Probability

Session 1

Module 1 Probability & Statistical Inference

The Summer Institutes
DEPARTMENT OF BIOSTATISTICS
SCHOOL OF PUBLIC HEALTH
UNIVERSITY *of* WASHINGTON

Liber de ludo aleae
(Book on Games of Chance)
by Gerolamo Cardano

- Written 1526 (published 1663)
- First systematic treatment of probability

NOTHING IN LIFE IS CERTAIN. IN EVERYTHING WE DO, WE GAUGE THE CHANCES OF SUCCESSFUL OUTCOMES, FROM BUSINESS TO MEDICINE TO THE WEATHER. BUT FOR MOST OF HUMAN HISTORY, ***PROBABILITY***, THE FORMAL STUDY OF THE LAWS OF CHANCE, WAS USED FOR ONLY ONE THING: ***GAMBLING***.

Image credit: https://newsinhealth.nih.gov/2011/05/when-stakes-turn-toxic
https://www.sciencephoto.com/media/224088/view/coloured-portrait-of-the-italian-girolamo-cardano

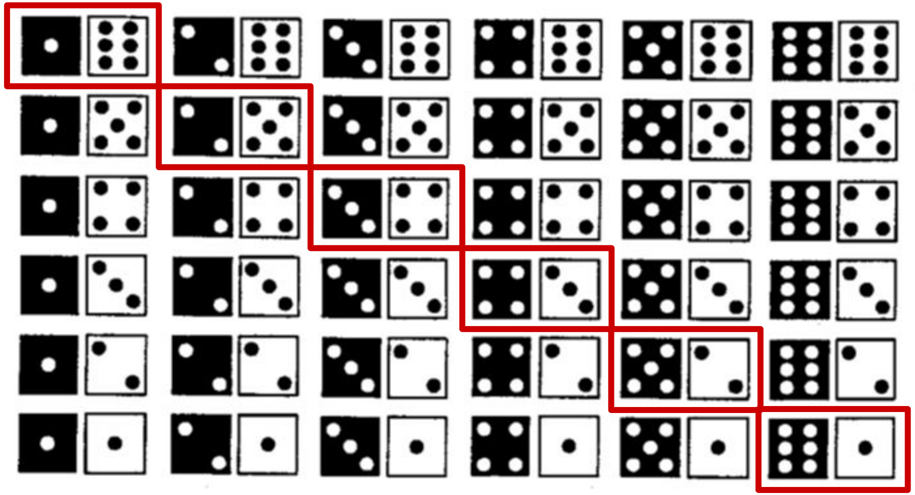
1 Probability

A measure of uncertainty associated with the occurrence of events or outcomes.

APPROACH 1

⭐ **Classical**: P(E) = m/N

If an event can occur in N <u>mutually exclusive,</u> <u>equally likely</u> ways, and if **m** of these possess characteristic **E**, then the probability of **E** is equal to **m/N**



Example
What is the probability of rolling a total of 7 on two dice?

E = two dice sum to 7
N = 36
m = 6
P(E) = m / N = 6/36 = 1/6

Session 1
PROBABILITY AND INFERENTIAL STATISTICS
UNIVERSITY *of* WASHINGTON

1 Probability

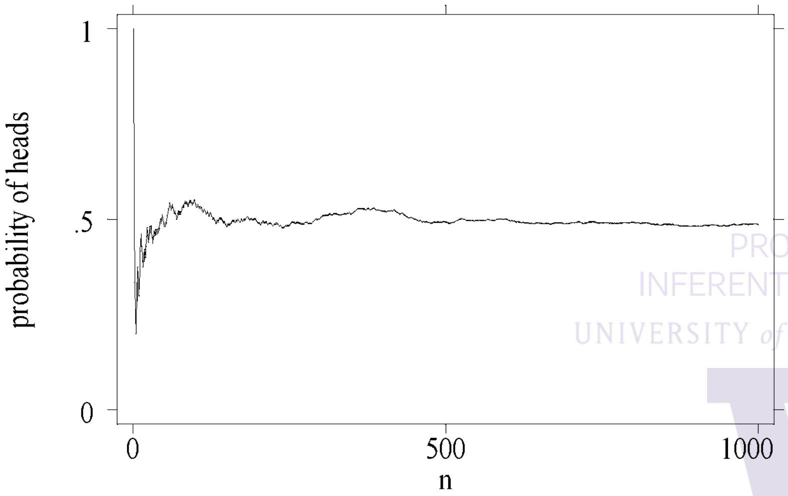
A measure of uncertainty associated with the occurrence of events or outcomes.

Example
Around 1900, the English statistician Karl Pearson heroically tossed a coin 24,000 times and recorded 12,012 heads, giving a proportion of 0.5005.

APPROACH 2

Relative Frequency: Pr(E) ≈ m/n

If a process or an experiment is repeated a large number of times **N**, and if the characteristic **E**, occurs m times, then the relative frequency, **m/N**, of **E** will be approximately equal to the probability of **E**.



Session 1
PROBABILITY AND INFERENTIAL STATISTICS
UNIVERSITY of WASHINGTON

1 Probability

A measure of uncertainty associated with the occurrence of events or outcomes.

APPROACH 3

Personal Probability

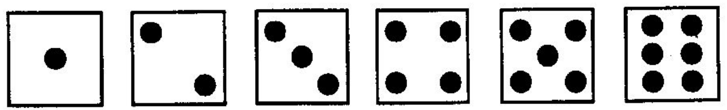
What is the probability of life on Mars?

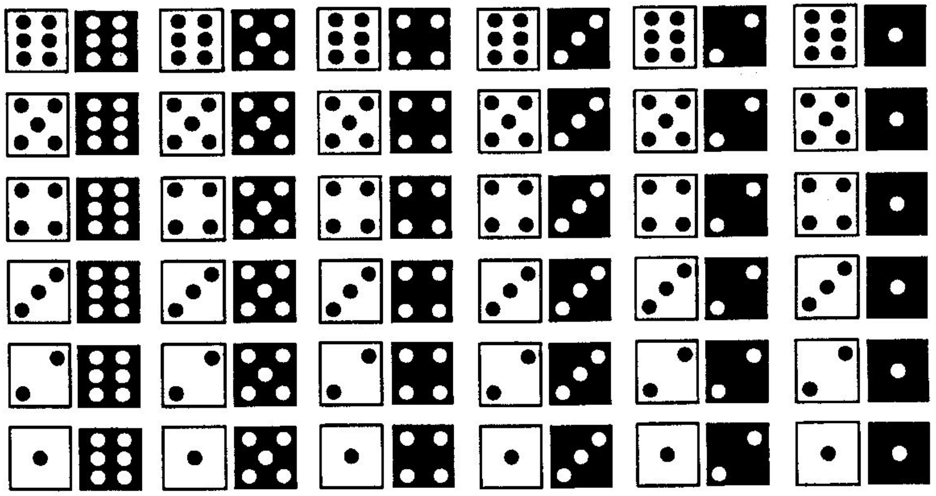
DEFINITION

Sample Space

The **sample space** consists of the possible outcomes of an experiment. An event is an outcome or set of outcomes.

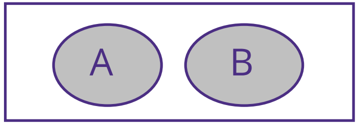
For a coin flip the sample space is (H,T).





Basic Properties of Probability

- Two events A and B are said to be **mutually exclusive (disjoint)** if only one or the other, but not both, can occur in a particular experiment.



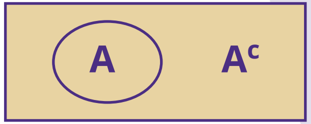
- Given an experiment with n mutually exclusive events, E_1, E_2, …., E_n, the **probability of any event is non-negative and less than 1**:

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

- The **sum** of the probabilities of an exhaustive collection (i.e., at least one must occur) of mutually exclusive outcomes is **1**:

$$\sum_{i=1}^{n} P(E_i) = P(E_1) + P(E_2) + \cdots + P(E_n) = 1$$

- The probability of all events other than an event A is denoted by $P(A^c)$. Note that $P(A^c) = 1 - P(A)$



Break #1

Pause the video,
take a break, stretch,
then review relevant exercises from worksheet.

Afterwards, continue on!

Image Credit: indg0.com

Notation for Joint Probabilities

If A and B are any two events then we write

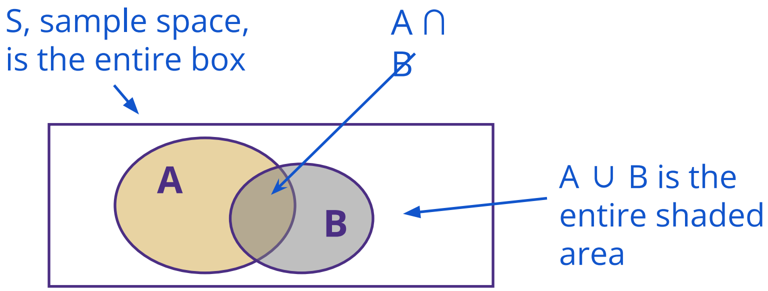
$$P(A \text{ or } B) \text{ or } P(A \cup B)$$

to indicate the probability that event A or event B (or both) occurred.

If A and B are any two events then we write

$$P(A \text{ and } B) \text{ or } P(AB) \text{ or } P(A \cap B)$$

to indicate the probability that both A and B occurred.

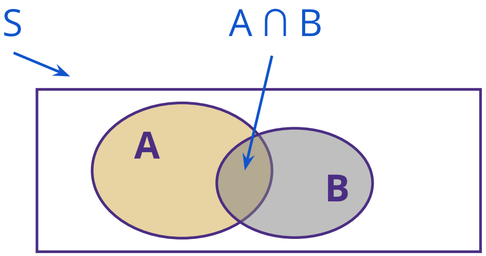


Notation for Joint Probabilities

If A and B are any two events then we write the conditional probability

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

to indicate the probability of A among the subset of cases in which B is known to have occurred.

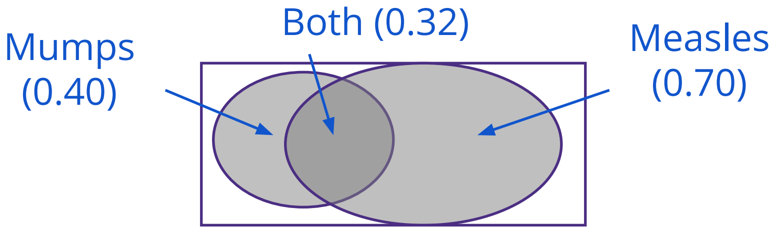


Addition Rule

If two events A and B are not mutually exclusive, then the probability that event A or event B occurs is:

$$P(A \text{ or } B) = P(A) + P(B) - P(AB)$$

Example Of the students at Anytown High school, 40% have had the mumps, 70% have had measles and 32% have had both. What is the probability that a randomly chosen student has had at least one of the above diseases?



General Probability Rules

Independence

Two events A and B are said to be **independent** if and only if

$$P(A|B) = P(A) \text{ or } P(B|A) = P(B) \text{ or } P(AB) = P(A)P(B)$$

(Note: If any one holds then all three hold)

Example Of the students at Anytown High school, 40% have had the mumps, 70% have had measles and 32% have had both.

Are the two events independent?

No, because P(mumps and measles) = 0.32 while P(mumps) P(measles) = 0.28

Multiplication Rule

Two events A and B are "independent" (probability of one does not depend on whether the other occurred) if

$$P(AB) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

Example Of the students at Anytown High school, 40% have had the mumps, 70% have had measles. The probability of having measles given you have mumps is 80%.

What's the probability of having both?

P(both) = P(measles | mumps) P(mumps) = 0.80 * 0.40 = 0.32

Break #2

**Pause the video,
take a break, stretch,
then review relevant exercises
from worksheet.**

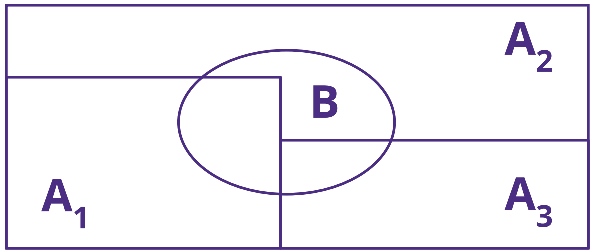
Afterwards, continue on!

Image Credit: indg0.com

Total Probability

If A_1, ..., A_n are mutually exclusive, exhaustive events, then

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i)$$



Bayes' Rule

Bayes' rule combines multiplication rule with total probability rule

$$P(A_j \mid B) = \frac{A_j \cap B}{P(B)} = \frac{P(B|A_j)\,P(A_j)}{P(B)} = \frac{P(B|A_j)\,P(A_j)}{\sum_{i=1}^{n} P(B \mid A_i)\,P(A_i)}$$

We will only apply this to the situation where A and B have two levels each, say, A and A^c, B and B^c. The formula becomes

$$P(A \mid B) = \frac{P(B|A)\,P(A)}{P(B \mid A)\,P(A) + P(B \mid A^c)\,P(A^c)}$$

EXAMPLE

An Application of Bayes' Rule

Screening

Suppose we have a random sample of 1100 people from a population…

A = disease pos.

B = test pos.

Prevalence = P(A) = 100/1100 = 0.091

Sensitivity = P(B | A) = 90/100 = 0.9

Specificity = $P(B^c \mid A^c)$ = 970/1000 = 0.97

PVP = P(A | B) = 90/120 = 0.75
(predictive value of a positive test)

PVN = $P(A^c \mid B^c)$ = 970/980 = 0.99
(predictive value of a negative test)

Test Result \ Disease Status	Positive	Negative	
Positive	90	30	120
Negative	10	970	980
	100	1000	1100