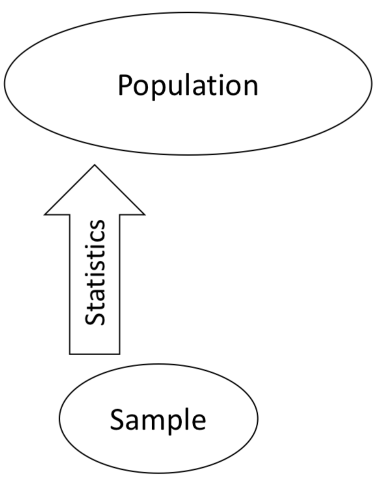
Stat 324 –
Introduction to Statistics for Engineers

LECTURE 3: PROBABILITY (COMPARED TO STATISTICS)

Probability vs **Statistics**



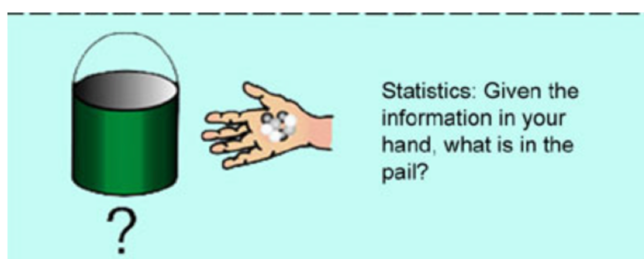


Diagram showing the difference between statistics and probability. (Image by MIT OCW. Based on Gilbert, Norma. *Statistics*. Philadelphia, PA: W. B. Saunders Co., 1976. ISBN: 072164127X.)

Probability vs Statistics

Population

Probability considers information from a population to understand how random samples from that population behave.

Sample

Diagram showing the difference between statistics and probability. (Image by MIT OCW. Based on Gilbert, Norma. *Statistics*. Philadelphia, PA: W. B. Saunders Co., 1976. ISBN: 072164127X.)

Probability vs **Statistics**

e.g. Suppose that Claire doesn't know how many of the ants in her colony of 100 are poisonous. She takes a random sample of 20 ants and finds that 3 are poisonous. Statistics would help her get an __estamate__ for,

"What proportion of the 100 ants are poisonous?"

She would use information from the sample to make a claim/educated guess about __what is true about the pop__.

The estimate is most useful if paired with a measure of __uncertainty__:

e.g 15% give or take 2% is more informative than 15% *give or take* 5%

The estimate __will__ change sample to sample (as may the uncertainty). Claire can reduce the variability of her estimate by __increasing the size of her sample of ants she gets her estimate from

The only way we can know what is true about the colony of ants is to __take a census__

Probability vs Statistics

e.g. Claire's ant farm supports 100 ants. She knows 10 of the 100 ants are poisonous, but they are indistinguishable by sight.

Probability helps us answer:

"If I select one ant at random, what is the chance that it is poisonous?"

Even though we know the percent of ants that are poisonous, we can't predict what will happen in any single sample (sampling variability),

e.g. each ant will either be poisonous or not.

We will typically interpret probability as what will tend to happen over many, many samples.

Some Terminology

- A **random process** is any process that generates data in which the result has some amount of random chance involved.
  - e.g, Select one ant at random from the ant farm and observe whether it is poisonous
  - e.g. coin flip, playing cards, sample of 3 students' heights
  - e.g. randomly assign half of patients to treatment A and half to treatment B and observe weight loss
- An [elementary] **outcome** is a distinct result of a random process.
  - e.g. the outcome of "poisonous" when evaluating an ant
  - e.g. H when flipping a coin, Queen of Hearts when choosing a card from a deck, (64, 61, 77)
  - e.g. observed weight loss of patients on treatments
- A **sample space S** is the set of all possible outcomes of a random process.
  - e.g. {poisonous, nonpoisonous}, {H,T}, {1,2,3,4,5,6}, {52 cards from a deck}, {all possible trios of heights}
- An **event E** is a collection of outcomes (a subset of the sample space)
  - e.g. E1={Club}, E2={2 heights < 60 inches}, E3={mean weight loss of group A< mean weight loss of group B}
  - an event is said to have occurred if at least one of the outcomes in this collection occurs; not all outcomes in the collection need occur for the event to be said to occur.

Two Interpretations of Probability

(1) Subjective (Bayesian)

- Probability of an event is a _degree of belief_ about the changes that the event will occur.
- Valuable in random processes that can never be _exactly repeated_.
  e.g. probability of Iceland winning the 2018 World Cup
- Use prior information to inform decision
  e.g. Toddler chooses not to touch hot stove since mom told me it was hot
- Consistent with how humans use info in informal, every-day decision making, but mathematics is _advanced_.
- With sufficient information, Bayesian and frequentist statistics _agree well_.

Two Interpretations of Probability

(2) Long-Run Frequency (frequentist)

- If a random process is repeated infinitely many times, the probability of an event is the _proportion of times_ that the event occurs.
- Valuable in situations where it would be theoretically possible to repeat a random process many times
  - e.g. drawing 1 ant from the ant farm many times (and determining if it is poisonous), or flipping a coin (looking at whether heads landed up)
- If we do not know P(A) exactly, it can be estimated accurately by _repeating the experiment many times_

Applying Terminology

Ex: An engineer has a box containing 4 bolts and another box containing 4 nuts. The diameters of the bolts are 4,6,8, and 10 mm, and the diameters of the nuts are 6,10, 12, and 14 mm. One bolt and one nut are chosen. Choose 1 bolt & 1 nut

1. How large is the sample space S? (4,6) 6 8 10

bolts · nuts (4,10)

4 4 = 16 (4,12)

(4,14)

Specify the subsets corresponding to the events A and B

1. Let A be the event that the bolt diameter is less than 8.

(4,6) (6,6)
(4,10) (6,10)
(4,12) (6,12)
(4,14) (6,14)

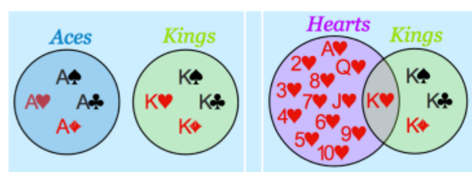
2. Let B be the event that the bolt and the nut have the same diameter.

(6,6) (10,10)

Some Terminology

- The **union** of two events A and B, denoted $A \cup B$, "A or B" is the set of outcomes that belong to A, to B, or __to both__
- The __intersection__ of two events A and B, denoted $A \cap B$, "A and B" is the set of outcomes that belong to both A and B.
- The **complement** of an event A denoted A^c, "not A" is the set of outcomes that are not in A.
- Events A and B are **mutually exclusive** if they have __no outcomes in common__.

e.g. chance of Ace and King are mutually exclusive



Aces and Kings are **Mutually Exclusive** (can't be both)

Hearts and Kings are **not** Mutually Exclusive (can be both)

Properties of Probability (in the context of a finite sample space)

- The probability of an **event** is the __Sum______ of the probabilities of the [elementary] __outcomes_______ in that event.

  e.g: Let A be the event that the bolt diameter is less than 8.

  - P(A)=P(4,6)+ P(4,10)+P(4,12)... P(6, 14)

  - Or P(A)= $\dfrac{\text{number of favorable outcomes for our event}}{\text{total \# of outcomes}}$

(Depends on assumption that all [elementary] outcomes are _equally likely__)

- The sum of all outcomes in the sample space is __1____.

Properties of Probability (in the context of a finite sample space)

- The probability of an event is between 0 and 1 ($0 \leq P(E) \leq 1$).
  - Where a probability of 0 indicates an event that will never occur.

    e.g Probability that I will get a bolt with diameter 2.

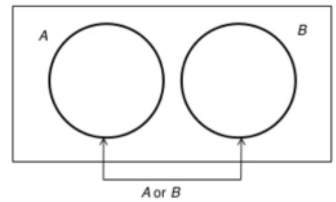
  - Where a probability of 1 indicates an event that will always occur.

- The probability that an event does not occur is 1 minus the probability that it does occur. $P(B^C) = 1 - P(B)$,

  e.g. probability I do not get a bolt and nut that have the same diameter.

  P(B)=2/16, so $P(B^C)$=1-2/16=14/16

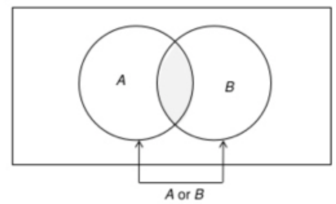
Properties of Probability (in the context of a finite sample space)



Probability of "A or B"

- If A and B are mutually exclusive events, then

  $P(A \cup B) =$ P(A) + P(B)



- If A and B are not mutually exclusive events, then

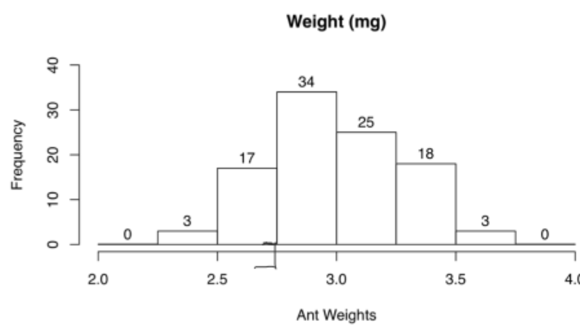
  $P(A \cup B) =$ P(A) + P(B) − P(A∩B)

  includes middle — includes middle — subtract one middle

Probability of "A or B" is the ___sum___ of the event probabilities minus the probability of their overlap (and when events are mutually exclusive, there is ___overlap___)

Applying Basic Probability

Ex 1: Consider the 100 ants again. Here are their weights.
(using the convention that values on borderline are included in upper bin)
If one ant is selected at random, what is the probability…



a. Its weight is less than 2.75mg?

$20/100 = 0.20$

b. Its weight is at least 2.75 mg?

$80/100 = 0.80$

c. Its weight is between 2.5 and 3.5?

$94/100 = 0.94$

c. Its weight is at least 3.5 or less than 2.5 mg?

$6/100 = 0.06$

*Check against relative frequency with repeated sampling.

Applying Basic Probability

Ex 2: Fifteen persons reporting to a Red Cross center one day are types for blood, and the following counts are found.

Blood Group	O	A	B	AB	Total
No. of Persons	3	5	6	1	15

If a person is randomly selected, what is the probability that this person's blood group is:

a. AB? 1/15 b. either A or B? 11/15 c. Not O? 12/15 d. contains B? 7/15

Ex 3: An extrusion die used to produce aluminum dowel is given specifications for the length and diameter of dowel to be produced. The 1000 dowels produced on a given day are classified as follows.

Diameter

Length	Too Thin	OK	Too Thick	
Too Short	10	3	5 double count	18
OK	38	900	4	
Too Long	2	25	13	

22

A dowel is sampled at random on that day. What is the probability..

a. That it is too short?

18/1000

b. That it is too short or too thick?

35/1000

$P(A \cap B)$

c. That it is too short and too thick?

5/1000

Extending Basic Probability

llows.

		Diameter			
		Too Thin	OK	Too Thick	
Length	Too Short	10	3	5	
	OK	38	900	4	942
	Too Long	2	25	13	
			928		

Ex 3 continued:

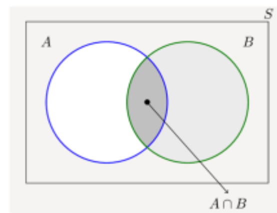
d. What is the probability that a randomly chosen rod has a diameter that meets specifications? **P(Diameter OK)** $= 928/1000$

This is called an unconditional probability since it is based on full sample space.

e. What is the probability that a randomly chosen rod meets the diameter specification **given** that it meets the length specification? **P(Diameter OK| Length OK)** $= \frac{900}{942}$

This is called a conditional probability since it is only based on a part of the sample space.

P(Diameter OK | Length OK)=$\frac{P(Diameter\ OK\ and\ Length\ OK)}{Length\ OK}$ $\quad P(A\,|B) = \frac{P(A \cap B)}{P(B)}$



f. Is P(Diameter OK)=P(Diameter OK|Length OK)? no

$928/1000 \neq \frac{900}{942}$

dependence between Diameter OK and length OK

Properties of Probability- Independence

- Two events are said to be **independent** if the occurrence of one does not change the probability of the other.
  - If P(A) and P(B) >0, P(B|A)=P(B) or, equivalently $P(A|B) = P(A)$
  - Otherwise, they are **dependent.**

Length		Diameter		
		Too Thin	OK	Too Thick
	Too Short	10	3	5
	OK	38	900	4
	Too Long	2	25	13

e.g. If an aluminum dowel is randomly sampled from the day,

Compare the unconditional probability of the dowel being too long and the conditional probability of it being too long given it is too thin.

What does that tell us about the events "Dowel is too thin" and "Dowel is too long" ?

$P(\text{Too thin}) == P(\text{Too thin} \mid \text{too long})$

$50/1000 \qquad \frac{2}{40}$

$5\% = 5\%$

Too thin + too long is independent in this set

$40/1000 = 2/50$

$4\% \quad 4\%$

equal too long + to thin independent in this sample

ways to check independence

Properties of Probability – Multiplication Rule

We can rearrange our equation $P(A|B) = \frac{P(A \cap B)}{P(B)}$ to obtain an equation for calculating the probability of A and B:

- In General: $P(A \cap B) = P(B) * P(A|B)$
- And if A and B are independent: $P(A \cap B) = P(B) * P(A)$

Length		Diameter			
		Too Thin	OK	Too Thick	Total
	Too Short	10	3	5	18
	OK	38	900	4	942
	Too Long	2	25	13	40
	Total	50	928	22	1000

Check that P(Too Thin)*P(Too Long)=P(Too Thin and Too Long)

$\frac{2\cancel{000}}{1000\cancel{000}} = \frac{50}{1000} * \frac{40}{1000} == \frac{2}{1000}$ ✓

- Ex 4: It is known that 5% of the cars and 10% of the light trucks produced by a certain manufacturer require warranty service. If someone purchases both a car and a light truck from this manufacturer, then assuming the vehicles function independently, what is the probability that they both require warranty service? Is independence a "good" assumption in this scenario?

$P(\text{car and truck require service}) = P(C \text{ and } T)$ C = car requires T = truck requires

$= P(C) * P(T|C)$
$= P(C) * P(T)$ } assumed independence
$= 0.05 * 0.1 = 0.005$

Reasonableness of Independence Assumption

In most cases, the best way to determine whether independence of events is a good assumption is through understanding the process that produces the events.

1. A die is rolled. Is it reasonable to believe that the outcome of the second roll is independent of the first roll? Yes
2. A card is chosen. Is it reasonable to believe that the outcome of the second card chosen is independent of the first card? no, depends if card is put back
3. A chemical reaction is run twice,
   1. Using different equipment each time - yes
   2. Using the same equipment each time – no
4. The items in a simple random sample may be treated as independent (even though this is sampling without replacement), unless the population is finite and the sample comprises more than about 5-10% of the population (you will look at this in discussion next week).

Probability Practice

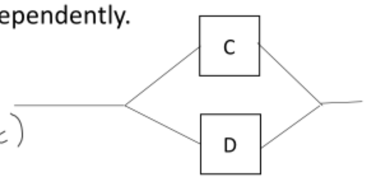
Ex 5: A system contains two components, A and B, connected in series as shown in the following diagram. The system will function only if both components function. The probability that A functions is P(A)=0.98, and the probability that B functions is P(B)=0.95. Assume that A and B function independently. Find the probability that the system functions.

P(system functions) = P(A and B function)

= P(A functions) * P(B functions | A functions)

= P(A function) * P(B function) = 0.98 * 0.95 = 0.931

Ex 6: A system contains two components, C and D, connected in parallel as shown in the following diagram. The system will function if either C or D functions. The probability that C functions is 0.90 and the probability that D functions is 0.85. Assume that C and D function independently. Find the probability that the system functions.



added 2 overlap

subtract 1

P(system func) = P(C func or D func)

= P(C func) + P(D func) − P(C and D func)

= 0.90 + 0.85 − (0.90 * 0.85)

= 0.985

Probability Practice

Ex 7: A geneticist is studying two genes. Each gene can be either dominant or recessive. A sample of 100 individuals is categorized in the table below.

		Gene 2		
		Dominant	Recessive	
Gene 1	Dominant	56	24	80
	Recessive	14	6	20
		70	30	100

a. What is the probability that in a randomly sampled Individual, Gene 1 is dominant?

80/100

b. What is the probability that in a randomly sampled Individual, Gene 2 is dominant?

70/100

c. Given that Gene 1 is dominant, what is the probability that Gene 2 is dominant?

P(Gene 2 Dom | Gene 1 Dom) = 56/80 = 0.70

d. The genes are said to be in linkage equilibrium if the event that Gene 1 is dominant is independent of the event that Gene 2 is dominant. Are these genes in linkage equilibrium?

is Gene 1 dominance independent of Gene 2 dominance?

P(Gene 2 Dom | Gene 1 Dom) == P(Gene 2 dom)

0.7 == 0.7 ✓ independent

Probability Practice

Ex 8: Consider Rolling 6-sided die. Find the probability for these events:

a. Observing an odd number with one roll of the die. $\frac{3}{6}$ odd $= 1/2$

b. Observing a number greater than 4 on one roll of the die. $\frac{2}{6} = 1/3$

c. Observing a number that is even and a number that is greater than 2 on one roll of the die.

$\frac{2}{6}$

X

d. Rolling exactly 1 six in 4 rolls of the die.

6XXX
X6XX
XX6X
XXX6

$P(6XXX) = P(6) * P(X) * P(X) * P(X)$
$= 1/6 * 5/6 * 5/6 * 5/6$

$= 4 * \frac{1}{6} * (5/6)^3$

e. Rolling at least 1 six in 4 rolls of the die.

$1 - P(\text{no } 6\text{'s})$ $1 - (5/6)^4$