3778394

slido

Join at slido.com
#3778394

Click **Present with Slido** or install our Chrome extension to display joining instructions for participants while presenting.

CC BY NC SA

3778394

LECTURE 17

Random Variables

Representations of random events and their properties; sampling variability.

Data 100/Data 200, Spring 2025 @ UC Berkeley

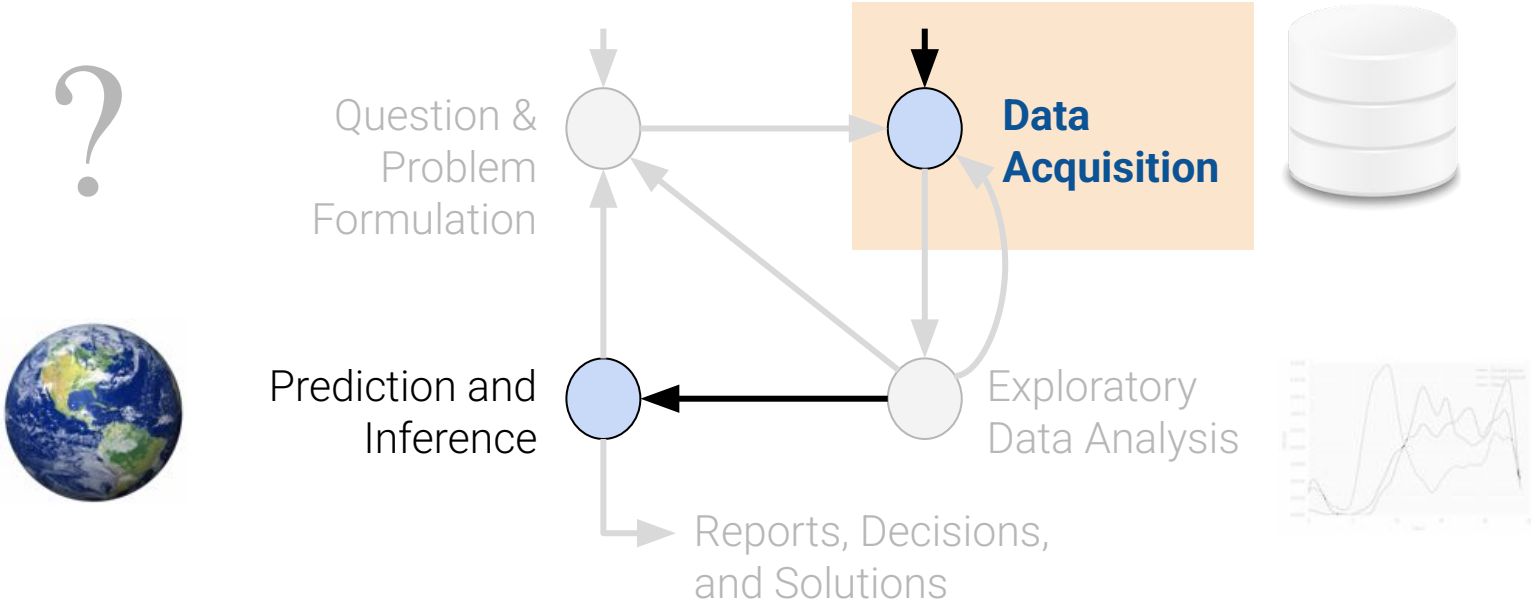
Narges Norouzi and Josh Grossman

Content credit: Acknowledgments

CC BY NC SA

Why Probability?

3778394



(today)

Model Selection Basics:
Cross Validation
Regularization

Probability I:
Random Variables
Estimators

Probability II:
Bias and Variance
Inference/Multicollinearity

CC BY NC SA

3778394

Today's Roadmap

Lecture 17, Data 100 Spring 2025

Random Variables and Distributions

Expectation and Variance

Sums of Random Variables

- Example: Bernoulli and Binomial Random Variables

Populations and Samples

CC BY NC SA

3778394

Random Variables and Distributions

Lecture 17, Data 100 Spring 2025

Random Variables and Distributions

Expectation and Variance

Sums of Random Variables

- Example: Bernoulli and Binomial Random Variables

Populations and Samples

CC BY NC SA

An Investing Conundrum (From a real interview!)

3778394

Imagine you are an analyst at a small-but-mighty venture capital firm.

You identify a hot new start-up and are considering an investment of **$1 million**.

Your analysis estimates a **90%** chance that the start-up **fails** and you lose your $1 million.

But, there's a **10%** chance the start-up **succeeds** and your investment grows to **$10 million**.

You can only invest the $1 million, or do nothing. Do you invest?

CC BY NC SA

3778394

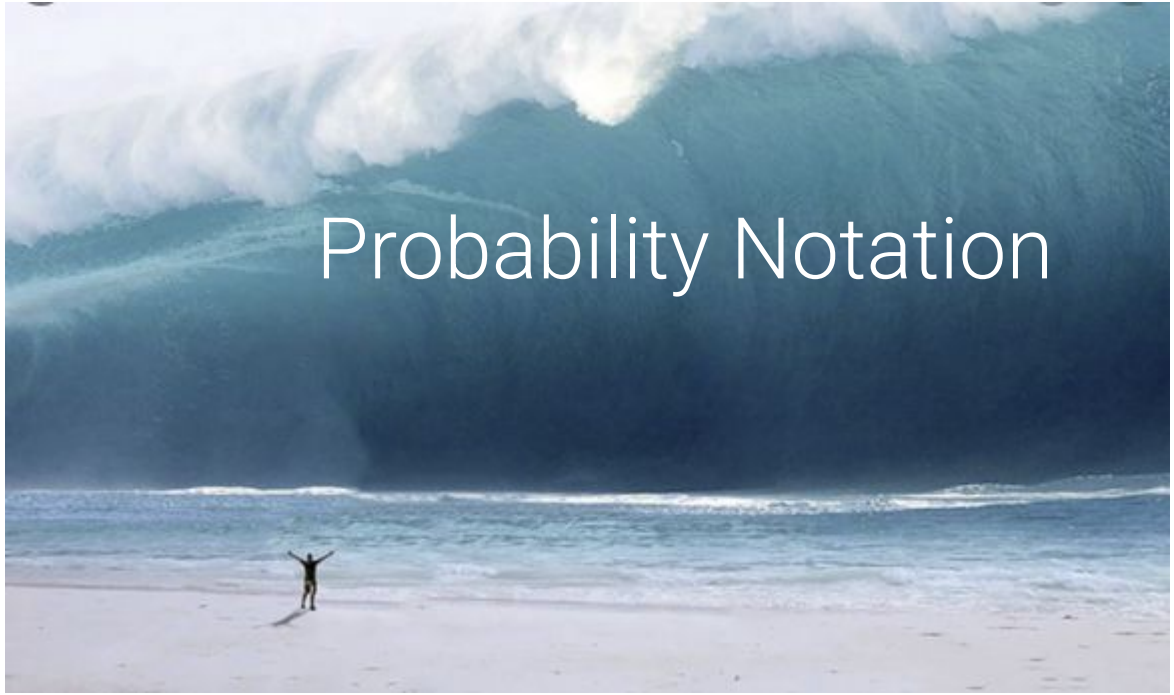
You are considering an investment of $1 million. You calculate a 90% chance you lose your $1 million and a 10% chance your investment grows to $10 million. Do you invest?

Presenting with animations, GIFs or speaker notes? Enable our Chrome extension

slido

Incoming Tsunami of Notation!

3778394



Be kind to yourself 💙

CC BY NC SA

[Terminology] Random Variable (RV)

3778394

A **random variable (RV)** represents the **outcome of a random event**.*

For example, a fair coin can land either heads (H) or tails (T).

Let the random variable (RV) X_i represent the outcome of the ith coin flip in a series of flips:

$$X_i = \begin{cases} 1, & \text{if coin lands heads} \\ 0, & \text{if coin lands tails} \end{cases}$$

Suppose we repeatedly flip a coin.

- The first flip is heads → $X_1 = 1$
- The second flip is tails → $X_2 = 0$
- The third flip is tails → $X_3 = 0$
- And so on.

*Technically there is a more precise definition but it is out of scope for this class.

CC BY NC SA

Coin flips are more interesting than you might think!

4

Abstract

We analyze the natural process of flipping a coin which is caught in the hand. We prove that vigorously-flipped coins are biased to come up the same way they started. The amount of bias depends on a single parameter, the angle between the normal to the coin and the angular momentum vector. Measurements of this parameter based on high-speed photography are reported. For natural flips, the chance of coming up as started is about .51.

[Diaconis et al., 2007]

CC BY NC SA

Coin flips are more interesting than you might think!

3778394

Abstract

Many people have flipped coins but few have stopped to ponder the statistical and physical intricacies of the process. In a preregistered study we collected 350,757 coin flips to test the counterintuitive prediction from a physics model of human coin tossing developed by Diaconis, Holmes, and Montgomery (D-H-M; 2007). The model asserts that when people flip an ordinary coin, it tends to land on the same side it started—D-H-M estimated the probability of a same-side outcome to be about 51%. Our data lend strong support to this precise prediction: the coins landed on the same side more often than not, $\Pr(\text{same side}) = 0.508$, 95% credible interval (CI) $[0.506, 0.509]$, $\text{BF}_{\text{same-side bias}} = 2364$.

[Bartos et al., 2023]

But, in Data 100, we will assume heads and tails are equally likely for a fair coin!

CC BY NC SA

[Terminology] Discrete and Continuous Random Variables

3778394

Random variables (RVs) are either **discrete** or **continuous**.

Discrete: **Finite** # of possible outcomes. For example, coin flips (heads or tails). $\mathbf{X_i \in \{0, 1\}}$

Continuous: **Infinite** # of possible outcomes. For example, let $\mathbf{Y_i}$ be an RV representing the number of seconds a coin is in the air during the ith coin flip.

First flip takes 2.134 seconds. $\mathbf{Y_1 = 2.134}$. Second flip takes 1.789 seconds. $\mathbf{Y_2 = 1.789}$

Y_i can (technically) be any real number greater than 0. $\mathbf{Y_i \in \mathbb{R}^+}$

CC BY NC SA

[Terminology] Probability and Random Variables (RVs)

3778394

The outcomes of RVs have corresponding **probabilities**.

Discrete RV: Each possible outcome has an associated probability.

For coin flips, **P(X=1) = 0.5** and **P(X=0) = 0.5**

Continuous RV: Individual outcomes have probability 0. No single event is "special"!

For example, if Y represents airtime of coin flips, **P(Y = 2.712324 seconds) = 0**

But, **ranges** of outcomes have probability ≥ 0.

For example, we could have a situation where **P(0 seconds < Y < 2 seconds) = 0.8**

CC BY NC SA

[Terminology] Probability Distribution of a Discrete Random Variable

3778394

The **probability distribution** of a discrete RV provides the probability associated with each possible outcome.

For coin flips:

Outcome	P(X=Outcome)
Heads (1)	0.5
Tails (0)	0.5

More compactly:

x	P(X=x)
1	0.5
0	0.5

We typically use **capital letters (X)** as a placeholder for the outcome of a random variable.

We typically use **lowercase letters (x)** to denote possible values that an RV can take on.

CC BY NC SA

Working with probability distributions of discrete RVs

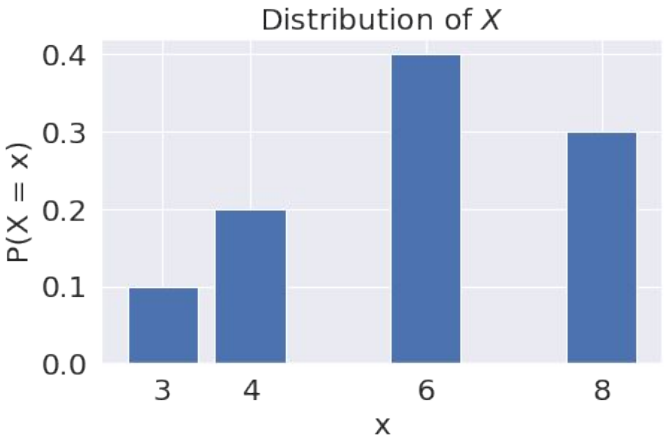
3778394

Compute the following probabilities for the discrete RV X

Notice that there is no i subscript! This problem is framed as a single event, not a series of events.

x	P(X = x)
3	0.1
4	0.2
6	0.4
8	0.3

1. $P(X = 4) =$
2. $P(X < 6) =$
3. $P(X = 6 \text{ or } X = 8) =$
4. $P(X \leq 8) =$



CC BY NC SA

Working with probability distributions of discrete RVs

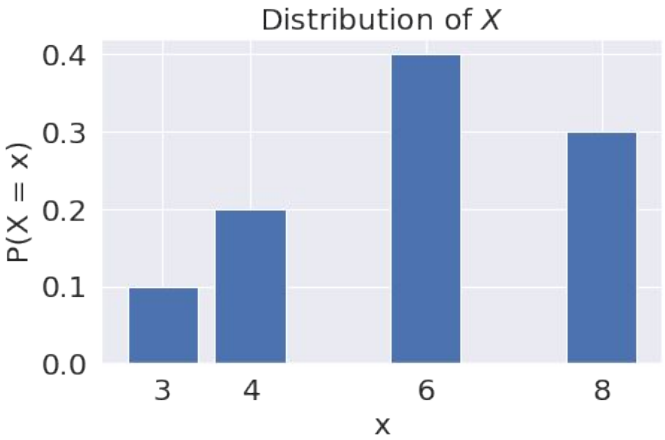
3778394

Compute the following probabilities for the discrete RV X

Notice that there is no i subscript! This problem is framed as a single event, not a series of events.

x	P(X = x)
3	0.1
4	0.2
6	0.4
8	0.3

1. $P(X = 4) =$ 0.2
2. $P(X < 6) =$ $0.1 + 0.2 = 0.3$
3. $P(X = 6 \text{ or } X = 8) =$ $0.4 + 0.3 = 0.7$
3. $P(X \leq 8) =$ 1



CC BY NC SA

[Terminology] Probability Distribution of a Discrete Random Variable

3778394

The probabilities associated with each outcome of a <u>discrete</u> RV **sum to 1**.

x	P(X=x)
1	0.5
0	0.5

P(X = 1) + P(X = 0) = 1

In other words,

$$\sum_{\substack{\text{all possible} \\ \text{values of x}}} P(X = x) = 1$$

More compactly,

$$\sum_{x} P(X = x) = 1$$

CC BY NC SA

[Terminology] Probability Distribution of a Continuous Random Variable

3778394

Probability tables don't work for continuous RVs!

Let Y be an RV that represents the seconds of airtime for a coin flip.

y	P(Y=y)
0.000000001	0
0.000000002	0
0.000000003	0
0.000000004	0
. . .	. . .

← This doesn't make any sense!

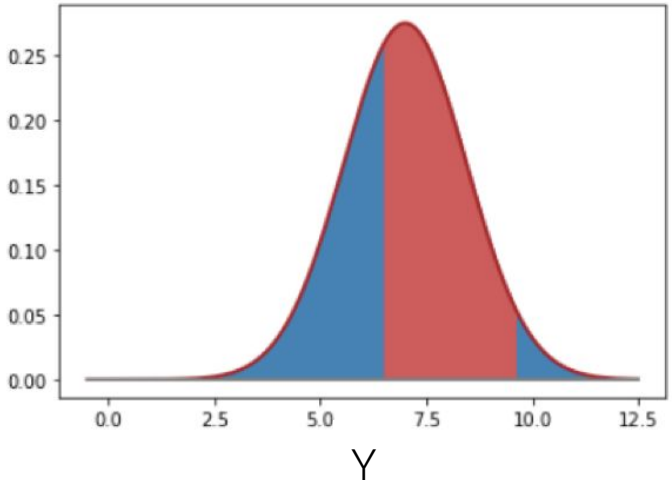
CC BY NC SA

For continuous random variables, probability comes from area

3778394

Distribution of **continuous** random variable Y

Density
(not probability!)



The **red area** under the curve is $P(6.8 \leq Y \leq 9.5)$

The **total area** under the **density** curve of the RV Y is 1.

Continuous RVs work a little differently than discrete RVs.

- Take a probability class (Data 140) to learn more about continuous distributions.

CC BY NC SA

Do not edit
How to change the design

Let Y be an RV representing your total cash after the start-up investment. What is the probability distribution of Y?

(Recall: It costs $1 million to invest, with a 10% chance of the investment growing to $10 million, and a 90% chance of failure.)

Presenting with animations, GIFs or speaker notes? Enable our Chrome extension

slido

$1 million investment. 10% chance it succeeds and turns into $10 million.

3778394

y	P(Y=y)
0	0.9
1	0.1

This is the distribution of the RV of **whether or not** the investment succeeds.

y	P(Y=y)
$0	0.9
$10 million	0.1

Remember, we lose $1 million if the investment fails.

Y is a continuous RV. We can't show it in a table.

In reality, investments are usually continuously distributed. But we've set up a world where there are only two possible outcomes.

CC BY NC SA

[Terminology] Probability Distribution

3778394

A **probability distribution** fully defines an RV. With the distribution of a RV, you can:

- Compute **properties** of the RV (e.g., mean, median, variance, …)
- **Simulate** random *draws* of the RV (e.g., flip 100 coins, what happens?)

Very often, we don't know the true distribution. We can only observe a sample drawn from it.

The **true** distribution of X:

x	P(X=x)
1	0.5
0	0.5

Suppose I flip 100 coins using Python. 57 are heads.

The **empirical** distribution of X, for this sample:

x	P(X=x)
1	0.57
0	0.43

CC BY NC SA

Named Distributions: **Bernoulli**(p)

3778394

Binary random variable that takes on **two values** 0 or 1.

$$X_i = \begin{cases} 1, & \text{for some outcome} \\ 0, & \text{any other outcome} \end{cases}$$

The Bernoulli distribution is **parameterized** by ***p*** **,** the probability that $P(X_i=1)$.

x	$P(X_i=x)$
1	p
0	1-p

Suppose an RV X_i follows a Bernoulli distribution with p=0.5

Then, we can write $\mathbf{X_i \sim Bernoulli(0.5)}$ or $\mathbf{X_i \sim Bern(0.5)}$

CC BY NC SA

Other Named Distributions and their Parameters

3778394

Bernoulli(p) → Discrete

- 1 with probability **p**, and 0 with probability 1 - p.

Binomial(n, p) → Discrete

- Probability of 0 heads, 1 heads, … or n heads out of **n** coin flips, where P(Heads) = **p**

Categorical(p_1, …, p_k) → Discrete

- Probability that an RV takes on the value of each possible category labeled 1,2,…,**k**

Uniform(a,b) → Continuous

- All values between **a** and **b** have the same *density*.

Normal(mean=$\boldsymbol{\mu}$, variance=$\boldsymbol{\sigma^2}$) → Continuous

CC BY NC SA

3778394

Expectation and Variance

Lecture 17, Data 100 Spring 2025

Random Variables and Distributions

Expectation and Variance

Sums of Random Variables

- Example: Bernoulli and Binomial Random Variables

Populations and Samples

CC BY NC SA

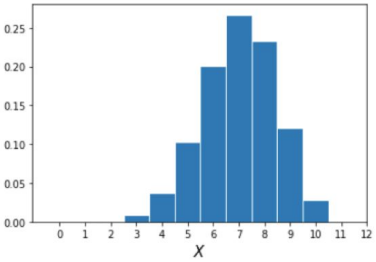
Communicating Information About Random Variables

3778394

List of sampled values
$X_1 = 1, X_2 = 0, \ldots$

Probability distribution
P(X=1) = p
P(X=0) = 1-p

Histogram of sampled values



Descriptions

Properties (i.e., summaries)

Random Variable X

Expectation
the "average value" of X
$E[X]$

Variance
the "spread" of X
$\mathrm{Var}(X)$

Other properties:
Skewness
Kurtosis
And more!
(Beyond Data 100)

CC BY NC SA

Population and Sample Averages

3778394

Let $X_i \sim$ Bernoulli(0.5). In other words, we're flipping a fair coin.

If I were to flip <u>many</u> coins, I **expect** the average of those 0/1 flips to be **0.5**.

If I flip 100 coins and 57 are heads, the average of this **sample** of 0/1 flips is **0.57**.

0.5 is the **expected value** (i.e., true average) of the RV X_i .

- The expected value is a fixed property of X_i (i.e., a property of the **coin flipping process**)

0.57 is the **sample** average, or **empirical** average, of $X_1, \ldots, X_{100}$

- The sample average can change from sample to sample. The expected value does not.

CC BY NC SA

3778394

Suppose you made 100 of these investments. On average, how much should you expect to earn per investment?

Presenting with animations, GIFs or speaker notes? Enable our Chrome extension

slido

Expected Value of a Categorical Distribution

3778394

x	Pr(X=x)
- $1 million	0.9
$10 million	0.1

CC BY NC SA

Expected Value of a Categorical Distribution

3778394

x	Pr(X=x)
- $1 million	0.9
$10 million	0.1

Over 100 investments, we would expect to have, on average:

-$1 million for 90 of the investments

$10 million for 10 of the investments

100 investment total
= (- $1 million)(90) + ($10 million)(10)
= - $90 million + $100 million = $10 million

Average over 100 investments = $10 million / 100 = **$100,000**

CC BY NC SA

Expected Value of a Categorical Distribution

3778394

x	Pr(X=x)
- $1 million	0.9
$10 million	0.1

Our investment earns $100,000 in **expectation** (i.e., on average).

Our computation:

[(- $1 million)**(90)** + ($10 million)**(10)**] / **100** = $100,000

Equivalently:

[(- $1 million)**(0.9)** + ($10 million)**(0.1)**] = $100,000

The expected value (i.e., the expectation) of any discrete RV X:

$$\mathbb{E}(X) = \sum_{x} x \cdot P(X = x)$$

CC BY NC SA

3778394

Suppose P(X=1) = 0.5. What is E(X)?

Presenting with animations, GIFs or speaker notes? Enable our Chrome extension

slido

Expected value of a Bernoulli RV

3778394

$$X = \begin{cases} 1, & \text{if coin lands heads} \\ 0, & \text{if coin lands tails} \end{cases}$$

Suppose P(X=1) = 0.5. What is E(X)?

$$\mathbb{E}(X) = \sum_x x \cdot P(X = x)$$

x	P(X=x)
1	0.5
0	0.5

CC BY NC SA

Expected value of a Bernoulli RV

3778394

$$X = \begin{cases} 1, & \text{if coin lands heads} \\ 0, & \text{if coin lands tails} \end{cases}$$

Suppose P(X=1) = 0.5. What is E(X)?

$$\mathbb{E}(X) = \sum_x x \cdot P(X = x)$$

$$= 1 \cdot 0.5 + 0 \cdot 0.5$$

$$= 0.5$$

x	P(X=x)
1	0.5
0	0.5

The expectation of X **does not** need to be a possible value of X.

It's just an average.

CC BY NC SA

3778394

Suppose X ~ Bernoulli(p). What is E(X)?

Presenting with animations, GIFs or speaker notes? Enable our Chrome extension

slido

Expected value of a Bernoulli RV

3778394

Suppose X ~ Bernoulli(p). What is E(X)?

$$\mathbb{E}(X) = \sum_x x \cdot P(X = x)$$

x	P(X=x)
1	p
0	1-p

CC BY NC SA

Expected value of a Bernoulli RV

3778394

Suppose X ~ Bernoulli(p). What is E(X)?

$$\mathbb{E}(X) = \sum_x x \cdot P(X = x)$$

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

E(X) can be computed as a function of the parameters.

x	P(X=x)
1	p
0	1-p

CC BY NC SA

Spread of a Random Variable

3778394

Suppose an RV X has a mean of μ . In other words, E(X) = μ .

- μ is a fixed property of X (i.e., a constant), not an RV

How much do we expect a random draw of X to **differ** from μ ?

<u>Proposal 1</u>: E(X - μ) is the expected difference between X and the value μ.

x	P(X=x)
1	0.5
-1	0.5

E(X) = μ = 0. 50% of the time, X - μ will be 1, else -1. So, E(X - μ) = 0

x	P(X=x)
100	0.5
-100	0.5

E(X) = μ = 0. 50% of the time, X - μ will be 100, else -100. So, E(X - μ) = 0

As it turns out, **E(X - μ) is always 0 for any distribution***. Not very helpful!

*As long as E(X) is not undefined. See Cauchy distribution.

CC BY NC SA

Spread of a Random Variable

3778394

Suppose an RV X has a mean of μ . In other words, E(X) = μ .

- μ is a fixed property of X (i.e., a constant), not an RV

How much do we expect a random draw of X to **differ** from μ ?

<u>Proposal 2</u>: **E(| X - μ |)** is the expected **absolute** difference between X and μ.

This is better! But absolute values have nasty math.

<u>Final proposal</u>: **E((X - μ)^2)** is the expected **squared** difference between X and μ.

Since **E(X) = μ** , the expression above can be rewritten as **E[(X - E(X))^2].**

You can draw parallels to MSE and the squared residuals of OLS!

CC BY NC SA

Variance of a Random Variable

3778394

Variance is the expected **squared** deviation from the mean of X (i.e, E(X)).

$$\mathbb{E}(X) = \sum_x x \cdot P(X = x)$$

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

Sometimes written like this:

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

The units of variance are the **square** of the units of X.

To get back to the natural scale, take the square root to get the **standard deviation** of X:

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

CC BY NC SA

Variance of a Categorical Distribution

3778394

Variance is the expected **squared** deviation from the mean of X (i.e, E(X)).

$$\mathbb{E}(X) = \sum_x x \cdot P(X = x)$$

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

What is the variance of the investment? Recall that **E(X) = $100,000**.

x	**Pr(X=x)**	**x - E(X)**	**(x - E(X))^2**
- $1 million	0.9		
$10 million	0.1		

CC BY NC SA

Variance of a Categorical Distribution

3778394

Variance is the expected **squared** deviation from the mean of X (i.e, E(X)).

$$\mathbb{E}(X) = \sum_x x \cdot P(X = x)$$

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

What is the variance of the investment? Recall that **E(X) = $100,000**.

x	Pr(X=x)	x - E(X)	(x - E(X))^2
- $1 million	0.9	- $1.1 million	$1.21 trillion
$10 million	0.1	$9.9 million	$98 trillion

Var(X) = E[(x - E(X))^2] =

(1.21 trillion)(0.9) +

(98 trillion)(0.1)

≈ 10.89 trillion dollars^2

SD(X)^2 = Var(X). So, SD(X) ≈ $3.3 million.

Typically, our **actual** outcome deviates from our **average** outcome by about $3.3 million.

CC BY NC SA

Variance: Alternate Calculation

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

3778394

Variance is more conveniently calculated with this formula (proof in appendix):

$$\mathbb{E}(X) = \sum_x x \cdot P(X = x)$$

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

What is the variance of the investment? Recall that **E(X) = $100,000**.

x	**Pr(X=x)**	**x^2**
- $1 million	0.9	
$10 million	0.1	

CC BY NC SA

Variance: Alternate Calculation

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

3778394

Variance is more conveniently calculated with this formula (proof in appendix):

$$\mathbb{E}(X) = \sum_x x \cdot P(X = x)$$

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

What is the variance of the investment? Recall that **E(X) = $100,000**.

x	**Pr(X=x)**	**x^2**
- $1 million	0.9	$1 trillion
$10 million	0.1	$100 trillion

$E(X^2)$ =	$Var(X) = E(X^2)-E(X)^2$
(1 trillion)(0.9) +	10.9 trillion -
(100 trillion)(0.1)	$100{,}000^2$
= 10.9 trillion dollars^2	≈ 10.89 trillion dollars^2

CC BY NC SA

Properties of Bernoulli Random Variables

$$\mathbb{E}[X] = \sum_x xP(X = x)$$
$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2\end{aligned}$$

3778394

Let $X \sim$ **Bernoulli**(p).

Expectation:

x	P(X=x)
1	p
0	1-p

Variance:

CC BY NC SA

Properties of Bernoulli Random Variables

$$\mathbb{E}[X] = \sum_x xP(X = x)$$
$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$
$$= \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

3778394

Let $X \sim$ **Bernoulli**(p).

Expectation:

$$\mathbb{E}[X] = 1 \cdot p + 0 \cdot (1 - p) = p$$

We will get an average value of p across many, many samples

x	P(X=x)
1	p
0	1-p

Variance:

$$\mathbb{E}[X^2] = 1^2 \cdot p + 0 \cdot (1 - p) = p$$
$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$
$$= p - p^2 = p(1 - p)$$

Lower Var: p = 0.1 or 0.9
Higher Var: p close to 0.5

More info: google("plot x(1 - x)")

CC BY NC SA

Lots of Named Distributions on Wikipedia

3778394

WIKIPEDIA
The Free Encyclopedia

Search Wikipedia

Bernoulli distribution

Mean	p
Median	$\begin{cases} 0 & \text{if } p < 1/2 \\ [0,1] & \text{if } p = 1/2 \\ 1 & \text{if } p > 1/2 \end{cases}$
Mode	$\begin{cases} 0 & \text{if } p < 1/2 \\ 0, 1 & \text{if } p = 1/2 \\ 1 & \text{if } p > 1/2 \end{cases}$
Variance	$p(1-p) = pq$

wikipedia.org/wiki/Bernoulli_distribution

- Bernoulli distribution, for the ou
- Binomial distribution, for the nu number of independent occurre
- Negative binomial distribution, failures before a given number
- Geometric distribution, for bino before the first success; a spec

Related to sampling schemes over

- Hypergeometric distribution, fo number of total occurrences, us
- Beta-binomial distribution, for tl number of total occurrences, sa replacement)

Poisson distribution, for

Exponential distribution,

Gamma distribution, for

Categorical distribution, for a single categorical o

Bernoulli distribution

Multinomial distribution, for the number of each ty generalization of the binomial distribution

Multivariate hypergeometric distribution, similar to generalization of the hypergeometric distribution

wikipedia.org/wiki/Probability_distribution

Scientists use lots of different distributions to model phenomena in the real world.

Returning to the Investing Conundrum

3778394

Your analysis estimates a **90%** chance that the start-up **fails** and you lose your $1 million.

But, there's a **10%** chance the start-up **succeeds** and your investment grows to **$10 million**.

The investment has an expected value of $100,000. **Do you invest?**

Answer: It depends on your "runway".

If you only have $1 million, there's a 90% chance you end up bankrupt. Variance is your enemy.

If you have $1 billion in the bank and you can make many investments like this, you'll almost certainly profit in the long run.

This is why venture capitalists tend to be successful even though start-ups have a high failure rate. Venture capital can "take the hit" of high variance investments with positive expected value.

CC BY NC SA

3778394

Sums of Random Variables

Lecture 17, Data 100 Spring 2025

Random Variables and Distributions

Expectation and Variance

Sums of Random Variables

- Example: Bernoulli and Binomial Random Variables

Populations and Samples

CC BY NC SA

Functions of Multiple Random Variables

3778394

A function of an RV is also an RV!

For instance, if $X_1, X_2, \ldots, X_n$ are RVs, then these are RVs too:

$$X_n^2 \qquad \frac{1}{n}\sum_{i=1}^{n}(X_i - c)^2$$

$$\max(X_1, X_2, \ldots, X_n)$$

$$\boxed{\frac{1}{n}\sum_{i=1}^{n} X_i}$$

Counts+means involve **sums of RVs**, so good to learn how sums behave!

E(X) and Var(X) are not RVs → They are constant properties of an RV X

CC BY NC SA

A new mystery: Symmetric dice?

3778394

Let X be an RV denoting the number facing up on the roll of a die.

What is E(X)? All sides are equally likely.

(1+2+3+4+5+6)/6 = **3.5**

Now, roll two dice.

What is $\mathbf{E(2X_1)}$?

2 ✖

What is $\mathbf{E(X_1 + X_2)}$?

✚

CC BY NC SA

Properties of Expectation

3778394

Recall definition of expectation:

$$E[X] = \sum_{\text{all possible } x} x P(X = x)$$

Properties:

1. **Expectation is linear**.
Intuition: Summations are linear.
Out-of-scope proof in appendix.

$$E[aX + b] = aE[X] + b$$

2. **Expectation of sum = Sum of expectations.**
Intuition: Summations are linear.
Out-of-scope proof in appendix.

$$E[X + Y] = E[X] + E[Y]$$

3. **g(E(X)) ≠ E(g(X)). For example, $E(X)^2 \neq E(X^2)$.**
Out-of-scope proof in appendix.

$$\mathbb{E}[X]^2 \neq \mathbb{E}[X^2]$$

CC BY NC SA

3778394

Roll two dice. For one roll X, recall that E(X)=3.5.

What is E(2*X_1)? What is E(X_1 + X_2)?

Presenting with animations, GIFs or speaker notes? Enable our Chrome extension

slido

Symmetric dice?

3778394

Roll two dice. For one roll X, recall that E(X)=3.5

$$E[aX+b] = aE[X]+b$$

$$E[X+Y] = E[X]+E[Y]$$

$E(2X_1)$ =

$E(X_1 + X_2)$ =

CC BY NC SA

Symmetric dice?

3778394

Roll two dice. For one roll X, recall that E(X)=3.5

$$E[aX+b] = aE[X]+b$$

$$E[X+Y] = E[X]+E[Y]$$

$E(2X_1) = 2E(X_1) = 2*3.5 = 7$

$E(X_1 + X_2) = E(X_1) + E(X_2) = 3.5 + 3.5 = 7$

CC BY NC SA

Variance of our symmetric dice

3778394

Let's return to our dice mystery.

We saw that $E(2X_1) = E(X_1 + X_2) = 7$.

How much should we expect $2X_1$ and X_1+X_2 to vary around their **identical** means?

What is $Var(2X_1)$?

What is $Var(X_1 + X_2)$

CC BY NC SA

3778394

How much should we expect 2*X_1 and X_1+X_2 to vary around their identical means?

Presenting with animations, GIFs or speaker notes? Enable our Chrome extension

slido

Properties of Variance [1/2]

3778394

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

Properties:

1. **Variance is non-linear**:

Shifting (b) doesn't change variance, scaling (a) does.

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

$$\mathrm{SD}(aX + b) = |a|\mathrm{SD}(X)$$

CC BY NC SA

Properties of Variance [1/2]

3778394

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

Properties:

1. **Variance is non-linear**:

Shifting (b) doesn't change variance, scaling (a) does.

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

$$\mathrm{SD}(aX + b) = |a|\mathrm{SD}(X)$$

Intuition for a: Think of unit conversions.
X is an RV for height in meters.
Y is the same as X, but in centimeters. So, Y=100X.
Suppose SD(X) = 1.
Then, SD(Y) = 100.
100 SD(X) = SD(Y) = SD(100X)

Var(Y) = [100 SD(X)]^2 = 100^2 SD(X)^2 = 100^2 Var(X)

Intuition for b: If I add a constant to each element of a list, the elements remain spread out the same amount. So, SD and variance are unaffected by adding a constant.

CC BY NC SA

Identically Distributed vs. i.i.d.

3778394

Suppose that we have two random variables X and Y.

X and Y are **identically distributed** if the distribution of X is the same as the distribution of Y.

X and Y are **independent and identically distributed (i.i.d.)** if:

- X and Y are identically distributed, *and*
- Knowing the outcome of X does not influence your belief of the outcome of Y, and vice versa ("X and Y are **independent**.")

In Data 100, we will always tell you when RVs are i.i.d.

i.i.d. RVs

CC BY NC SA

Properties of Variance [2/2]

3778394

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

Properties:

1. **Variance is non-linear**:

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

$$\mathrm{SD}(aX + b) = |a|\mathrm{SD}(X)$$

2. Variance of sums of RVs is affected by the **independence** of the RVs:

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \underbrace{2\mathrm{Cov}(X, Y)}_{\text{Covariance of X and Y (next slide).}}$$

CC BY NC SA

Covariance and Correlation: The Basics

3778394

Covariance is the expected product of deviations from expectation.

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

Think of covariance as unstandardized **correlation** (i.e., correlation on a raw scale, not [-1,1])

$$r(X,Y) = \mathbb{E}\left[\underbrace{\left(\frac{X - \mathbb{E}[X]}{\mathrm{SD}(X)}\right)}_{X \text{ in } \textbf{standard units}}\left(\frac{Y - \mathbb{E}[Y]}{\mathrm{SD}(Y)}\right)\right] = \frac{\boxed{\mathrm{Cov}(X,Y)}}{\mathrm{SD}(X)\mathrm{SD}(Y)}$$

If r=0 for X and Y, then Cov(X,Y) = 0.

If knowing X tells you nothing about Y, then X and Y are **independent**, and Cov(X,Y) = 0.

We want you to generally understand what covariance is. Don't worry about computing it by hand.

r=0 does not guarantee independence.

CC BY NC SA

Properties of Variance [2/2]

3778394

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

Properties:

1. **Variance is non-linear**: $$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

2. Variance of sums of RVs is affected by the **independence** of the RVs:

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$$

Intuition: Suppose I have two similar investments X and Y that are positively **correlated**: If X is successful, Y is very likely successful, and vice-versa. X+Y will likely be really big or really small → Correlation makes the summed outcome more likely extreme. This "extremeness" is covariance.

Covariance can be negative, just like correlation. Imagine the opposite scenario: If one investment is successful, the other one is likely unsuccessful, and vice versa. In this case, X+Y will often be close to 0, since the investments often cancel each other out. This pull towards 0 is represented by a negative covariance term.

CC BY NC SA

Expectation of $2X_1$ vs $X_1 + X_2$

3778394

Let X_1 and X_2 be numbers on two dice rolls.

X_1, X_2 are i.i.d., so $E[X_1] = E[X_2]$ and $\mathrm{Var}(X_1) = \mathrm{Var}(X_2)$

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$$

$\mathrm{Var}(2X_1) =$

$\mathrm{Var}(X_1+X_2) =$

CC BY NC SA

Expectation of $2X_1$ vs $X_1 + X_2$

3778394

Let X_1 and X_2 be numbers on two dice rolls.

X_1, X_2 are i.i.d., so $E[X_1] = E[X_2]$ and $\mathrm{Var}(X_1) = \mathrm{Var}(X_2)$

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$$

$\mathrm{Var}(2X_1) = 2^2\mathrm{Var}(X_1) =$ **$4\mathrm{Var}(X_1)$**

$$\mathrm{Var}(X_1+X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \underbrace{\cancel{2\mathrm{Cov}(X_1, X_2)}}_{X_1,\, X_2 \text{ independent}}^{0} = \boldsymbol{2\mathrm{Var}(X_1)}$$

Alternate Solution:
$\mathrm{Var}(X_1 + X_1) = \mathrm{Var}(X_1) + \mathrm{Var}(X_1) + 2\mathrm{Cov}(X_1,X_1) = 4\mathrm{Var}(X_1)$.
Note that $\mathrm{Cov}(X,X) = E[(X - \mu)(X - \mu)] = E[(X - \mu)^2] = \mathrm{Var}(X)$

$\mathrm{Var}(X_1)=\mathrm{Var}(X_2)=35/12$. See appendix.

CC BY NC SA

Distributions of Sums

3778394

Let X_1 and X_2 be numbers on rolls of two dice.

- X_1, X_2 are **i.i.d.**, so X_1, X_2 have the same distribution.
- But $2X_1$ and $X_1 + X_2$ have different distributions!

Let's show this through simulation.

Demo

CC BY NC SA

Variance of $2X_1$ vs $X_1 + X_2$

3778394

How does this dice example relate to investments?

Often, you're advised to **diversify** investments to reduce **risk**. Don't put all $ in one investment.

In other words, don't gamble everything on the first dice roll. Spread your $ over the rolls.

Diversification reduces the **variance** of your portfolio of investments.

This is also why we pay for insurance. We pay premiums (i.e., reduce our expected value) in exchange for help in a catastrophe (i.e., greatly reduced variance).

CC BY NC SA

[Summary] Expectation and Variance for Linear Functions of Random Variables

3778394

Let X be a random variable with distribution $P(X = x)$.

$$\mathbb{E}[X] = \sum_x x P(X = x)$$

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \quad \text{(definition)}$$

$$= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \quad \text{(easier computation)}$$

Let a and b be scalar values.

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

Let Y be another random variable.

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

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \underbrace{2\mathrm{Cov}(X, Y)}_{\text{Zero if X, Y independent.}}$$

CC BY NC SA

3778394

Lecture 17 Ended Here!

BY NC SA

3778394

Bernoulli and Binomial Random Variables

Lecture 17, Data 100 Spring 2025

Random Variables and Distributions

Expectation and Variance

Sums of Random Variables

- **Example: Bernoulli and Binomial Random Variables**

Populations and Samples

CC BY NC SA

Continuing our investment story

3778394

Suppose we make **two** investments, each with a **50%** chance of success. (Coin flips)

We could denote the success of each investment as an RV X_i, where $\mathbf{X_i \sim Bernoulli(0.5)}$.

Assume the success of each investment is independently determined.

Let Y be the **total** number of successful investments. What's the probability distribution of Y?

Four equally likely possibilities:

1. Both fail. Y = 0
2. First one succeeds, second one fails. Y = 1
3. First one fails, second one succeeds. Y = 1
4. Both succeed. Y = 2

y	P(Y=y)
0	1 * 1/4
1	2 * 1/4
2	1 * 1/4

CC BY NC SA

Continuing our investment story

3778394

Suppose we make **three** investments, each with a 50% chance of success.

Let Y be the total number of successful investments. What's the probability distribution of Y?

$2^3 = 8$ equally likely possibilities:

1. All fail. Y = 0
2. First succeeds, others fail. Y = 1
3. Second succeeds, others fail. Y = 1
4. Third succeeds, others fail. Y = 1
5. First fails, others succeed. Y = 2
6. Second fails, others succeed. Y = 2
7. Third fails, others succeed. Y = 2
8. All succeed. Y = 3

y	P(Y=y)
0	1 * 1/8
1	3 * 1/8
2	3 * 1/8
3	1 * 1/8

CC BY NC SA

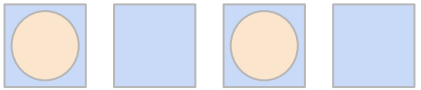
Pascal's Triangle and the Binomial Coefficient

3778394

y	P(Y=y)
0	1 * 1/4
1	2 * 1/4
2	1 * 1/4

y	P(Y=y)
0	1 * 1/8
1	3 * 1/8
2	3 * 1/8
3	1 * 1/8

$$1$$
$$1 \quad 1$$
$$1 \quad 2 \quad 1$$
$$1 \quad 3 \quad 3 \quad 1$$
$$1 \quad 4 \quad 6 \quad 4 \quad 1$$
$$1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1$$
$$1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1$$
$$1 \quad 7 \quad 21 \quad 35 \quad 35 \quad 21 \quad 7 \quad 1$$



Row **4**, Index **2**: "How many ways can I arrange **2** balls in **4** boxes?"

$$= \binom{4}{2} = 6$$

$$\binom{\text{row } \#}{\text{index}} = \text{element}$$

$$\binom{n}{y} = \frac{n!}{y!(n-y)!}$$

CC BY NC SA

Continuing our investment story

3778394

We've discovered the binomial distribution**.** If **Y ~ Binomial(n,p)**, then:

$$\mathbf{P}(Y = y) = \binom{n}{y} p^y (1-p)^{n-y} \qquad \binom{n}{y} = \frac{n!}{y!(n-y)!}$$

In **n Bernoulli trials**, each with **p** probability of success, what's the probability of **y** successes?

Intuition for the formula above:

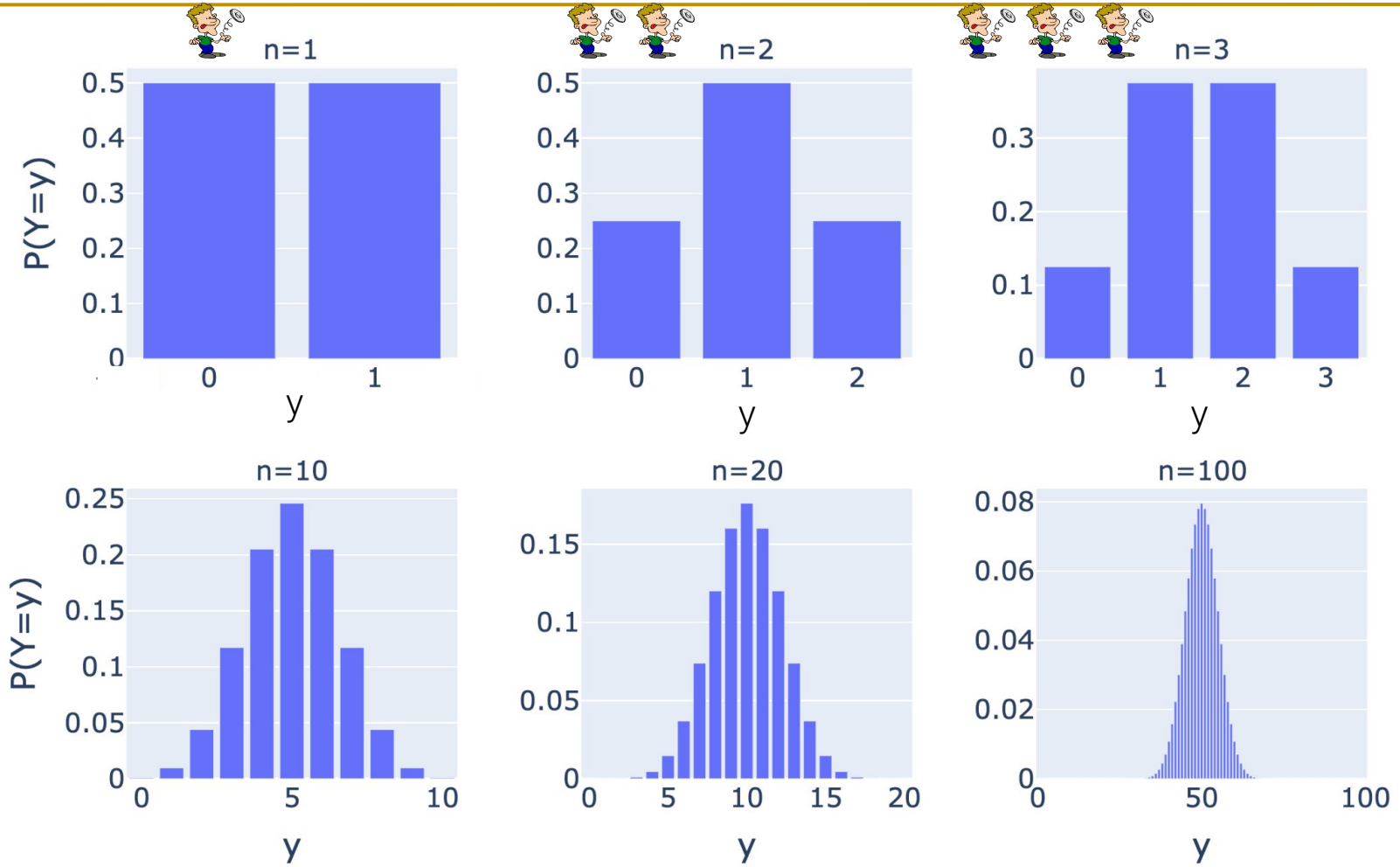
- **(n y)** is the number of ways to arrange **y** successes among **n** trials.
- **p** is the probability of success on a single trial. $\mathbf{p^y}$ represents **y** successes.
- There are **n** total trials. If **y** are successful, then **n-y** are failures. $\mathbf{(1-p)^{n-y}}$ represents n-y failures.

We don't expect you to compute complex binomial probabilities by hand. Just make sure to understand the logic!

CC BY NC SA

Visualizing the Binomial Distribution for p=0.5 and various n

3778394



CC BY NC SA

Continuing our investment story

3778394

Suppose we now make **100** independent investments, each with **10%** probability of success.

The total number of successful investments is an RV Y, where **Y ~ Binomial(n=100, p=0.1)**.

Equivalently, $\mathbf{Y = X_1 + X_2 + \ldots + X_{100}}$, where $\mathbf{X_i \sim Bernoulli(p=0.1)}$ and X_i's are iid.

What is E(Y), the expected number of successful investments?

CC BY NC SA

3778394

Suppose we now make 100 independent investments, each with 10% probability of success. The total number of successful investments is an RV Y, where Y ~ Binomial(n=100, p=0.1).

What is E(Y), the expected number of successful investments?

Presenting with animations, GIFs or speaker notes? Enable our Chrome extension

slido

Continuing our investment story

3778394

Suppose we now make 100 independent investments, each with **10%** probability of success.

The total number of successful investments is an RV Y, where **Y ~ Binomial(n=100, p=0.1)**.

Equivalently, $Y = X_1 + X_2 + \ldots + X_{100}$, where $X_i \sim$ Bernoulli(p=0.1) and X_i's are iid.

What is E(Y), the expected number of successful investments?

A. 10 investments
B. Less than 10 investments
C. Greater than 10 investments

Intuitively, we might expect 10% of the 100 investments to be successful.

CC BY NC SA

Properties of Binomial Random Variables

3778394

Let Y be a **Binomial**(n, p) random variable.

$$Y = \sum_{i=1}^{n} X_i$$, where X_i's are **Bernoulli**(p) and **i.i.d.** Recall that **$E(X_i)=p$** and **$Var(X_i)=p(1-p)$**.

Expectation:

Variance:

CC BY NC SA

Properties of Binomial Random Variables

3778394

Let Y be a **Binomial**(n, p) random variable.

$$Y = \sum_{i=1}^{n} X_i$$, where X_i's are **Bernoulli**(p) and **i.i.d.** Recall that **$E(X_i)=p$** and **$Var(X_i)=p(1-p)$**.

Expectation:

$$\mathbb{E}(Y) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}(X_i) = np$$

Variance: Because all X_i's are independent, $Cov(X_i, X_j) = 0$, so $Var(X_i + X_j) = Var(X_i) + Var(X_j)$

$$\mathrm{Var}(Y) = \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = np(1-p)$$

CC BY NC SA

3778394

Populations and Samples

Lecture 17, Data 100 Spring 2025

Random Variables and Distributions

Expectation and Variance

Sums of Random Variables

- Example: Bernoulli and Binomial Random Variables

Populations and Samples

CC BY NC SA

From Population to Sample

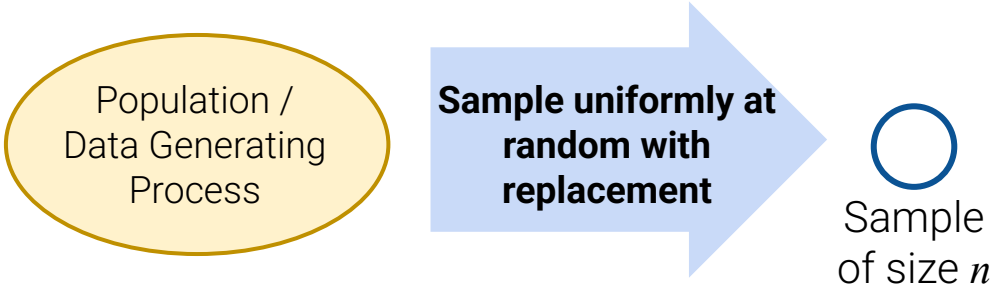
3778394

If we know the **distribution of an RV**, we can reliably compute expectation, variance, etc.

But, in reality, we typically don't know this **true** distribution (i.e., **population** distribution).

Instead, we typically collect a **sample** from some population distribution, and then use that sample to make **inferences** about population properties (e.g., its mean, variance, etc.)

The **big assumption** we often make in modeling/inference:



For large populations, this is effectively a simple random sample (SRS).

CC BY NC SA

We often assume that the sample is a set of i.i.d. RVs

3778394

Population / Data Generating Process

Sample uniformly at random with replacement

Sample of size n

x	P(X = x)
3	0.1
4	0.2
6	0.4
8	0.3

or

	x
0	6
1	6
2	4
3	3
4	6
...	...
99995	6
99996	6
99997	4
99998	8
99999	8
...	...

Data-generating distribution

Population (really large or infinite N)

```
df.sample(n,
replace=True)
```
[documentation]

	x
0	6
1	8
2	4
3	6
4	6
...	...
95	6
96	4
97	8
98	6
99	8

Each observation in our sample is an RV drawn **i.i.d.** from our population distribution.

Sample (n << N)

$$\boxed{X_1, X_2, \ldots, X_n}$$

CC BY NC SA

We often assume that the sample is a set of i.i.d. RVs

3778394

Population / Data Generating Process

Sample uniformly at random with replacement

Sample of size n

x	P(X = x)
3	0.1
4	0.2
6	0.4
8	0.3

or

	X
0	6
1	6
2	4
3	3
4	6
...	...
99995	6
99996	6
99997	4
99998	8
99999	8
...	...

E[X]= 5.9

Population Mean
A **constant**,
i.e., fixed value

`df.sample(n, replace=True)`
[documentation]

	X
0	6
1	8
2	4
3	6
4	6
...	...
95	6
96	4
97	8
98	6
99	8

Sample Mean
A **random variable**!
Depends on our randomly drawn sample.

`np.mean(...) = 5.71`

Sample
(n << N)

$X_1, X_2, \ldots, X_n$

CC BY NC SA

Using the Sample Mean to Estimate the Population Mean

3778394

We often want to **estimate** some characteristic of a population. For example, what is the average height of all Cal undergraduates?

To estimate the population average, we typically collect a **single sample** and find its average.

We might sample 3 undergrads: X_1 = 64in, X_2= 70in, X_3= 66in. <u>Sample mean</u>: 66.7in

Since the sample is random, it *could have* come out differently. → e.g., sample mean of 68.1in

Can we quantify how different the sample mean could have been (i.e., in a parallel universe)?

CC BY NC SA

Central Limit Theorem (Data 8 + today's terminology)

3778394

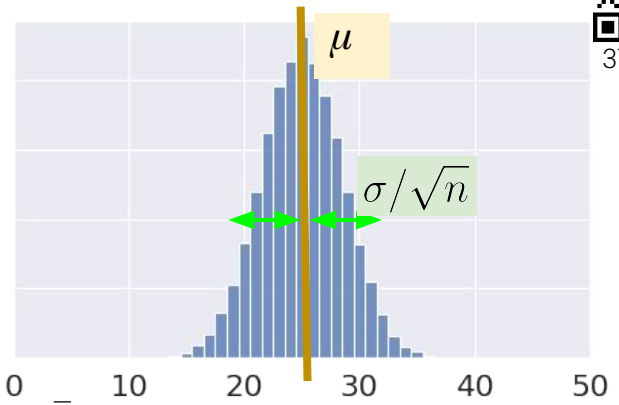
For an i.i.d. sample of X_i's of size n ($X_1, \ldots, X_n$),

Where n is "big enough",

and $X_i \sim$ Unknown, where $E(X_i)=\mu$ and $SD(X_i)=\sigma$,

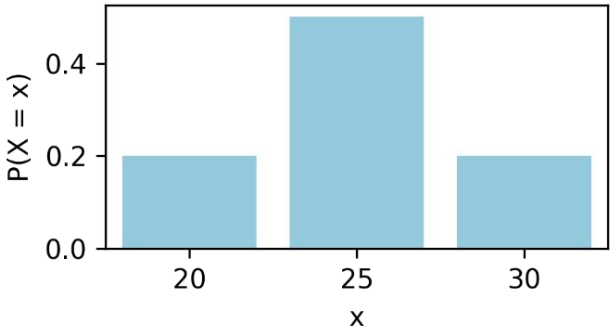
the distribution of $\bar{X}_n$, the **sample mean** of X_i's ,

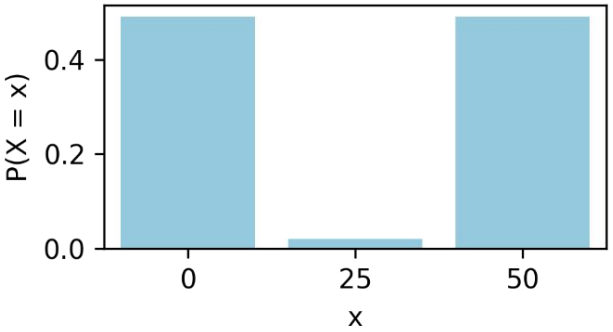
is **roughly normal** with mean μ and SD $\sigma/\sqrt{n}$.

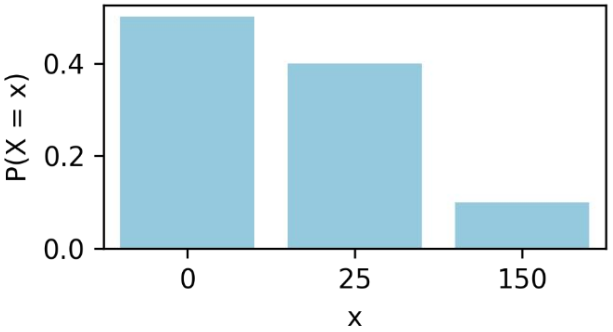


$\bar{X}_n$ for multiple samples of size n

Three possible distributions of X_i where μ=25







How large does n have to be? If true distribution of X_i's is **roughly symmetric and unimodal**/uniform, could need as few as **n = 20**.
If population is very skewed, you will need bigger n.

CC BY NC SA

Central Limit Theorem (Data 8 + today's terminology)

3778394

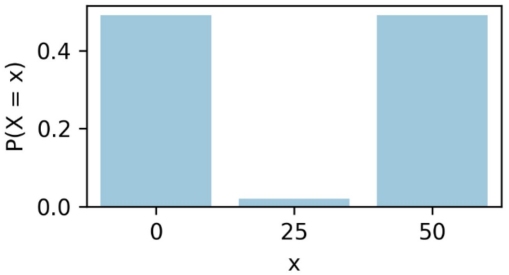
$$X_i \sim \text{Unknown}$$

$$\mathbb{E}(X_i) = \mu$$

$$\text{SD}(X_i) = \sigma$$

Population / Data Generating Process

A possible distribution of X

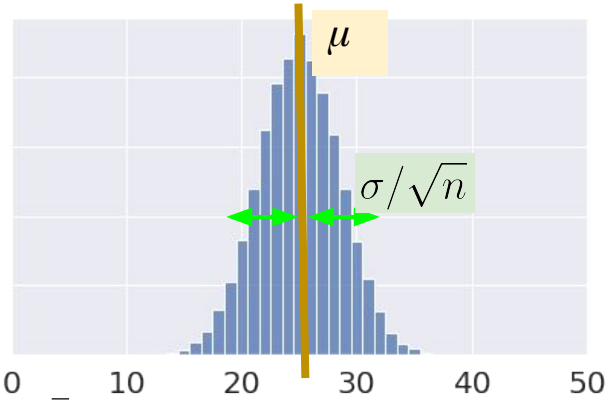


$$\bar{X}_n \sim \text{Normal}(\mu, \frac{\sigma}{\sqrt{n}})$$

$$\mathbb{E}(\bar{X}_n) = \mu$$

$$\text{SD}(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$$

Many samples of size n



$\bar{X}_n$ for many samples of size n

CC BY NC SA

Central Limit Theorem (Data 8 + today's terminology)

3778394

For an i.i.d. sample of X_i's of size n,

Where n is "big enough",

and $X_i \sim$ Unknown, where $E(X_i)=\mu$ and $SD(X_i)=\sigma$,

the distribution of $\bar{X}_n$, the **sample mean** of X_i's ,

is **roughly normal** with mean μ and SD $\sigma/\sqrt{n}$.

Proof out of scope

(Let's prove it!)

Sample mean of X

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

CC BY NC SA

Central Limit Theorem (Data 8 + today's terminology)

3778394

For an i.i.d. sample of X_i's of size n,

Where n is "big enough",

and $X_i \sim$ Unknown, where $E(X_i)=\mu$ and $SD(X_i)=\sigma$,

the distribution of $\bar{X}_n$, the **sample mean** of X_i's ,

is **roughly normal** with $\underbrace{\text{mean } \mu \text{ and SD } \sigma/\sqrt{n}}_{\text{(Let's prove it!)}}$

Sample mean of X

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Expectation:

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i]$$

$$= \frac{1}{n}(n\mu) = \mu$$

Variance/Standard Deviation:

$$\mathrm{Var}(\bar{X}_n) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\left(\underbrace{\sum_{i=1}^{n}\mathrm{Var}(X_i)}_{\text{IID} \rightarrow \mathrm{Cov}(X_i, X_j) = 0}\right)$$

$$= \frac{1}{n^2}\left(n\sigma^2\right) = \frac{\sigma^2}{n}$$

$$\mathrm{SD}(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$$

CC BY NC SA

3778394

LECTURE 17

Random Variables

Content credit: Acknowledgments

CC BY NC SA

3778394

Appendix

Supplemental Slides and Extra Practice

BY NC SA

Extra Practice: Expected Value of a Categorical Distribution

3778394

Suppose a waiter earns tips according to this **categorical** distribution:

x	Pr(X=x)
$0	0.1
$10	0.2
$20	0.3
$100	0.4

On a random day, how much should the waiter **expect** to earn?

A. $100
B. ($0 + $10 + $20 + $100) / 4 = **$32.50**
C. $0(0.1) + $10(0.2) + $20(0.3) + $100(0.4) = **$48**
D. Something else

CC BY NC SA

Extra Practice: Expected Value of a Categorical Distribution

3778394

Suppose a waiter earns tips according to this **categorical** distribution:

x	Pr(X=x)
$0	0.1
$10	0.2
$20	0.3
$100	0.4

On a random day, how much should the waiter **expect** to earn?

A. $100

B. ($0 + $10 + $20 + $100) / 4 = **$32.50**

C. $0(0.1) + $10(0.2) + $20(0.3) + $100(0.4) = **$48**

D. Something else

CC BY NC SA

Extra Practice: Expected Value of a Categorical Distribution

3778394

x	Pr(X=x)
$0	0.1
$10	0.2
$20	0.3
$100	0.4

In 100 nights of tips, we would expect to see, on average:

$0 on 10 nights

$10 on 20 nights

$20 on 30 nights

$100 on 40 nights

100 night total = ($0)(10) + ($10)(20) + ($20)(30) + ($100)(40) = $4800

Average over 100 nights = $4800 / 100 = **$48**

CC BY NC SA

Extra Practice: Expected Value of a Categorical Distribution

3778394

x	Pr(X=x)
$0	0.1
$10	0.2
$20	0.3
$100	0.4

The waiter earns $48 a night, in **expectation** (i.e., on average).

Our computation:

[($0)(10) + ($10)(20) + ($20)(30) + ($100)(40)] / 100 = $48

Equivalent:

[($0)(0.1) + ($10)(0.2) + ($20)(0.3) + ($100)(0.4)] = $48

The expected value (i.e., the expectation) of any discrete RV X:

$$\mathbb{E}(X) = \sum_{x} x \cdot P(X = x)$$

CC BY NC SA

Extra Practice: Variance of a Categorical Distribution

3778394

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

What is the variance of the tips earned by the waiter? Recall that E(X) = $48. $\mathbb{E}(X) = \sum_x x \cdot P(X = x)$

x	Pr(X=x)	x - E(X)	(x - E(X))^2
$0	0.1		
$10	0.2		
$20	0.3		
$100	0.4		

CC BY NC SA

Extra Practice: Variance of a Categorical Distribution

3778394

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

What is the variance of the tips earned by the waiter? Recall that E(X) = $48. $\mathbb{E}(X) = \sum_x x \cdot P(X = x)$

x	Pr(X=x)	x - E(X)	(x - E(X))^2
$0	0.1	-$48	2304
$10	0.2	-$38	1444
$20	0.3	-$28	784
$100	0.4	$52	2704

2304(0.1) +

1444(0.2) +

784(0.3) +

2704(0.4)

= 1836 dollars^2

CC BY NC SA

Extra Practice: Variance: Alternate Calculation

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

3778394

Variance is more conveniently calculated with this formula (proof coming shortly):

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

What is the variance of the tips earned by the waiter? Recall that E(X) = $48. $\mathbb{E}(X) = \sum_x x \cdot P(X = x)$

x	Pr(X=x)	x^2
$0	0.1	
$10	0.2	
$20	0.3	
$100	0.4	

CC BY NC SA

Extra Practice: Variance: Alternate Calculation

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

3778394

Variance is more conveniently calculated with this formula (proof coming shortly):

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

What is the variance of the tips earned by the waiter? Recall that E(X) = $48. $\mathbb{E}(X) = \sum_x x \cdot P(X = x)$

x	Pr(X=x)	x^2
$0	0.1	0
$10	0.2	100
$20	0.3	400
$100	0.4	10000

0(0.1) +

100(0.2) +

400(0.3) +

10000(0.4)

= 4140 dollars^2

4140 - 48^2 =

1836 dollars^2

CC BY NC SA

Extra Practice: Expectation and Variance of a Single Die Roll

3778394

Let X be the outcome of a single die roll.

X is a random variable.

$$P(X = x) = \begin{cases} 1/6 & \text{if } x \in \{1, 2, 3, 4, 5, 6\} \\ 0 & \text{otherwise} \end{cases}$$

1. What is the expectation, $E[X]$?

$$\mathbb{E}[X] = \sum_x x P(X = x)$$

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

(definitions/properties)

2. What is the variance, $\mathrm{Var}(X)$?

CC BY NC SA

Extra Practice: Expectation and Variance of a Single Die Roll

3778394

Let X be the outcome of a single die roll.

X is a random variable.

$$P(X = x) = \begin{cases} 1/6 & \text{if } x \in \{1, 2, 3, 4, 5, 6\} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}[X] = \sum_x x P(X = x)$$

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

1. What is the expectation, $E[X]$?

$$\begin{aligned} \mathbb{E}[X] &= 1(1/6) + 2(1/6) + 3(1/6) + 4(1/6) + 5(1/6) + 6(1/6) \\ &= (1/6)(1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2} \end{aligned}$$

2. What is the variance, $\mathrm{Var}(X)$?

CC BY NC SA

Extra Practice: Expectation and Variance of a Single Die Roll

3778394

Let X be the outcome of a single die roll.

X is a random variable.

$$P(X = x) = \begin{cases} 1/6 & \text{if } x \in \{1, 2, 3, 4, 5, 6\} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}[X] = \sum_x x P(X = x)$$

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

1. What is the expectation, $E[X]$?

$$\begin{aligned} \mathbb{E}[X] &= 1(1/6) + 2(1/6) + 3(1/6) + 4(1/6) + 5(1/6) + 6(1/6) \\ &= (1/6)(1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2} \end{aligned}$$

2. What is the variance, $\mathrm{Var}(X)$?

Approach 1: Definition

$$\begin{aligned} \mathrm{Var}(X) = (1/6)\big((1 - 7/2)^2 &+ (2 - 7/2)^2 \\ &+ (3 - 7/2)^2 + (4 - 7/2)^2 \\ &+ (5 - 7/2)^2 + (6 - 7/2)^2\big) \\ = 35/12 \end{aligned}$$

Approach 2: Property

$$\mathbb{E}[X^2] = \sum_x x^2 P(X = x) \quad = 91/6$$

$$\mathrm{Var}(X) = 91/6 - (7/2)^2 = 35/12$$

CC BY NC SA

Which Would You Pick?

$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X, Y)$
$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$

2214265

Suppose you win cash based on the number of heads you get in a series of 20 coin flips.

Let $X_1, X_2, \ldots, X_{20}$ be 20 **IID** Bernoulli(0.5) random variables.

- Since X_i s are independent: $\mathrm{Cov}(X_i, X_j) = 0$ for all i, j.
- Since X_i is Bernoulli(p = 0.5): $E[X_i] = p = 0.5$, $\mathrm{Var}(X_i) = p(1-p) = 0.25$.

	A. $Y_A = \$10 \bullet X_1 + \$10 \bullet X_2$	**B.** $Y_B = \$\left(\sum_{i=1}^{20} X_i\right)$	**C.** $Y_C = \$20 \bullet X_1$
Expectation	$E[Y_A] = 10(0.5) + 10(0.5) = 10$	$E[Y_B] = 0.5 + \ldots + 0.5 = 10$	$E[Y_C] = 20(0.5) = 10$
Variance	$\mathrm{Var}(Y_A) = 10^2(0.25) + 10^2(0.25)$ $= 50$	$\mathrm{Var}(Y_B) = 0.25 + \ldots + 0.25$ $= 20(0.25)$ $= 5$	$\mathrm{Var}(Y_C) = 20^2(0.25)$ $= 100$
Std. Deviation	$\mathrm{SD}(Y_A) \approx$ **7.07**	$\mathrm{SD}(Y_B) \approx$ **2.24**	$\mathrm{SD}(Y_C) =$ **10**

CC BY NC SA

Proof of Property of Expectation #1 (Out of Scope for Data 100)

3778394

Recall definition of expectation: $\mathbb{E}[X] = \sum_x xP(X = x)$

1. **Expectation is linear**:

(intuition: summations are linear)

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

$$\mathbb{E}[aX + b] = \sum_{ax+b} (ax + b)P(aX + b = ax + b)$$ Definition of expectation.

$$= \sum_x (ax + b)P(X = x)$$ a and b are constants. We only need to iterate over x. For P(...), subtract b and divide by a.

$$= \sum_x axP(X = x) + \sum_x bP(X = x)$$ Linearity of sums.

$$= a\sum_x xP(X = x) + b\sum_x P(X = x)$$ Constants in sums.

$$= a\mathbb{E}[X] + b$$ Definition of expectation. Sum of a distribution is 1.

CC BY NC SA

Proof of Property of Expectation #2 (Out of Scope for Data 100)

3778394

Recall definitions of expectation: $\mathbb{E}[X] = \sum_x xP(X = x)$

2. **Expectation is linear in sums of RVs**: $E[X+Y] = E[X] + E[Y]$

$$\mathbb{E}[X+Y] = \sum_x \sum_y (x+y)P(X = x \text{ and } Y = y)$$

Definition of expectation. Iterate over all possible x and y combinations.

$$= \sum_x \sum_y xP(X = x \text{ and } Y = y) + \sum_x \sum_y yP(X = x \text{ and } Y = y)$$

Linearity of sum.

$$= \sum_x x \sum_y P(X = x \text{ and } Y = y) + \sum_y y \sum_x P(X = x \text{ and } Y = y)$$

Move x and y out of sum.

$$= \sum_x xP(X = x) + \sum_y yP(Y = y)$$

Law of Total Probability.

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

CC BY NC SA

Proof of Property of Expectation #3 (Out of Scope for Data 100)

3778394

Recall definitions of expectation: $\mathbb{E}[X] = \sum_x xP(X = x)$

3. **g(E(X)) ≠ E(g(X)). For example, $E(X)^2 \neq E(X^2)$.**

$$\mathbb{E}[X]^2 \neq \mathbb{E}[X^2]$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \geq 0$$

Derived later in lecture.

$$\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$$

$$\text{Define } g(x) = x^2.$$

$$\text{Then, } \mathbb{E}[g(x)] \geq g(\mathbb{E}[X]).$$

CC BY NC SA

Derive the alternative variance definition. You should understand each step!

3778394

Original Definition:

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

Computational Definition:

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

Useful Properties:

Property 1:

$$E[aX + b] = aE[X] + b$$

Property 2:

$$E[X + Y] = E[X] + E[Y]$$

CC BY NC SA

Derive the alternative variance definition. You should understand each step!

3778394

Original Definition:

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

Computational Definition:

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

Useful Properties:

Property 1:

$$E[aX + b] = aE[X] + b$$

Property 2:

$$E[X + Y] = E[X] + E[Y]$$

$$Var(X) = E\left[(X - E[X])^2\right]$$

Constants inside expectations are colored **blue**.

$$= E\left[X^2 - 2E[X]X + E[X]^2\right]$$ Expanding the square

$$= E\left[X^2\right] - E\left[2E[X]X\right] + E\left[E[X]^2\right]$$ by Property 2

$$= E\left[X^2\right] - 2E[X]E\left[X\right] + E[X]^2$$ by Property 1

$$= E\left[X^2\right] - 2E[X]^2 + E[X]^2$$

$$= E\left[X^2\right] - E[X]^2$$ Algebra

CC BY NC SA

Derive the alternative variance definition. You should understand each step!

3778394

Original Definition:

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

Computational Definition:

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

Useful Properties:

Property 1:

$$E[aX + b] = aE[X] + b$$

Property 2:

$$E[X + Y] = E[X] + E[Y]$$

$$Var(X) = E\left[(X - E[X])^2\right]$$ Constants inside expectations are colored **blue**.

$$= E\left[X^2 - 2E[X]X + E[X]^2\right]$$ Expanding the square

$$= E\left[X^2\right] - E\left[2E[X]X\right] + E\left[E[X]^2\right]$$ by Property 2

$$= E\left[X^2\right] - 2E[X]E\left[X\right] + E[X]^2$$ by Property 1

$$= E\left[X^2\right] - 2E[X]^2 + E[X]^2$$

$$= E\left[X^2\right] - E[X]^2$$ Algebra

CC BY NC SA

Proof of Non-Linearity of Variance (You Should Understand Each Step)

3778394

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

Properties:

1. **Variance is non-linear**:

Shifting doesn't change variance, scaling does

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

$$\mathrm{SD}(aX + b) = |a|\mathrm{SD}(X)$$

Proof:

$$\begin{aligned}
\mathbf{Var}\left[aX + b\right] &= \mathbf{E}\left[(aX + b)^2\right] - \left(\mathbf{E}\left[aX + b\right]\right)^2 \\
&= \mathbf{E}\left[(aX + b)^2\right] - \left(a\mathbf{E}\left[X\right] + b\right)^2 \\
&= \mathbf{E}\left[(aX + b)^2\right] - a^2\mathbf{E}[X]^2 - 2ab\mathbf{E}[X] - b^2 \\
&= \mathbf{E}\left[a^2X^2 + 2abX + b^2\right] - a^2\mathbf{E}[X]^2 - 2ab\mathbf{E}[X] - b^2 \\
&= a^2\mathbf{E}\left[X^2\right] + 2ab\mathbf{E}\left[X\right] + b^2 - a^2\mathbf{E}[X]^2 - 2ab\mathbf{E}[X] - b^2 \\
&= a^2\left(\mathbf{E}\left[X^2\right] - \mathbf{E}[X]^2\right) \\
&= a^2\mathbf{Var}\left[X\right]
\end{aligned}$$

CC BY NC SA

Proof of Non-Linearity of Variance (You Should Understand Each Step)

3778394

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

Properties:

1. **Variance is non-linear**:

Shifting doesn't change variance, scaling does

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

$$\mathrm{SD}(aX + b) = |a|\mathrm{SD}(X)$$

Proof:

$$\begin{aligned}
\mathbf{Var}\left[aX + b\right] &= \mathbf{E}\left[(aX + b)^2\right] - \left(\mathbf{E}\left[aX + b\right]\right)^2 && \text{Computational Formula} \\
\text{Linearity of Exp.} \quad &= \mathbf{E}\left[(aX + b)^2\right] - \left(a\mathbf{E}\left[X\right] + b\right)^2 \\
\text{Algebra} \quad &= \mathbf{E}\left[(aX + b)^2\right] - a^2\mathbf{E}[X]^2 - 2ab\mathbf{E}[X] - b^2 \\
\text{Algebra} \quad &= \mathbf{E}\left[a^2X^2 + 2abX + b^2\right] - a^2\mathbf{E}[X]^2 - 2ab\mathbf{E}[X] - b^2 \\
\text{Linearity of Exp.} \quad &= a^2\mathbf{E}\left[X^2\right] + 2ab\mathbf{E}\left[X\right] + b^2 - a^2\mathbf{E}[X]^2 - 2ab\mathbf{E}[X] - b^2 \\
\text{Algebra} \quad &= a^2\left(\mathbf{E}\left[X^2\right] - \mathbf{E}[X]^2\right) \\
\text{Computational Formula} \quad &= a^2\mathbf{Var}\left[X\right]
\end{aligned}$$

CC BY NC SA

Proof of Variance of a Sum (Out of Scope for Data 100)

Jump back: link

3778394

The variance of a sum is affected by the dependence between the two random variables that are being added. Let's expand out the definition of Var(X + Y) to see what's going on.

Let $\mu_x = E[X], \mu_y = E[Y]$

$$
\begin{aligned}
Var(X+Y) &= E\left[(X+Y-E(X+Y))^2\right] \\
&= E\left[((X-\mu_x)+(Y-\mu_y))^2\right] \\
&= E\left[(X-\mu_x)^2 + 2(X-\mu_x)(Y-\mu_y) + (Y-\mu_y)^2\right] \\
&= E\left[(X-\mu_x)^2\right] + E\left[(Y-\mu_y)^2\right] + 2E\left[(X-\mu_x)(Y-\mu_y)\right] \\
&= Var(X) + Var(Y) + 2E\left[(X-E(X))(Y-E(Y))\right]
\end{aligned}
$$

By the linearity of expectation, and substitution.

CC BY NC SA