Math 183
Statistical Methods

Eddie Aamari
S.E.W. Assistant Professor

eaamari@ucsd.edu
math.ucsd.edu/~eaamari/
AP&M 5880A

Today: Chapter 2 (end)

- Random variables
- Create a probability model for a random variable
- Expected value, variance, standard deviation of a random variable
- Rules on linear combinations of random variables

Answer Duration at Office Hours

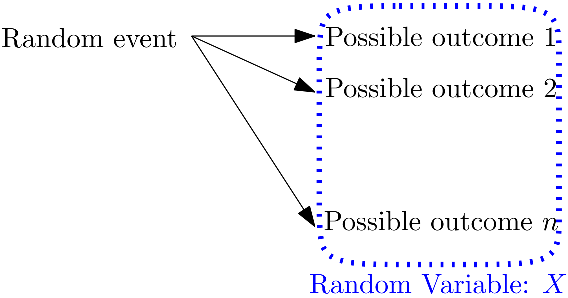
The instructor's office hours are very popular: you usually have to wait to ask your question. In the rare event that the instructor is in a bad mood (10% of the time), he answers your question in 8 minutes. When he is in a good mood, he answers in 2 minutes.
How long do you expect his answer to take?

We don't say 5 minutes (the average of 2 and 8) because the values 2 and 8 are not equally likely.

We need a weighted average to find our expected wait time:

$$\text{Expected time} = 0.1 \times 8 + 0.9 \times 2 = 2.6 \text{ minutes}$$

Random Variable

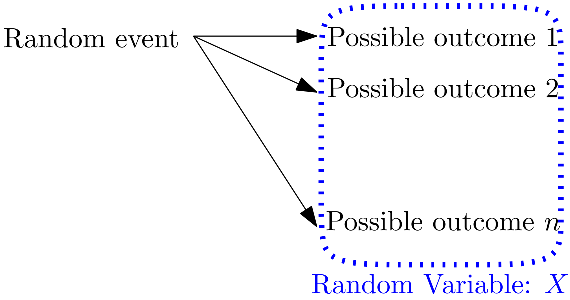


The outcomes of a random variable may not be equally likely
(= they have different probabilities)

A **random variable** is a quantity whose value depends on the outcome of a random event.

A **discrete random variables** is a random variable whose possible outcomes have gaps between them
Example: counts of the number of Heads in 7 coin flips.

Random Variable



Conventions of notation:

- We use capital letters, like X, to denote a random variable
- The values of a random variable are denoted with lower case letters, like x.
- For instance, $P(X = x)$ denotes the probability that the random variable X takes the particular value x.

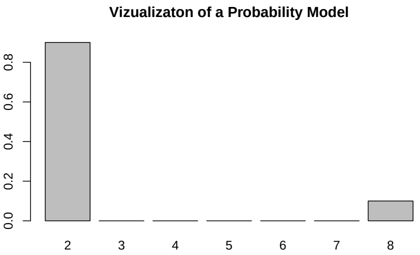
Probability Model

A **probability model** is:

- The list of all the possible outcomes of a random variable
- The probability of each outcome

Example: X = answer duration at office hours

Outcome:	x	2	8
Probability:	P(X = x)	0.9	0.1



Expected Value

X = answer duration at office hours

Outcome:	x	2	8
Probability:	P(X = x)	0.9	0.1

The expected value of X is $E(X) = 0.1 \times 8 + 0.9 \times 2 = 2.6$.

In general, the **expected value** of a discrete random variable is

$$E(X) = \sum_x P(X = x) \times x.$$

Synonyms: expectation, mean value, weighted average.

I Want To Play a Game...

A friend suggests you play a game with a six-sided die according to the following rules:

- If you roll an even number, you pay \$4.
- If you roll an odd number, then you get a new roll. If you match your first roll, you make \$72. But if you don't, you pay \$12.

Should you play?

Let X be the amount of money you make in one game.

Outcome:	x	-4\$	72\$	-12\$
Probability:	P(X = x)	1/2	$1/2 \times 1/6$	$1/2 \times 5/6$

$$
\begin{aligned}
E(X) &= \sum_x P(X = x) \times x \\
&= \frac{1}{2} \times (-4) + \frac{1}{2} \times \frac{1}{6} \times 72 + (-12) \times \frac{1}{2} \times \frac{5}{6} \\
&= -1\$
\end{aligned}
$$

I Want To Play a Game...

Another way to think about expected value is with the Law of Large Numbers: imagining you play infinitely many times. The money exchange might look like

-4\$,-12\$,-12\$,-4\$,-4\$,72\$,72\$,-4\$,-4\$,-4\$,-12\$,...

If you averaged all these values, you'd get -1\$ after an infinite number of games. In this point of view, the weighting already appears because outcomes have different probabilities.

Parameters of Models

We often denote by $\mu = E(X)$ the expected value of random variables.

μ (or $E(X)$) is called a **parameter** of the random variable and its associated probability model. A parameter is a value that helps summarize a probability model.

Another parameter we might care about is some measure of spread for the random variable.

Here, the spread should give some sense for how much the outcomes will vary from μ.

Variance and Standard Deviation

The **Variance** of X is

$$\sigma^2 = Var(X) = \sum_x (x - \mu)^2 P(X = x).$$

In our example

x	-4\$	72\$	-12\$
P(X = x)	1/2	1/2 × 1/6 = 1/12	1/2 × 5/6 = 5/12

$$\begin{aligned}\sigma^2 = Var(X) &= (-4 - (-1))^2 \frac{1}{2} + (72 - (-1))^2 \frac{1}{12} + (-12 - (-1))^2 \frac{5}{12} \\ &= 499\$^2.\end{aligned}$$

Remark: The unit of the variance $Var(X)$ is the square of the unit of the random variable X.

Variance and Standard Deviation

The **standard deviation** of X is the square root of the variance:

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

In our example, $\sigma = \sqrt{499\$^2} = \$22.34$

Remark: Unit of the standard deviation $SD(X)$ matches that of X.

Because $\mu = -\$1$ but the standard deviation is \$22.34, we see that some of the outcomes are good for the player, tempting the player to want to join in.

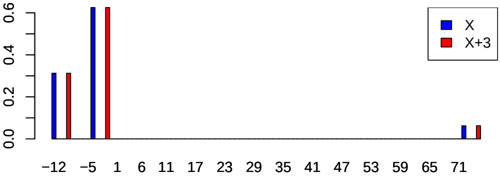
Combining Random Variables

Your friend feels bad about this nasty game. She offers to pay you \$3 any time you want to play it. Now, what is the expected value of the game?

New situation: $3 + X$. It seems like the expected value should be $39(-1) = \$2$.

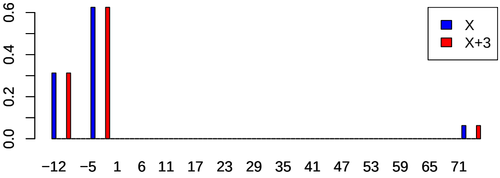
Indeed, for any constant c,

$$E(X + c) = E(X) + c.$$



Here, since $E(X + 3) = \$2 > 0$, you should play the game.

Combining Random Variables



Have the variance or standard deviation changed in our new version of the game?
No! The spread does not change. For any constant c,

$$Var(X \pm c) = Var(X) \qquad SD(X \pm c) = SD(X)$$

Combining Random Variables

Your friend realizes she is losing money under the new rules, so she stops with the \$3 bonus for playing.
Instead, she decides to double all the oucomes (paying 4 becomes paying 8; earning 72 becomes earning 144; ...)
What are μ, σ^2 and σ now?

$$E(aX) = aE(X)$$

$$Var(aX) = a^2 Var(X) \qquad SD(aX) = |a| SD(X)$$

For our game,

$$E(2X) = 2E(X) = 2 \times (-1) = -2\$$$
$$Var(2X) = 2^2 Var(X) = 4 \times 499 = 1996\$^2$$
$$SD(2X) = 2SD(X) = 2 \times 22.34 = 44.68\$$$

Combining Random Variables

You decide to play the (undoubled) game each weekday and record your weekly earnings. What value should you expect for the weekly earnings? How much variation do the weekly totals have?

As we are repeating the same random process an summing its outcomes, the problem is asking us to explore $X + X + X + X + X$.

Very tempting to write this as $5X$ and do the math with that, but random variables don't act like mathematical variables:

$$X + X + X + X + X \neq 5X.$$

To emphasize this, we write

$$X_1 + X_2 + X_3 + X_4 + X_5.$$

Combining Random Variables

For two random variables X and Y,

$$E(X \pm Y) = E(X) \pm E(Y)$$

$Var(X \pm Y) = Var(X) + Var(Y)$ if X and Y are independent.

$SD(X \pm Y) = \sqrt{Var(X) + Var(Y)}$ if X and Y are independent.

Remarks:

- Two random variables are independent if knowing the outcome of one has no effect on the outcome of the other.
- The fact about $E()$ is true even if X and Y are dependent.
- The Var fact has just a plus on the right hand side. If it had a minus, you'd be able to get negative variance, which makes no sense.

Combining Random Variables

Game	Expected Value	Variance
X	-1	499

We have

$$E(X_1 + X_2 + X_3 + X_4 + X_5)$$
$$= E(X_1) + E(X_2) + E(X_3) + E(X_4) + E(X_5)$$
$$= (-1) + (-1) + (-1) + (-1) + (-1) = -5\$$$

Note that $E(5X) = 5 \times (-1) = -5\$$ (same answer)

Since the payout of one day does not affect any other day, $X_1, \ldots, X_5$ are independent. Hence

$$Var(X_1 + X_2 + X_3 + X_4 + X_5)$$
$$= Var(X_1) + Var(X_2) + Var(X_3) + Var(X_4) + Var(X_5)$$
$$= 499 + 499 + 499 + 499 + 499 = 2495\$^2$$

Note that $Var(5X) = 5^2 \times 499 = 12475\$^2$ (different answer)

Thus, $SD(X_1 + X_2 + X_3 + X_4 + X_5) = \sqrt{2495} \simeq 49.95\$$.

Note that $SD(5X) = 5 \times 22.39 = 111.95\$$ (different answer)

Why are $2X$ and $X_1 + X_2$ Different?

Possible values for $X_1 + X_2$ are

$X_1 \setminus X_2$	-12	-4	72
-12	-24	-16	60
-4	-16	-8	68
72	60	68	144

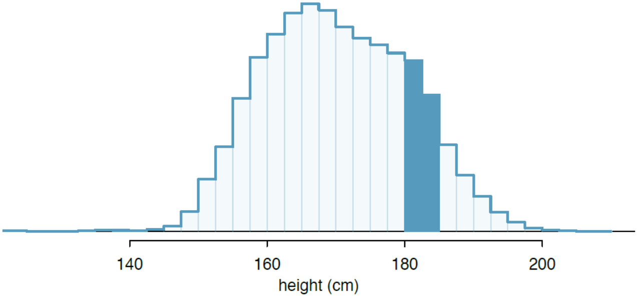
Possible values for $2X$ are $-24, -8$ and 144.

In the $X_1 + X_2$ scenario, we often add winning situations and losing situations which diminishes the influence of one another, creating less dramatic outcomes (= less variance!) than $2X$.

From Histograms to Continuous Distributions

Let X denote the height of a randomly selected US adults.

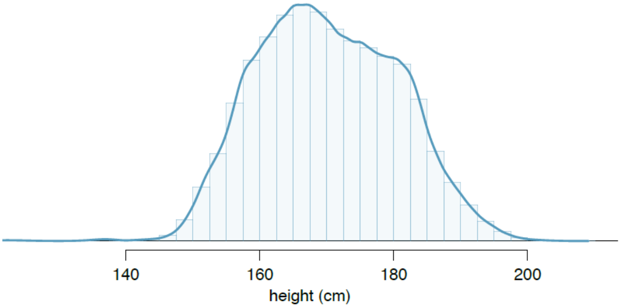
Here is a histogram of the distribution of X.



The proportion of data that fall in the shaded bins gives the probability that a X falls between 180cm and 185cm.

From Histograms to Continuous Distributions

Since height is a continuous random variable, its **probability density function** is a smooth curve



From Histograms to Continuous Distributions

This means that the probability that a randomly sampled US adult is between 180cm and 185cm can be estimated a the shaded area under the curve.