STA220H1: The Practice of Statistics I

Elena Tuzhilina

January 31, 2023

Please turn on your videos :)

Figure 1: [picture source]

Announcements

1. Use this form to provide your feedback for STA220 class. It is anonymous.
2. February 7 (in-person): 1 hour of review + 20 mins break + 80 mins test.
3. Quiz 5 will be optional (for extra grade!). I will post a lab in R and you can practice coding.

Agenda for today

- Recap: sample space, additions and multiplication rules, conditional probability
- Probability: Bayes' rule
- Random variables: discrete, binomial, normal
- Random variables: expectation and variance

Recap: probability vocabulary

Experiment - any activity that produces an outcome

Sample space - the collection of all possible outcomes of an experiment

$$S = \{O_1, O_2, \ldots, O_N\}$$

Event - a subset of the sample space (could be one or more of possible outcomes in the sample space)

$$A = \{O_1, O_3, O_{10}\}$$

$P(A) = ?$

Recap: probability of events

Probability of an outcome is

$$0 \leq P(O_i) \leq 1$$

The total probability of all outcomes is

$$\sum_{i=1}^{N} P(O_i) = 1$$

To compute the probability of a complex event, add together the probabilities of the outcomes

$$P(A) = P(O_1) + P(O_3) + P(O_{10})$$

Recap: probability of events

Often we can assume that all outcomes in the sample space are equally likely

$$\sum_{i=1}^{N} P(O_i) = 1 \qquad P(O_i) = \frac{1}{total\ number\ of\ outcomes}$$

In this case the probability of an event is

$$P(A) = \frac{number\ of\ outcomes\ favorable\ to\ event\ A}{total\ number\ of\ outcomes}$$

Recap: probability of events

- rolling a 6-sided die and getting an odd number

$S = \{1, 2, 3, 4, 5, 6\}$

$P(O_i) = 1/6$

$O =$	1	2	3	4	5	6
$P(O) =$	1/6	1/6	1/6	1/6	1/6	1/6

- giving birth to more than 2 kids

$S = \{0, 1, 2, 3, \ldots 20\}$

$P(O_i) =$ different

$O =$	0	1	2	3	4	....	20
$P(O) =$	0.05	0.1	0.3	0.2	0.05		0.001

Recap: probability rules

Subtraction - probability of event A not happening is one minus the probability of the event happening.

complement of A

$$P(\bar{A}) = 1 - P(A)$$

Multiplication - if events A and B are independent then

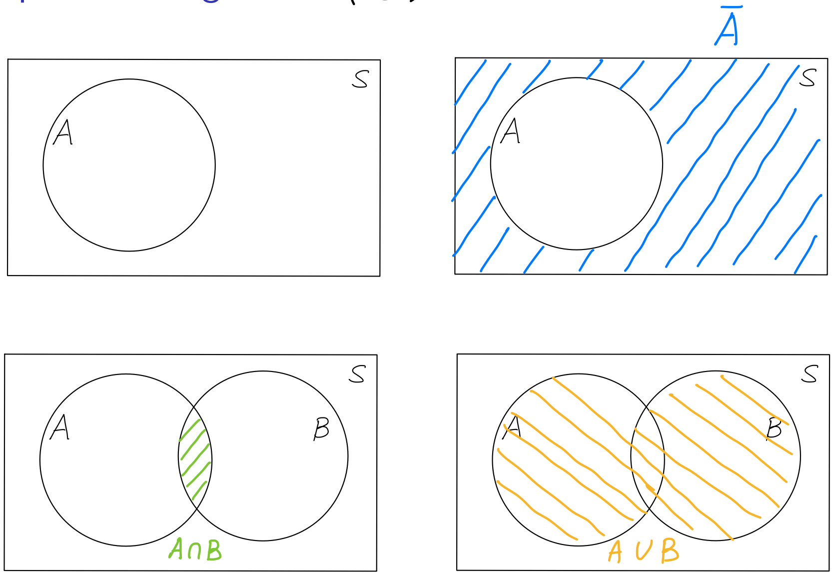
$$P(both\ A\ and\ B\ will\ occur) = P(A) \cdot P(B) = P(A \cap B)$$

Addition - to find the probability of either of two events occurring, add together the individual probabilities, then subtract the probability of both occurring together

$$P(either\ A\ or\ B\ occur) = P(A) + P(B) - P(A \cap B) = P(A \cup B)$$

Recap: Venn diagram

$P(S) = 1$

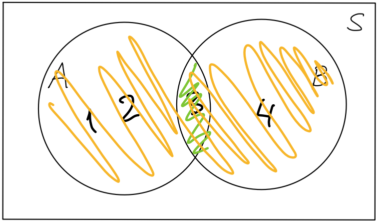


Venn diagram

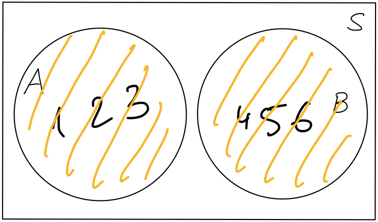
$S = \{1, 2, 3, 4, 5, 6\}$

$P(A \cap B) = 0$

$A \cap B$ – empty

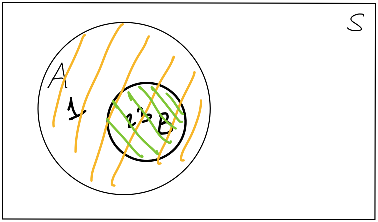


$A = \{1, 2, 3\}$ $B = \{3, 4\}$



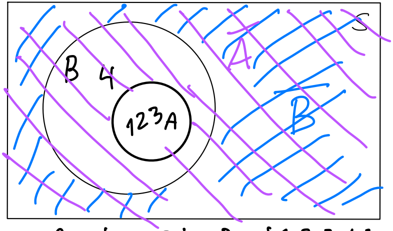
$A = \{1, 2, 3\}$ $B = \{4, 5, 6\}$

$A \cup B = A$ $A \cap B = B$



$A = \{1, 2, 3\}$ $B = \{2, 3\}$

$A \cup B = B$ $A \cap B = A$



$A = \{1, 2, 3\}$ $B = \{1, 2, 3, 4\}$

Conditional probability

We often wish to determine the probability of some event given that some other event has occurred, which are known as **conditional probabilities**.

To compute the conditional probability of A given B we need to know the joint probability $P(A \cap B)$ as well as the marginal probability P(B)

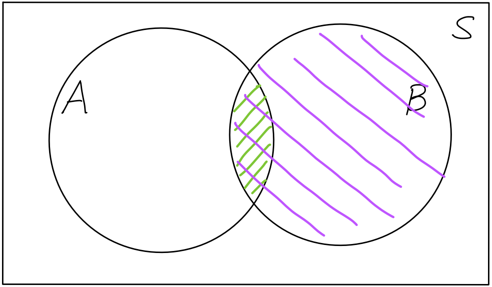
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

A given B

? If A and B are independent, what is $P(A|B)$?

$$P(A|B) = \frac{P(A) \cdot \cancel{P(B)}}{\cancel{P(B)}} = P(A)$$

Venn diagram



$$P(A \mid B) = \frac{\quad}{\quad} = \text{"proportion of As in B"}$$

Exercise

A

What is the probability of $\underbrace{\text{having a heart attack}}_{A}$ given that you are $\underbrace{\text{active}}_{B}$?

$A \cap B$

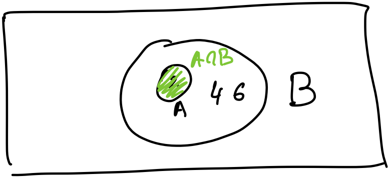
```
##                   active not-active
## heart attack        0.01       0.09
## no heart attack     0.49       0.41
```

$+$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \boxed{\frac{0.01}{0.01 + 0.49}}$$

$$= \frac{1}{50}$$

Exercise



$A = \{2\}$

$B = \{2, 4, 6\}$

$A \cap B = \{2\}$

Suppose we have a standard 6-sided die and roll it once. What is the probability that $\underbrace{\text{we rolled a 2}}_{A}$ given that $\underbrace{\text{the number we rolled was even}}_{B}$?

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/2} = \frac{1}{3}$$

$$\underbrace{\frac{5}{6}} \cdot \underbrace{\frac{5}{6}} \cdot \underbrace{\frac{5}{6}} = P(A \cap B \cap C)$$

A = {do not get 6 in 1st}

B = {... 2nd}

C = {... 3rd}

Independence

Statistical **independence** between events A and B means that knowing about A does not tell us anything about B.

That is, *the probability of A given some value of B is just the same as the overall probability of A*

$$P(A|B) = P(A)$$

This is equivalent to $\frac{P(A \cap B)}{P(B)} = P(A) \Rightarrow$

$$P(A \cap B) = P(A) \cdot P(B)$$

Exercise

Are events *being active* and *heart attack* independent?

A over *being active*, B over *heart attack*

```
##                     active not-active
## heart attack          0.01       0.09
## no heart attack       0.49       0.41
```

$$0.01 = P(A \cap B) \overset{?}{\neq} P(A) \cdot P(B) = 0.5 \cdot 0.1$$

$\Rightarrow$ not independent!

Conditional probability: tree diagram

$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow$

$\Rightarrow P(A \cap B) = P(A|B) \cdot P(B)$

- The probability of getting COVID-19 is

$P(covid) \cdot P(positive) = P(covid \cap positive)$

$$P(covid) = 0.5$$

$P(not\ covid) = 1 - P(covid) = 1 - 0.5$

- Given that you have COVID-19, the probability of getting a positive test is

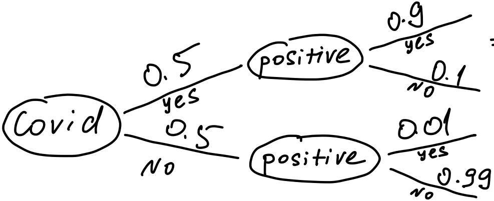
$$P(positive|covid) = 0.9$$

- Given that you do not have COVID-19, the probability of getting a positive test is

$P(positive|covid) \cdot P(covid)$

$$P(positive|no\ covid) = 0.01$$

$\parallel$



$P(covid \cap positive) =$
$= 0.9 \cdot 0.5$

$P(covid \cap negative) =$
$= 0.5 \cdot 0.1$

$P(no\ covid \cap positive) =$
$= 0.5 \cdot 0.01$

$P(no\ covid \cap negative)$

Bayes' rule

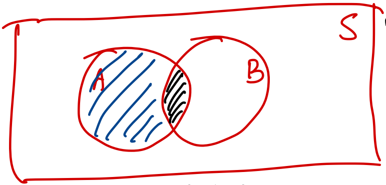
What is the probability to have COVID given that your test is positive?

$$P(covid|positive) = ?$$

$$= \frac{P(covid \cap positive)}{P(positive)} = \frac{0.5 \cdot 0.9}{0.5 \cdot 0.9 + 0.5 \cdot 0.01}$$

$$P(positive) = P(positive \cap covid) + P(positive \cap no\ covid)$$

Bayes' rule



$P(A)$ $P(A \cap B) =$
$P(A|B) \cdot P(B)$
$P(A \cap B) =$
$P(A) \cdot P(B)$

In many cases, we know $P(A|B)$ but we really want to know $P(B|A)$.

In order to reverse a conditional probability, we can use **Bayes' rule**

$$P(B|A) = \frac{\overbrace{P(A|B) \cdot P(B)}^{P(A \cap B)}}{\underbrace{\underbrace{P(A|B) \cdot P(B)}_{P(A \cap B)} + \underbrace{P(A|\bar{B}) \cdot P(\bar{B})}_{P(A \cap \bar{B})}}_{P(A)}}$$

$$P(A \cap B) + P(A \cap \overline{B}) = P(A)$$

Random variable

Random variable takes an outcome from random a experiment and gives a numerical number.

$$O_1 \to 1 \quad O_2 \to 3 \quad O_3 \to 2 \ldots O_N \to 1$$

- number of heads when tossing two coins

$O =$	TT	TH	HT	HH
$X =$	0	1	1	2

- total score when rolling two dice

$O =$	(1,1)	(1,2)	(2,1)	...	(6,5)	(6,6)
$X =$	2	3	3	....	11	12

- number of zeroes in your student ID

$O =$	100841234S	1101020030	....
$X =$	2	5	....

Random variable

Random variable can be discrete or continuous.

Discrete - takes one of a countable list of distinct values

- number of heads when tossing two coins $0, 1, 2$
- total score when rolling two dice $2, \ldots, 12$
- number of zeroes in your student ID $0, \ldots, 10$

Continuous - takes any value in an interval or collection of intervals

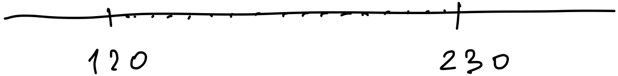
- the wait time for the next bus
- birth weights of babies

Exercise

$\{0, 0.5, 1\}$

Discrete or continuous?

- number of facebook friends $\{0, \ldots, 5000\}$
- heights of the U of T students $\{120\,cm, \ldots, 2.30\,cm\}$



Random variable

The probability distribution for a random variable describes how the ***probabilities*** *are distributed over the* ***values*** *of the random variable.*

- number of heads when tossing two coins

$O=$	TT	TH	HT	HH
$X=$	0	1	1	2
$P(O)=$	$1/4$	$1/4$	$1/4$	$1/4$

$X=$	0	1	2
$P(X)=$	$1/4$	$1/2$	$1/4$

- total score when rolling two dice

$O=$	$(1,1)$	$(1,2)$	$(2,1)$	$\dots$	$(6,5)$	$(6,6)$
$X=$	2	3	3	$\dots$	11	12
$P(O)=$	$1/36$	$1/36$	$1/36$		$1/36$	$1/36$

$(2,2)$ $(1,3)$
$(3,1)$

$X=$	2	3	4	$\dots$	12
$P(X)=$	$1/36$	$1/18$	$1/12$	$\dots$	$1/36$

$(6,6)$

Discrete: Bernoulli random variables

$$\boxed{p = P(X=1)}$$

Has two values 0 and 1 that happen with probabilities ~~*p*~~ $1-p$ *and* ~~*1-p*~~ p*. Usually used to represent experiments with two outcomes.*

- coin: 1 if heads, 0 if tails

  $P(X=1) = 0.5 \quad P(X=0) = 0.5 \quad \boxed{p = 0.5}$

- sex of a baby: 1 if female, 0 if male

  $P(X=1) = 0.52 \quad P(X=0) = 0.48 \quad \boxed{p = 0.52}$

- disease: 1 if sick, 0 if healthy

  $P(X=1) = 0.001 \quad P(X=0) = 0.999 \quad \boxed{p = 0.001}$

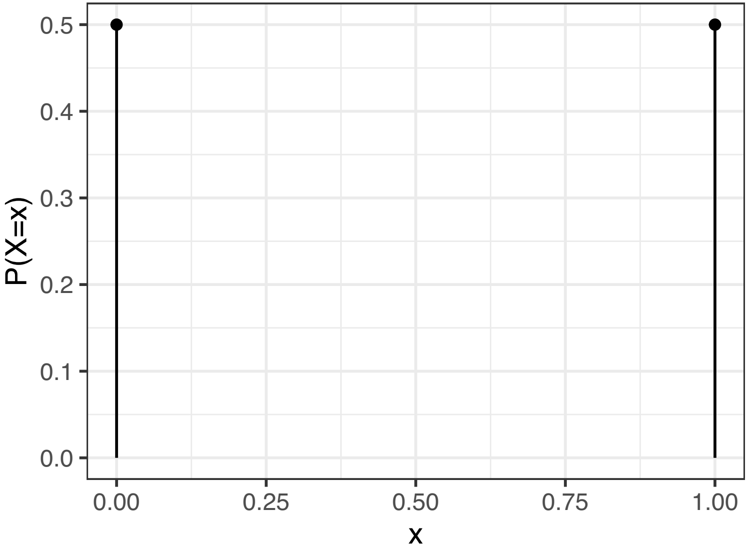
Discrete: Bernoulli random variables

Parameters: p (the probability of 1)

$$P(X = 1) = p \text{ and } P(X = 0) = 1 - p$$

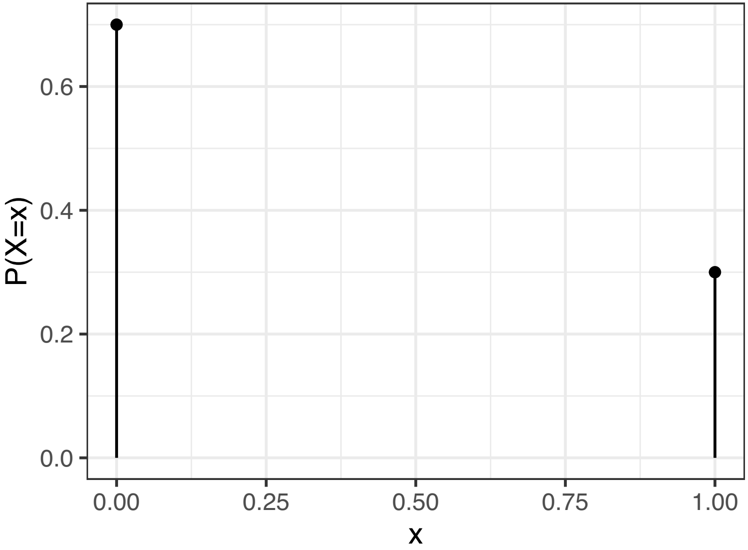
Discrete: Bernoulli random variables

Example: $p = 0.5$



Discrete: Bernoulli random variables

Example: $p = 0.3$



Discrete: Binomial random variable

Suppose we run n independent trials, each trial is "successful" with probability p and "unsuccessful" with probability 1-p (each trial can be represented by a Bernoulli random variable). Binomial random variable represents the number of "successful" trials.

- 5 babies were born in a hospital, number of girls is Binomial

  $n = 5 \quad p = 0.52 \quad X = 0, \dots, 5$

- 10 students were admitted to U of T, number of Canadian students is Binomial

  $n = 10 \quad p = 0.3 \quad X = 0, \dots, 10$

- 3 tomatoes were bought in a supermarket, number of rotten tomatoes is Binomial

  $0.1 \cdot 0.9 \cdot 0.9$

  $n = 3 \quad p = 0.1$

  $0.9 \quad 0.9 \quad 0.9$

$X =$	0	1	2	3
$P(x) =$	0.9^3	$3 \cdot 0.1 \cdot 0.9^2$	$3 \cdot 0.1^2 \cdot 0.9$	0.1^3

Discrete: Binomial random variables

Parameters: p (the probability of success), n (number of trials)

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \text{ for } x = 0, 1, ..., n$$

Discrete: Binomial random variables

Binomial Probability Table

n=Number of trials, k=Number of successes and p=Probability of success

$n = 5$ $k = 3$ $p = 0.2$

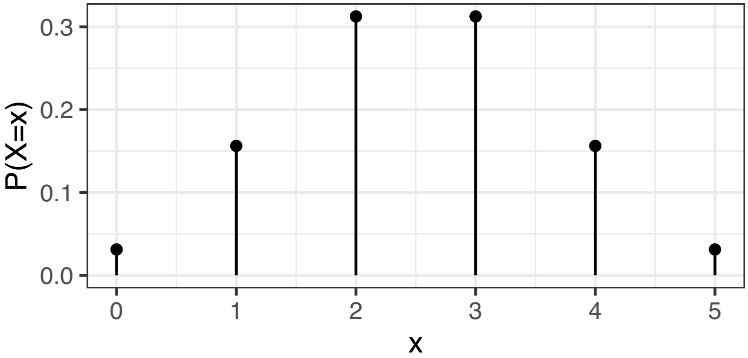
		p										
n	k	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
2	0	0.9801	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
	1	0.0198	0.0950	0.1800	0.2550	0.3200	0.3750	0.4200	0.4550	0.4800	0.4950	0.5000
	2	0.0001	0.0025	0.0100	0.0225	0.0400	0.0625	0.0900	0.1225	0.1600	0.2025	0.2500
3	0	0.9703	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
	1	0.0294	0.1354	0.2430	0.3251	0.3840	0.4219	0.4410	0.4436	0.4320	0.4084	0.3750
	2	0.0003	0.0071	0.0270	0.0574	0.0960	0.1406	0.1890	0.2389	0.2880	0.3341	0.3750
	3		0.0001	0.0010	0.0034	0.0080	0.0156	0.0270	0.0429	0.0640	0.0911	0.1250
4	0	0.9606	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
	1	0.0388	0.1715	0.2916	0.3685	0.4096	0.4219	0.4116	0.3845	0.3456	0.2995	0.2500
	2	0.0006	0.0135	0.0486	0.0975	0.1536	0.2109	0.2646	0.3105	0.3456	0.3675	0.3750
	3		0.0005	0.0036	0.0115	0.0256	0.0469	0.0756	0.1115	0.1536	0.2005	0.2500
	4			0.0001	0.0005	0.0016	0.0039	0.0081	0.0150	0.0256	0.0410	0.0625
5	0	0.9510	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0312
	1	0.0480	0.2036	0.3281	0.3915	0.4096	0.3955	0.3602	0.3124	0.2592	0.2059	0.1562
	2	0.0010	0.0214	0.0729	0.1382	0.2048	0.2637	0.3087	0.3364	0.3456	0.3369	0.3125
	3		0.0011	0.0081	0.0244	0.0512	0.0879	0.1323	0.1811	0.2304	0.2757	0.3125
	4			0.0005	0.0022	0.0064	0.0146	0.0284	0.0488	0.0768	0.1128	0.1562
	5				0.0001	0.0003	0.0010	0.0024	0.0053	0.0102	0.0185	0.0312

Discrete: Binomial random variables

Example: n = 5, p = 0.5

```r
dbinom(x = 0:5, size = 5, prob = 0.5)
```

```
## [1] 0.03125 0.15625 0.31250 0.31250 0.15625 0.03125
```

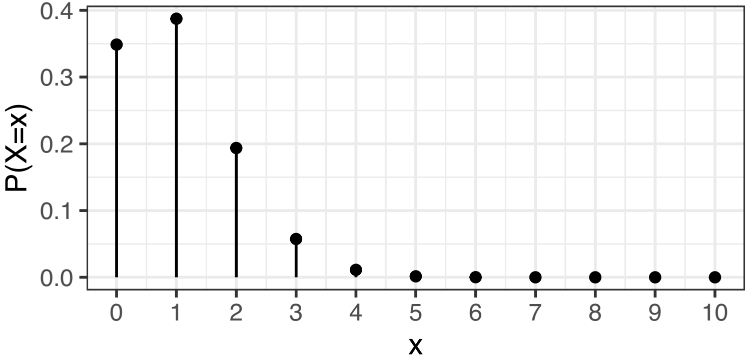


Discrete: Binomial random variables

Example: n = 10, p = 0.1

```
dbinom(x = 0:10, size = 10, prob = 0.1)
```

```
##  [1] 0.3486784401 0.3874204890 0.1937102445 0.0573956280 0.011160261
##  [6] 0.0014880348 0.0001377810 0.0000087480 0.0000003645 0.000000009
## [11] 0.0000000001
```



Discrete: Binomial vs Bernoulli

Actually, binomial can be represented as a sum of Bernoulli random variables.

$$2 = \underset{1}{Y_1} + \underset{0}{Y_2} + \underset{1}{Y_3}$$

$$X = Y_1 + Y_2 + ... + Y_n$$

- X - is Binomial with n trials and with and probability of success p
- $Y_1, Y_2, ..., Y_n$ - are Bernoulli random variables such that $P(Y_i = 1) = p$

Discrete: expected value

Expected value measures the average value of random variable in long term.

$$E(X) = \sum_x x \cdot P(X = x)$$

- number of heads when tossing two coins

$X =$	0	1	2
$P(x) =$	1/4	1/2	1/4

$$E(X) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4}$$

$$= \underline{1}$$

Exercise

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

What is the expectation of the score when rolling a 6-sided die?

$X =$	1	2	3	4	5	6
$P(x) =$	1/6	1/6	1/6	1/6	1/6	1/6

$$E(x) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} =$$

$$= 3.5$$

Exercise

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

What is the expectation of Bernoulli random variable with $p = 0.1$? What is the formula for general p?

$X=$	1	0
$P(x)=$		

$E(x) =$

$X=$	1	0
$P(x)=$		

$E(x) =$

Expected value vs. sample mean

- We have a random variable X

$X =$	0	1	2
$P(X) =$	1/4	1/2	1/4

$$E(X) = 1$$

- We generate a sample of size n using this random variable

$x_1, \ldots, x_n$

$\underbrace{0 \quad 1 \quad 0 \quad 2 \quad 1 \ldots \ldots 0}$

value	0	1	2
frequency	1/4	1/2	1/4

What is the relationship between $E(X)$ and $\bar{x} = \frac{x_1 + \ldots + x_n}{n}$?

$$\frac{x_1 + \ldots + x_n}{n} \simeq E(X)$$

Discrete: variance

Variance and standard deviation measure the spread of the values of a random variable in long term.

$$Var(X) = \sum_x (x - E(X))^2 \cdot P(X = x)$$

$$sd(X) = \sqrt{Var(X)}$$

Alternative formula for variance:

Exercise

$$Var(X) = \sum_x (x - E(X))^2 \cdot P(X = x)$$

What is the variance of the score when rolling a 6-sided die?

$X=$	1	2	3	4	5	6
$P(x)=$	1/6	1/6	1/6	1/6	1/6	1/6

$$E(x) = 3.5$$

$$Var(x) = (1 - 3.5)^2 \cdot \frac{1}{6} + (2 - 3.5)^2 \cdot \frac{1}{6} + \ldots$$

$$\ldots + (6 - 3.5)^2 \cdot \frac{1}{6}$$

Exercise

What is the variance of Bernoulli random variable with $p = 0.1$?
What is the formula for general p?

$X =$	1	0
$P(x) =$		

$E(x) =$

$Var(x) =$

$X =$	1	0
$P(x) =$		

$E(x) =$

$Var(x) =$

Important rules

Expectation

- If X is a random variable and a, b are some numbers then

$$E(a + b \cdot X) = a + b \cdot E(X)$$

- If Y is also a random variable then

$$E(X + Y) = E(X) + E(Y)$$

Variance

- If X is a random variable and a, b are some numbers then

$$Var(a + b \cdot X) = b^2 \cdot Var(X)$$

- If Y is also a random variable and it is independent of X then

$$Var(X + Y) = Var(X) + Var(Y)$$

Exercise

What is the expectation of Binomial random variable with $n = 5$ and $p = 0.1$? What is the formula for general n and p?

$E(x) =$

$E(x) =$

Variance vs. sample variance

$\boxed{0, 2, 1, \ldots}$

- We have a random variable X

  $Var(x)$

- We generate a sample of size n using this random variable $x_1, \ldots, x_n$

What is the relationship between $Var(X)$ and $s_x^2 = \frac{(x_1-\bar{x})^2+\ldots+(x_n-\bar{x})^2}{n-1}$?

Exercise

What is the variance of Binomial random variable with $n = 5$ and $p = 0.1$? What is the formula for general n and p?

$var(X) =$

$var(X) =$

Exercise

If we have 10 random variables $X_1, \ldots, X_{10}$ with $E(X_i) = 10$ and $Var(X_i) = 1$.

What would be the expectation and variance of $X_1 + \ldots + X_{10}$?

$E(X_1 + \ldots + X_{10}) =$

$Var(X_1 + \ldots + X_{10}) =$

What would be the expectation and variance of $\frac{X_1 + \ldots + X_{10}}{10}$?

$E\left(\frac{X_1 + \ldots + X_{10}}{10}\right) =$

$Var\left(\frac{X_1 + \ldots + X_{10}}{10}\right) =$

What are the general formulas for $\bar{X} = \frac{X_1 + \ldots + X_n}{n}$?

$E(\bar{X}) =$

$Var(\bar{X}) =$

TO DO

1. Module 2. Probability: Events and Module 2. Probability: Random Variables
2. Quiz 4 due Monday (February 6) @ 11:59 PM (EST)
3. Practice Problem Set 4