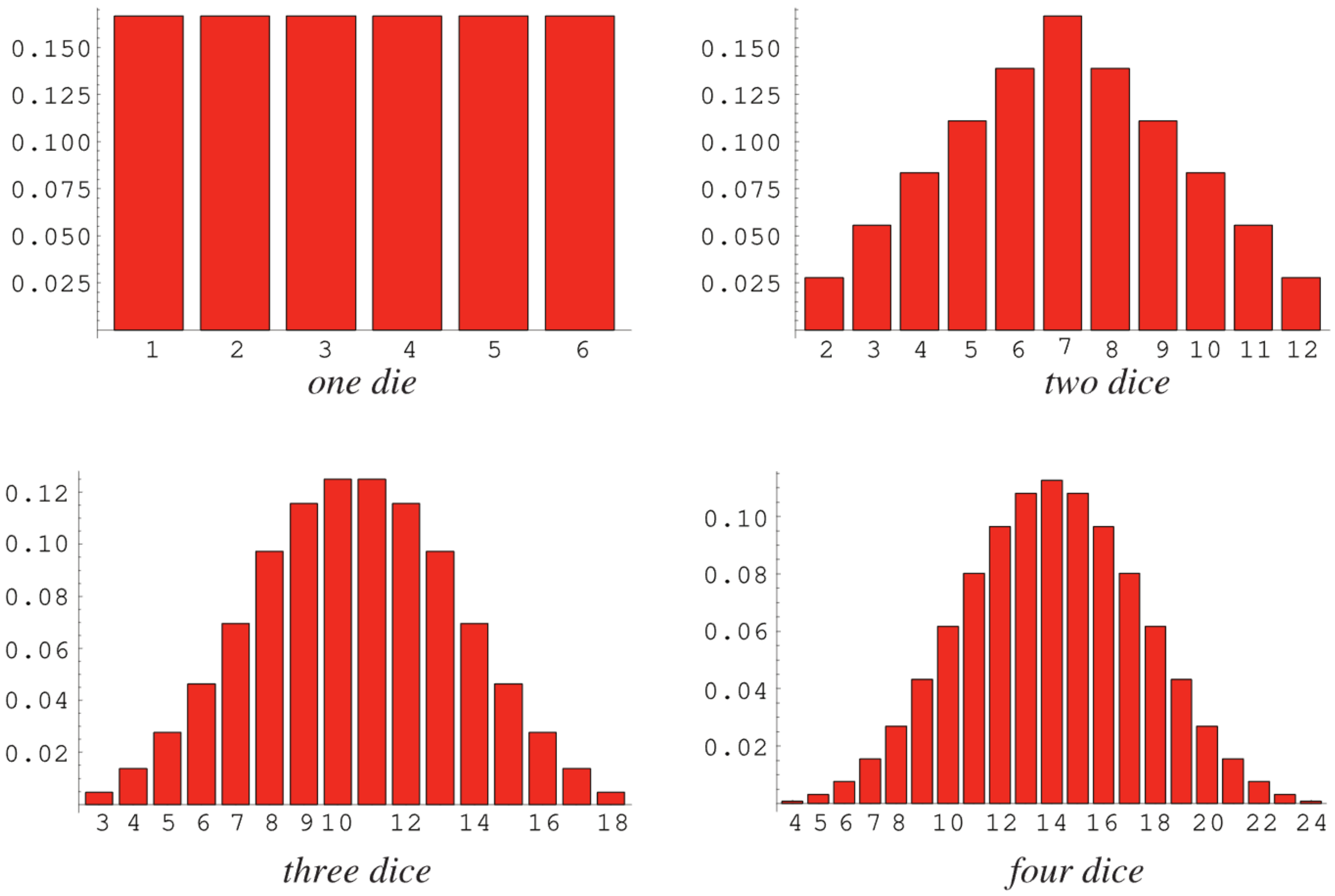
- What about the variance of the sum? $\text{Var}(S) = \lambda + \mu$

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.

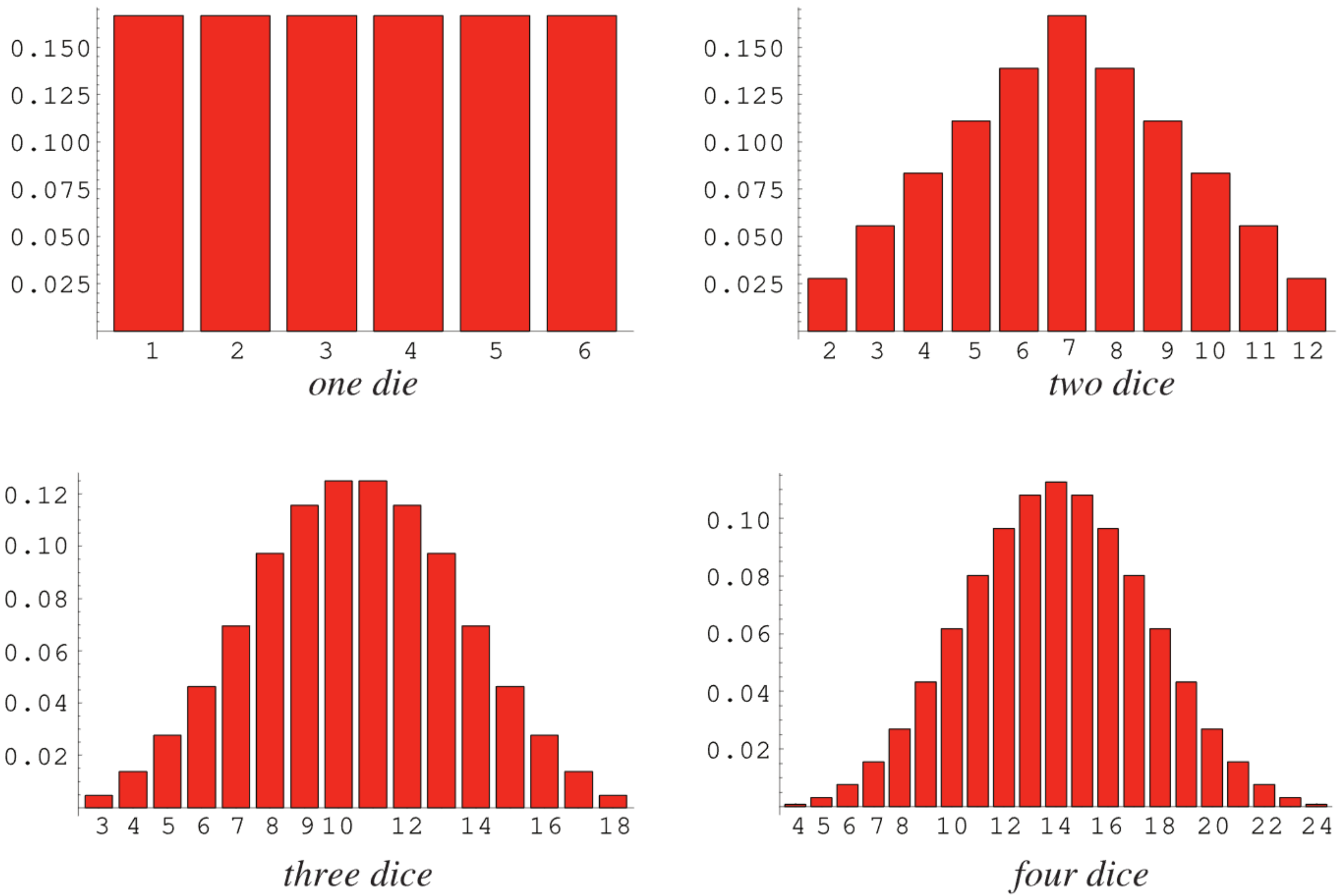
As an example, consider summing n rolls of fair six-sided dice (next slide).

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



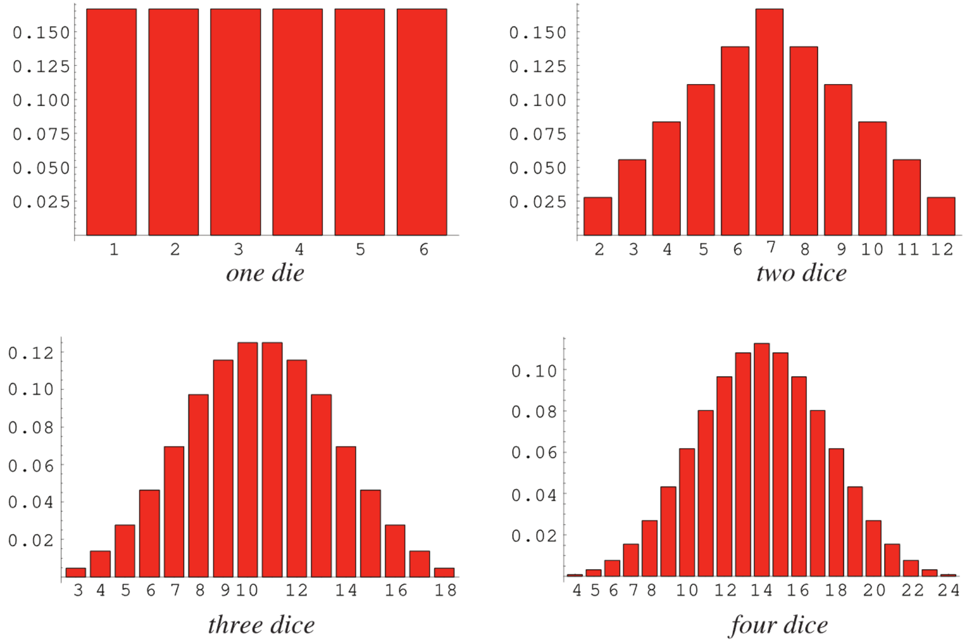
Note: The operation that is happening here is called “convolution”. We won’t discuss today.

Expected Value and Standard Deviation vs. N



n	$E[X]$	$\sigma(X)$
1	3.5	1.71
2	7	2.41
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	2.41
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 slow.
- In fact, $\sigma(n) \approx 1.7\sqrt{n}$, or equivalently $\mathrm{Var}(n) \approx 2.92n$

Expectation and Sums of Independent Random Variables

While we can't say concisely say much about 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 $E[X + Y] = E[X] + E[Y]$.

- 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.

Expectation and Sums of Independent Random Variables

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

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

$$E[(X+Y)^2] = E[X^2 + 2XY + Y^2] = E[X^2] + 2E[XY] + E[Y^2]$$

$$(E[X+Y])^2 = (E[X] + E[Y])^2 = E[X]^2 + 2E[X]E[Y] + E[Y]^2$$

With these two quantities expanded, we can write:

$$\begin{aligned}\mathrm{Var}(X+Y) &= E[(X+Y)^2] - (E[X+Y])^2 \\ &= E[X^2] - E[X]^2 + 2E[XY] - 2E[X]E[Y] + E[Y^2] - E[Y]^2 \\ &= \mathrm{Var}(X) + \underbrace{2E[XY] - 2E[X]E[Y]}_{\text{To deal with this, let's take a detour.}} + \mathrm{Var}(Y)\end{aligned}$$

Expectation and Sums of Independent Random Variables

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

Proof:

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

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

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

$$= E[X]\,E[Y]$$

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

$$2 \times 5 \times \mathrm{P}(\mathrm{D}_4 = 2, \mathrm{D}_6 = 5)$$

$$= 2 \times 5 \times \mathrm{P}(\mathrm{D}_4 = 2) \times (\mathrm{D}_6 = 5)$$

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

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

Expectation and Sums of Independent Random Variables

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

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

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

$$= E[X^2] - E[X]^2 + 2E[XY] - 2E[X]E[Y] + E[Y^2] - E[Y]^2$$

$$= \mathrm{Var}(X) + \underbrace{2E[XY] - 2E[X]E[Y]}_{} + \mathrm{Var}(Y)$$

We just showed $E[XY] = E[X]E[Y]$

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

Requires independence of X and Y

Summing i.i.d. Random Variables – Example 1

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:

- $E[S] = n\,E[X]$
- $\mathrm{Var}(S) = n\,\mathrm{Var}(X)$

Example: For rolling one six-sided die and n six-sided dice, we have:

$$E[X_i] = \frac{7}{2} \qquad E[S] = \frac{7}{2}n$$

$$\mathrm{Var}(X_i) = \frac{35}{12} \approx 2.92 \qquad \mathrm{Var}(S) = \frac{35}{12}n \approx 2.92n$$

$$\sigma(X_i) \approx 1.7 \qquad \sigma(S) \approx 1.7\sqrt{n}$$

Summing i.i.d. Random Variables – Example 2 (Bernoulli → Binomial)

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:

- $E[S] = n\, E[X_1]$
- $\mathrm{Var}(S) = n\, \mathrm{Var}(X_1)$

Example: For $X_1 \sim \mathrm{Bernoulli}(p)$:

$$E[X_1] = p$$

$$\mathrm{Var}(X_1) = p(1 - p)$$

For $S \sim \mathrm{Binomial}(n, p)$:

$$E[S] = np$$

$$\mathrm{Var}(S) = np(1 - p)$$

Covariance

Lecture 21, CS70 Summer 2025

Covariance

Earlier we showed that:

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

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

If X and Y are independent, then these black terms canceled out. But what if they are not independent?

- In that case, the variance of the sum will depend on this term. This term (excluding the factor of 2) is called the covariance.

Covariance Definition

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

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

We already know a couple of facts about the covariance:

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

Covariance Intuitively

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

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

- $E[R_1] = 1/2$
- $E[B_2] = 1/2$

Then we have:

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

Covariance Intuitively

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

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

- $E[R_1] = 1/2$
- $E[B_2] = 1/2$

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

$$P(R_1 = 1, B_2 = 1) = P(B_2 = 1 | R_1 = 1) \times P(R_1 = 1)$$
$$= 26/51 \times 26/52$$
$$= 13/51$$

Covariance Intuitively

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

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

- $E[R_1] = 1/2$
- $E[B_2] = 1/2$
- $E[R_1 B_2] = 13/51$

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

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

Covariance Intuitively (Example 2)

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

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

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

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

The negative covariance means that the random variables tend to move apart, e.g., if R_1 is 1, R_2 is less likely to be 1.

Covariance and Independence

Lecture 22, CS70 Summer 2025

Coupon Collector and Expectation

Variance

- Derivation and Definition
- Properties of Variance
- Example Variance Calculations
- Sums of Random Variables
- Covariance
- Covariance and Independence

Example 1

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?

Example 1

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.
- 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: $P(X = 1, Y = 1) \neq P(X = 1) \times P(Y = 1)$

Example 1

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

Example 1

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

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

Even though covariance is zero, variables are not independent!

Example 2

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

Z \ W	-1	0	1
1	1/7	1/7	0
0	1/7	1/7	1/7
-1	0	1/7	1/7

Questions:

- Are these variables independent?
- Is their covariance negative, zero, or positive?

Example 2

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

Z \ W	-1	0	1
1	1/7	1/7	0
0	1/7	1/7	1/7
-1	0	1/7	1/7

Questions:

- Are these variables independent?
    - No! If W = 1, Z can't be 1
    - $P(W = 1, Z = 1) \neq P(W = 1) \times P(Z = 1)$

Example 2

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

Z \ W	-1	0	1
1	1/7	1/7	0
0	1/7	1/7	1/7
-1	0	1/7	1/7

Questions:

- Is their covariance negative, zero, or positive?
  - Negative, tend to move in opposite directions.
  - Or in terms of the actual values: $\text{cov}(W,Z) = E[WZ] - E[W]\,E[Z]$

$$= -\frac{2}{7} - 0 \cdot 0$$

Summary

Key Discoveries of the Day

- For the coupon collector problem $X = X_1 + X_2 + \cdots + X_n$ where $X_i \sim \text{Geometric}$
  - $E[X] \approx n(\ln n + \gamma_E)$
- $\text{Var}(X) = E[(X - E[X])^2]$
- $\text{Var}(X) = E[X^2] - E[\text{X}]^2$
- $\sigma(X) = \sqrt{\text{Var}(X)}$
- $\text{Var}(X + Y) = \text{Var}(X) + 2E[XY] - 2E[X]E[Y] + \text{Var}(Y)$
- $\text{cov}(X, Y) = E[XY] - E[X]\,E[Y]$
- If X and Y independent:
  - $E[XY] = E[X]\,E[Y]$
  - $\text{cov}(X, Y) = 0$
  - $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$