Topic 2: Introduction to Probability

Optional Reading: Chapter 4

Xiner Zhou

Department of Statistics

University of California, Davis

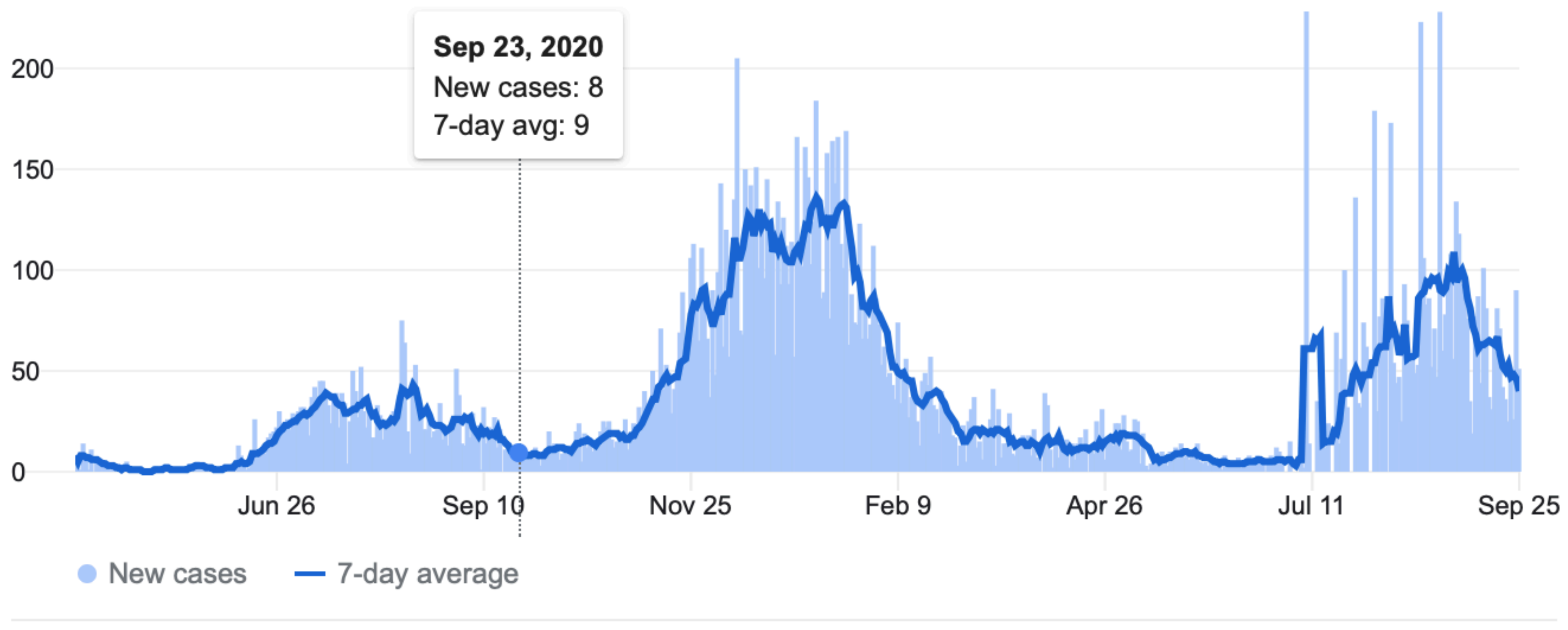
Part I: Probability

- Basic definitions in probability
- Counting rules for calculating probabilities
- Event Relations and Probability Rules

Why Study Probability?

Math is the logic of certainty, Probability is the logic of uncertainty.

casual conversation: “Luck, Coincidence, Random, Risk, Doubt, Chance……” — sense of Uncertainty



Despite its ubiquity, “probability” can be deeply counterintuitive.

Just reply on intuition, run serious risk of making inaccurate predictions or overconfident decisions.

That’s why probability is extremely useful in a wide variety of fields:

- Foundation for Statistics:
    - Powerful methods of using data to learn about the world:
        - Explaining variation: why some people live longer, why some students have higher performance
        - separating signal from noise: blood tests for the same person in the morning and afternoon
        - Modeling complex phenomenon: how to build a model that predict the evolution of COVID-19
- Applications in:
    - Physics, Biology, Computer Science, Gambling, Finance, Economics,
    - Political Science, Medicine, Astronomy….

An experiment is the process by which an observation (or measurement) is obtained.

- It is basically a method of data collection.
- Data are obtained by observing either uncontrolled events in nature or by observing events in controlled situations. We use the term experiment to describe either method of data collection.
- Examples:
    - Recording a test grade
    - Measuring daily rainfall
    - Interviewing a householder to obtain his or her opinion on ...
    - Tossing a coin and observing the face that appears

The sample space S of an experiment is: the collection of all possible outcomes of the experiment.

- Example:
  - Flip a coin S={H,T}

A simple event (also called a sample point) is the outcome that is observed on a single repetition of the experiment.

- Example:
  - Flip a coin: H, T both are a simple event or a sample point

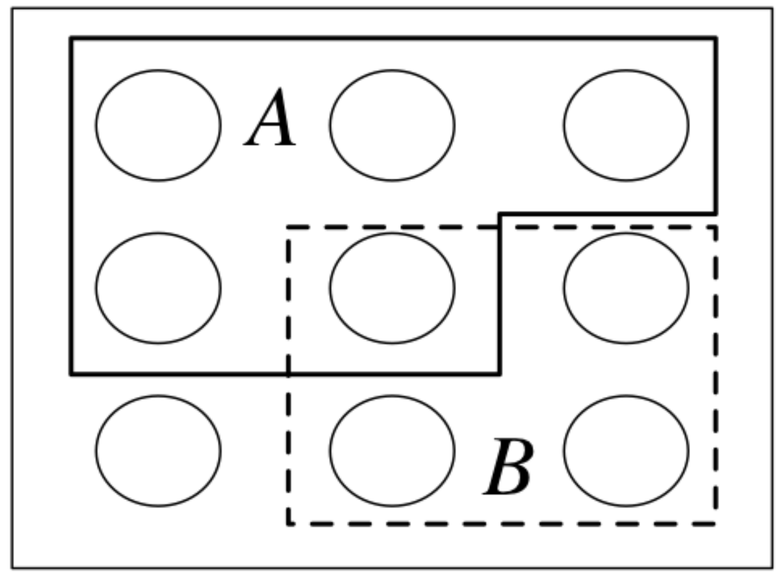
An event A is a collection of simple events.

- Example:
  - Getting a head: A={H}

We say that event A occurred or happened if: the actual outcome is in A

Historically, the earliest definition of probability of an event A was:

$$\frac{\text{number of ways the event could happen}}{\text{Total number of possible outcomes for the experiment}}$$



- Reply on assumptions that
  - every outcome is equally likely to happen:
    - when is/isn't appropriate?
    - You want to ask a friend for help, is everyone you know equally likely to be called?
  - Only have finite number of possible outcomes of an experiment:
    - when is/isn't appropriate?
    - Throw a coin into a fountain, probability inside a small circle
- Nevertheless, it's important to understand the first definition of probability and its many applications

Let A be an event for an experiment, the probability of A is:

$$P(A) = \frac{|A|}{|S|}$$

|A|: number of simple events in A

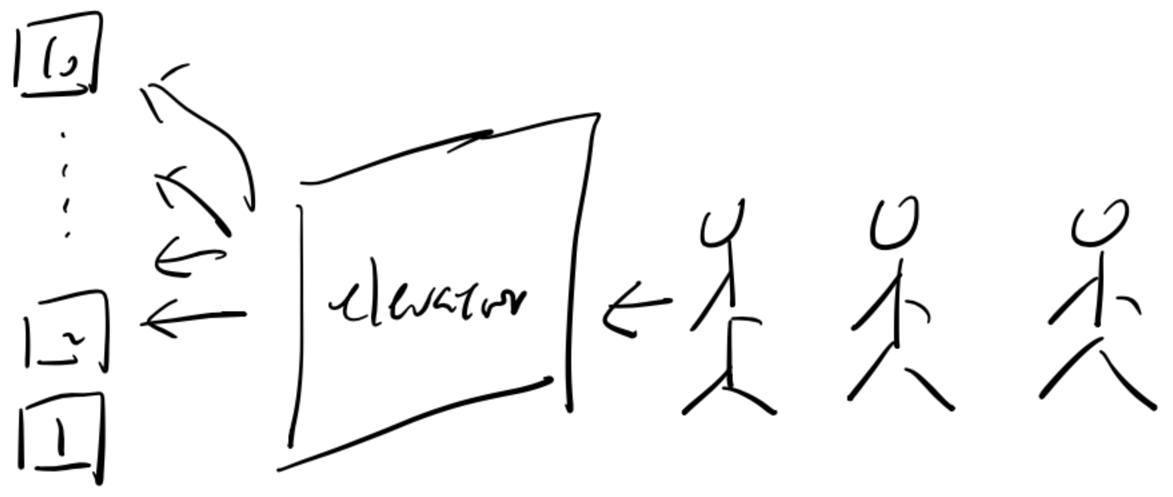
|S|: number of simple events in S

Elevator: people are equal likely to go to floor 2-10

Calculating probability of an event = Counting!

Elevator: 3 people are equal likely to go floor 2-10

What is the probability that buttons for 3 consecutive floors are pressed?



S={222,223,224,225,..........}

A={234,324,423,432,...........}

Wait... Can you imagine and count the sample space and the event A?

How to Count? When the sample space or the event is not so simple

Counting Rules for Probability

Multiplication rule (the fundamental counting rule)

If an experiment is performed in k stages,

with n_1 ways to accomplish the first stage,

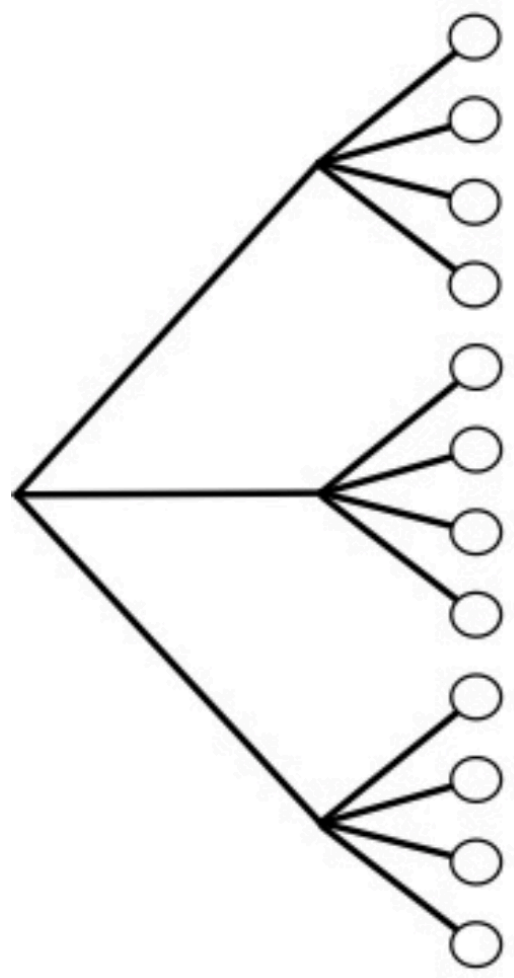
n_2 ways to accomplish the second stage, ...,

and n_k ways to accomplish the kth stage,

then the number of ways to accomplish the experiment is

$n_1 n_2 n_3 \cdots n_k$

Why:



If an experiment can be performed in 2 stages,

There is 3 ways to do the first stage, and 3 ways to do the second stage,

Then whichever you choose to do the first stage,

there are 3 multiple ways to do the second stage,

So, in total, if we count the number of circles in the end which represents unique ways to do the experiment

3+3+3=3 * 3 = n1 * n2

Note: The multiple rules says that an experiment can be break down into a sequence of sub-tasks,

It seems to suggest that the sub-tasks are taken in a specific order:

Sub-task 1-> sub-task 2 -> sub-task 3 ……

But: There is no requirement that sub-tasks are taken in a particular time order!

It is often easier to think about experiments in chronological order,

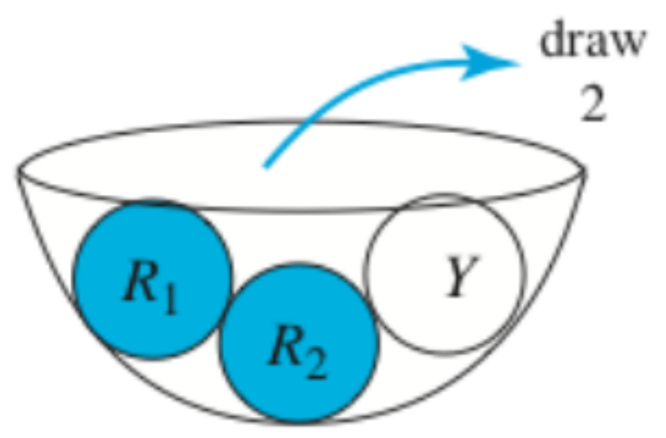
just to help you clarify mental barriers and think in a logical way.

A truck driver can take three routes from city Ato city B, four from city B to city C, and three from city C to city D.

If, when traveling from A to D, the driver must drive from A to B to C to D, how many possible A-to- D routes are available?

$$A \longrightarrow B \longrightarrow C \longrightarrow D$$

A candy dish contains one yellow and two red candies. Two candies are selected one at a time from the dish, and their colors are recorded. How many simple events are in the sample space S?



How many simple events are in the sample space when three coins are tossed?

You own 4 pairs of jeans, 12 clean T-shirts, and 4 wearable pairs of sneakers.

How many outfits (jeans, T-shirt, and sneakers) can you create?

Your family vacation involves a cross-country air flight, a rental car, and a hotel stay in Boston. If you can choose from four major air carriers, five car rental agencies, and three major hotel chains, how many options are available for your vacation accommodations?

There are 3 extremely common situations that cover most problems,

and can be dealt with by multiplication rule.

So it's a good idea to summarize them so that don't have to redo the same mental exercise every time.

We have: n distinct objects

Making k times of choices from them

1. When order matters & Sampling with replacement
    - Order matters: object 1 -> object 2 is different from object 2 -> object 1
    - Sampling with replacement: whichever object is taken, it will be returned to the mix so that it can be chosen again and again

How many simple events are in the sample space when three coins are tossed?

Solution:

subtask 1: flip the 1st coin in n1=2

subtask 2: flip the 1st coin in n2=2

subtask 3: flip the 1st coin in n3=2

=> the total number of possible results we can get from flipping 3 coins is n1*n2*n3=2*2*2=8

We have seen this example of coin flip before:

There are 2 objects : Head or Tail to choose from

We make 3 times of choices

We care about the order we see heads/tails: order matters

No matter we saw tail/head in previous flip, we can still see tail/head in next flip: sampling with replacement

We have: n distinct objects

Making k times of choices from them

1. When order matters & Sampling with replacement

- Order matters: object 1 -> object 2 is different from object 2 -> object 1
- Sampling with replacement: whichever object is taken, it will be returned to the mix so that it can be chosen again and again

=> There are $n^k = n \times \ldots \times n$ possible outcomes

We have: n distinct objects

Making k times of choices from them

2. When order matters & Sampling without replacement

- Order matters: object 1 -> object 2 is different from object 2 -> object 1
- Sampling without replacement: whichever object is taken, it will not be returned to the mix so that it can’t be chosen again

Subtask1: n1=n ways

Subtask2: n2=n-1 ways because 1 way has been chosen

Subtask 3:n3=n-2 ways

....

Subtask k: n-(k-1)=n-k+1 ways because k-1 ways has been chosen

=> There are $n \times (n-1) \times \ldots \times (n-k+1)$ possible outcomes

$P_k^n = n \times (n-1) \times \ldots \times (n-k+1)$ is called Permutation

Read as "n permute k"

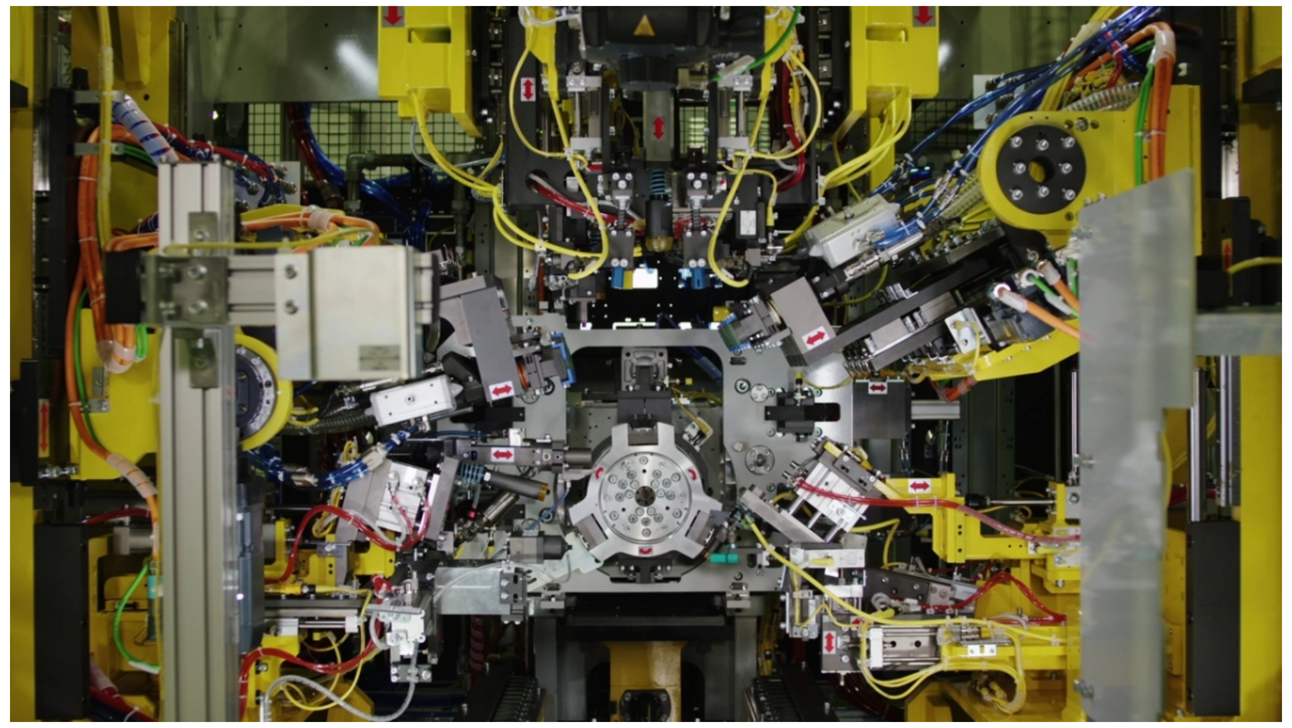
$P_k^n = \dfrac{k!}{(n-k)!}$ where $n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$ and $0! = 1$

Three lottery tickets are drawn from a total of 50 .

If the tickets will be distributed to each of three employees in the order in which they are drawn, the order will be important. How many simple events are associated with the experiment?

A piece of equipment is composed of five parts that can be assembled in any order.

A test is to be conducted to determine the time necessary for each order of assembly. If each order is to be tested once, how many tests must be conducted?



A businessman in New York is preparing an itinerary for a visit to six major cities. The distance traveled, and hence the cost of the trip, will depend on the order in which he plans his route.

How many different itineraries (and trip costs) are possible?

We have: n distinct objects

Making k times of choices from them

3. When order DOES NOT matters & Sampling without replacement

- Order does not matters: object 1 -> object 2 = object 2 -> object 1
- Sampling without replacement: whichever object is taken, it will be not returned so that it can't be chosen again

Compare with previous situation

2. When order matters & Sampling without replacement

Just one difference: whether order matters, (1,2) ?= (2,1)

$$\frac{k!}{(n-k)!k!} := C_k^n$$

The number of k-combinations from n objects, also called binomial coefficient

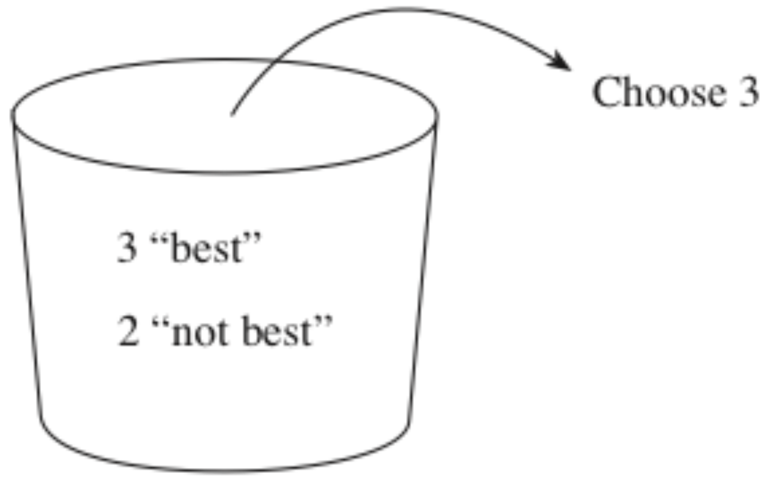
Read as "n choose k"

A printed circuit board may be purchased from five suppliers.

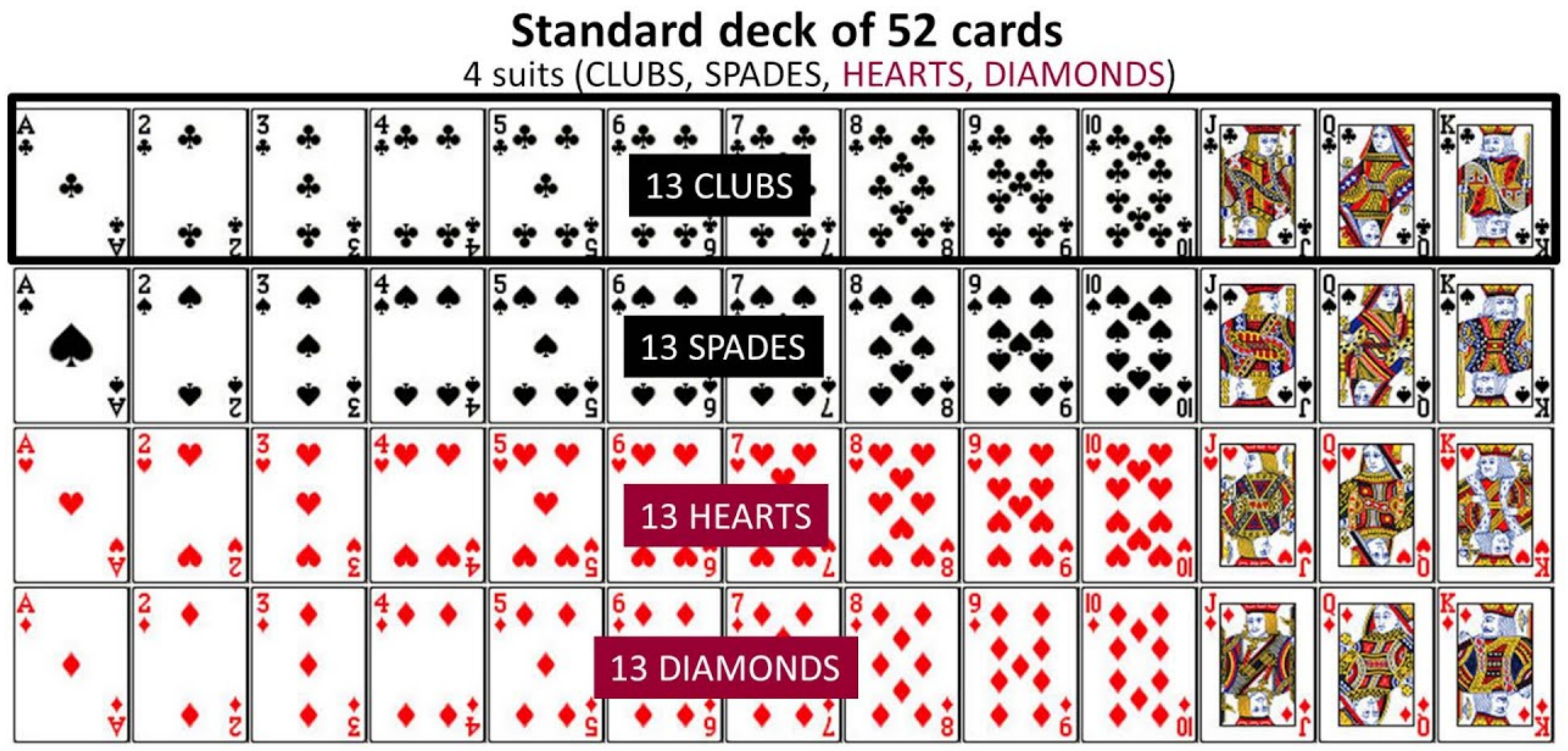
In how many ways can three suppliers be chosen from the five?

Five manufacturers produce a certain electronic device, whose quality varies from manufacturer to manufacturer.

If you were to select three manufacturers at random, what is the chance that the selection would contain exactly two of the best three?



(Full House in Poker)



A 5-card hand is dealt from a standard, well-shuffled 52-card deck.

The hand is called a full house in poker if:

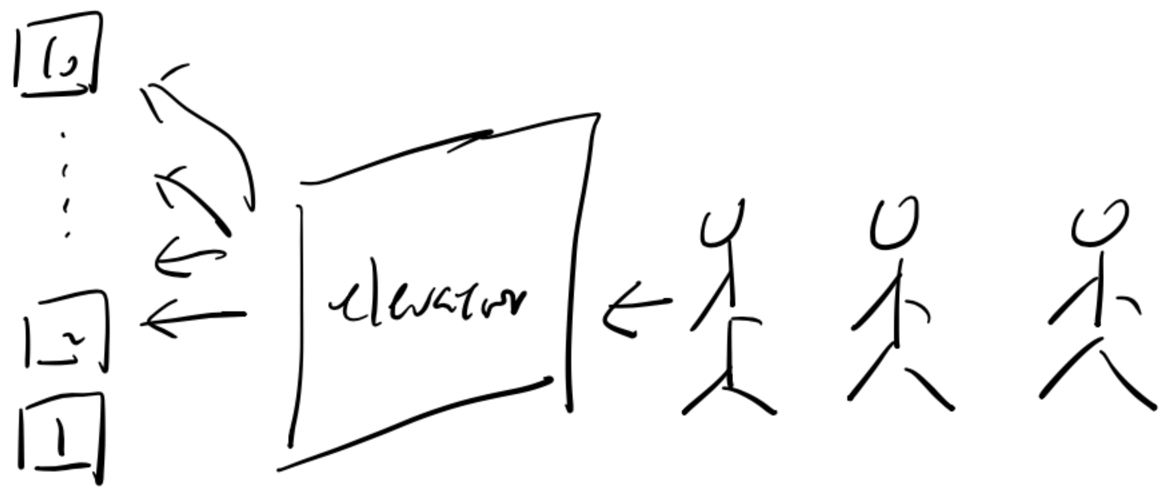
It consists of 3 cards of same rank + 2 cards of another rank

i.e. 3 7's and 2 10's in any order (order by which we draw cards does not matter)

What is the probability of a full house?

Elevator: 3 people are equal likely to go floor 2-10

What is the probability that buttons for 3 consecutive floors are pressed?



S={222,223,224,225,..........}

A={234,324,423,432,...........}

Wait… Can you imagine and count the sample space and the event A?

How to Count? When the sample space or the event is not so simple

General Definition of Probability and Properties

We've seen probabilities if naive definition of probability applies:

1. Finite sample space
2. Equal likely outcomes

To generalize probability to account for all situations.

General definition of Probability

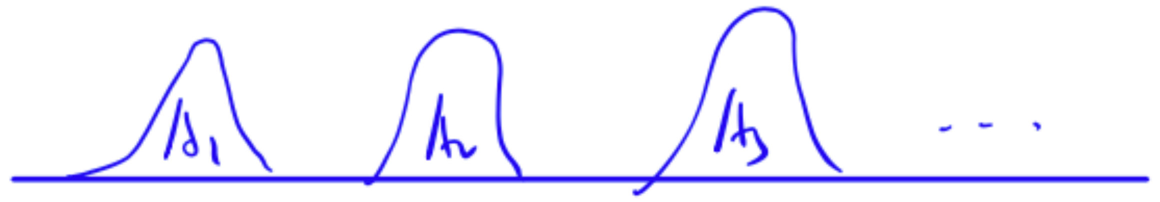
A probability P assigns an event A contained in the sample space S, a number between 0 and 1, and probability P must satisfy the following axioms:

1. $P(\varnothing) = 0, P(S) = 1$

mass of no sand = 0

total mass of the whole pile of sand = 1

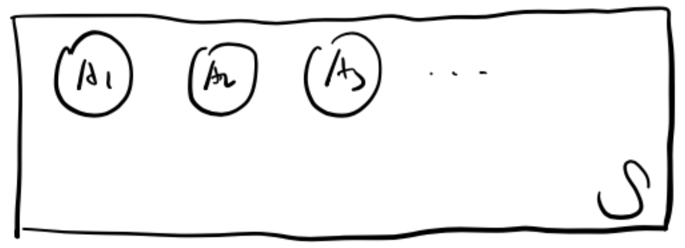
2. If $A_1, A_2, \ldots$ are disjoint events, then $P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P\left(A_j\right)$



total mass of non-overlapping piles of sand

= sum of individual mass

- Disjoint means: $A_i \cap A_j = \varnothing$ for $i \neq j$

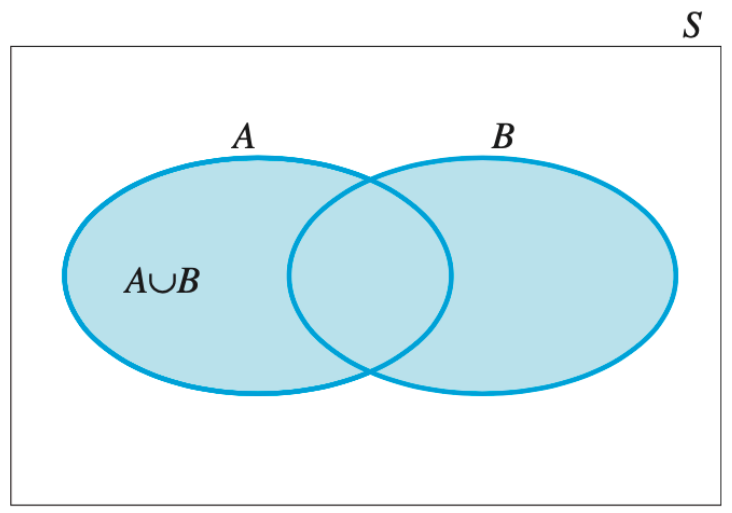


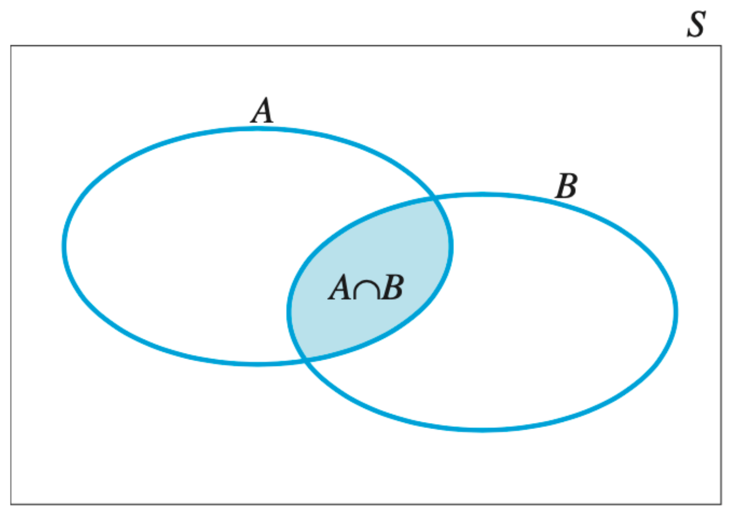
Event Relations and Probability Rules

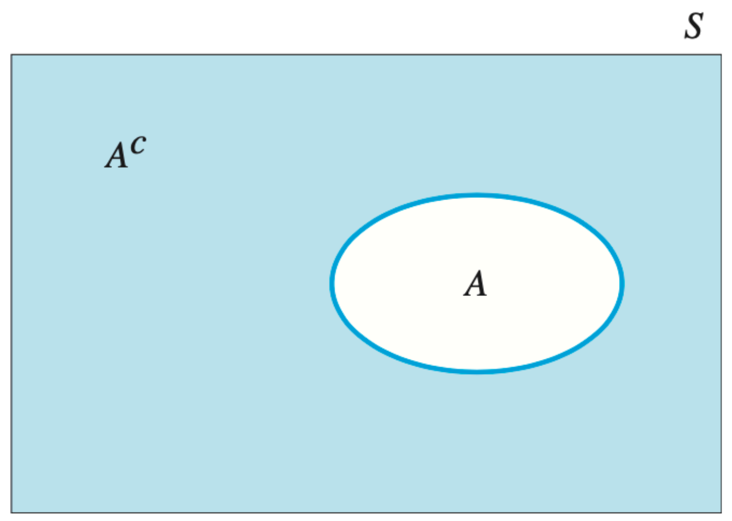
The union of events A and B, denoted by $A \cup B$, is the event that either A or B or both occur.

The intersection of events A and B, denoted by $A \cap B$, is the event that both A and B occur.

The complement of an event A, denoted by A^C, is the event that A does not occur.





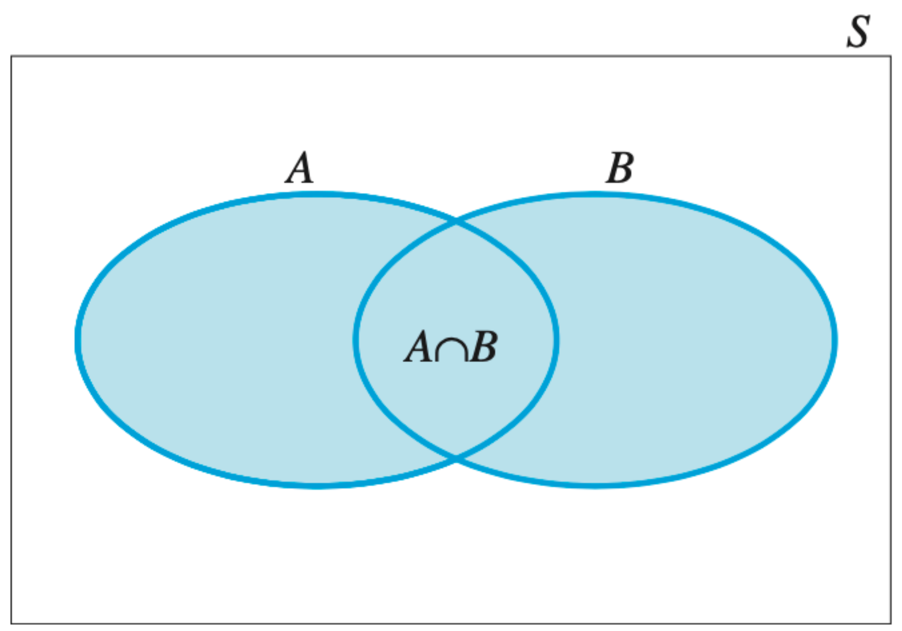


- The Addition Rule for calculating probability of unions

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

If A and B are mutually exclusive, $P(A \cap B) = 0$. Then

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



- Rule for Complements

$$P\left(A^c\right) = 1 - P(A)$$

In a telephone survey of 1000 adults, respondents were asked their opinion about the cost of a college education. The respondents were classified according to whether they currently had a child in college and whether they thought the loan burden for most college students is too high, the right amount, or too little. The proportions responding in each category are shown as:

	Too High (A)	Right Amount (B)	Too Little (C)
Child in College (D)	.35	.08	.01
No Child in College (E)	.25	.20	.11

Suppose one respondent is chosen at random from this group.

1. What is the probability that the respondent has a child in college?

2. What is the probability that the respondent does not have a child in college?

3.What is the probability that the respondent has a child in college or thinks that the loan burden is too high or both?

1. Birthday problem There are 10 people in a room. Assume each person's birthday is equally likely to be any of the 365 days of the year (we exclude February 29), and that people's birthdays are independent (we will define independence formally later, but intuitively it means that knowing some people's birthdays gives us no information about other people's birthdays; this would not hold if, e.g., we knew that two of the people were twins).

    1. How many ways are there to assign birthdays to all the people in the room?
    2. Can you easily count how many ways to assign birthdays so that at least one pair of people in the group have the same birthday?
    3. Can you easily count how many ways to assign birthdays so that no two people share a birthday?
    4. What is the probability that at least one pair of people in the group have the same birthday?

The first value of k for which the probability of a match exceeds 0.5 is $k = 23$. Thus, in a group of 23 people, there is a better than 50% chance that there is at least one birthday match. At $k = 57$, the probability of a match already exceeds 99%.

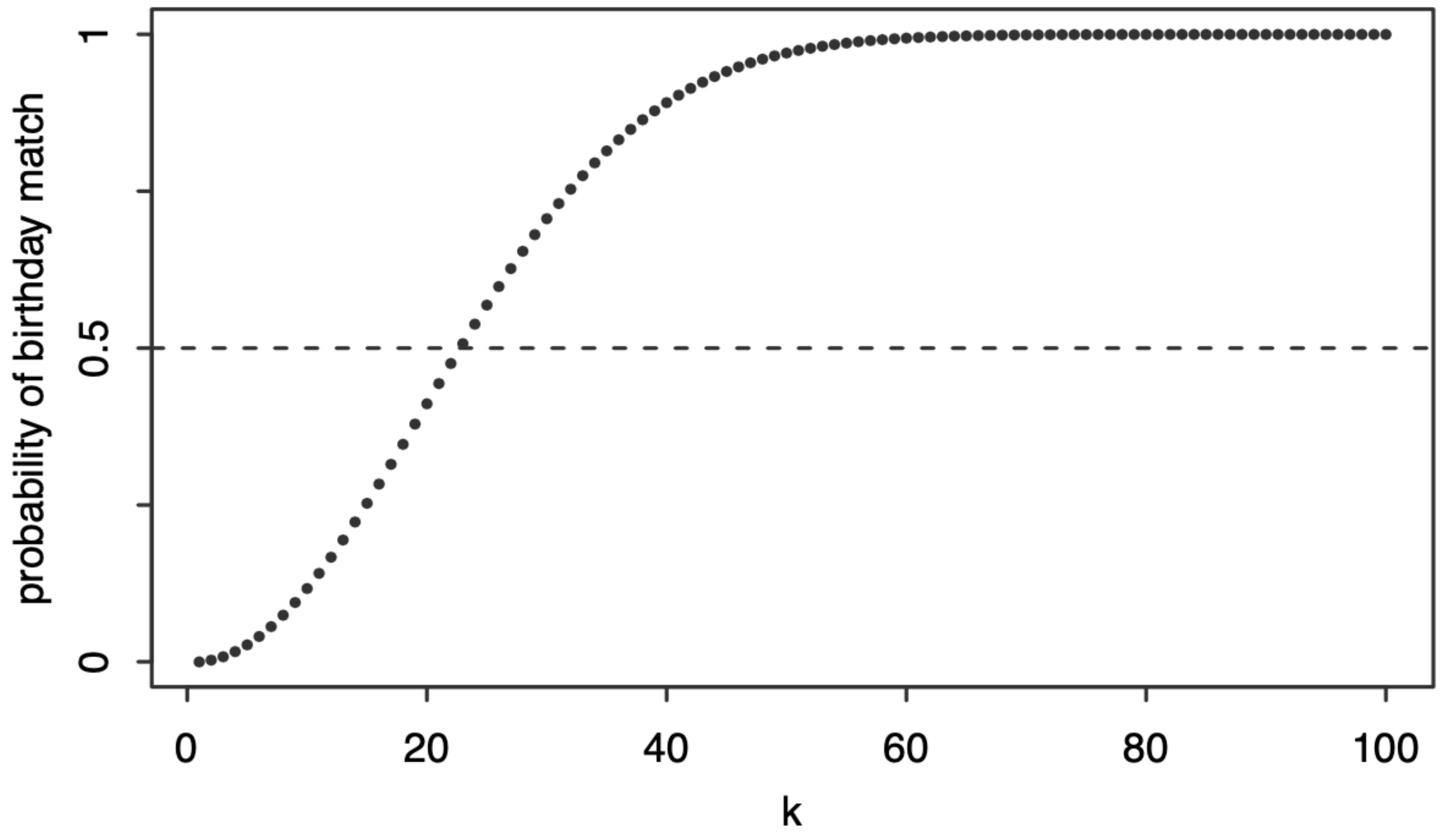


FIGURE 1.5
Probability that in a room of k people, at least two were born on the same day. This probability first exceeds 0.5 when $k = 23$.

Part II : Conditional Probability

- Definition of conditional probability
- Bayes Rule
- Law of Total Probability

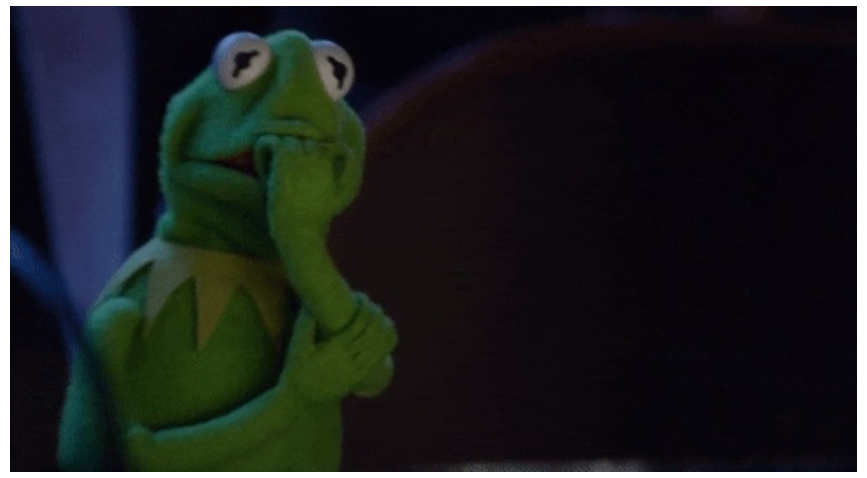
Imagine being awakened one night by the sound of your burglar alarm.

What is the probability or your degree of belief that a burglary attempt has taken place, i.e. someone has broken into your place?

A patient is tested for a disease that affects only 1% of the population.

The test result is positive, i.e. the test claims that the patient has the disease.

What is the probability that he indeed has the disease, given that the evidence provided by test result?



Bad things must be true!

We have introduced probability as a language for expressing uncertainty about events.

Probability can also be interpreted as “the degree of belief or dis-belief” of some claims or events.

Now we have a new situation:

- Whenever we observe new evidence, in the form of data/information (alarm, positive test result....)
- We acquire new information that may affect our belief about the events

How should we update our belief about an event in light of new evidence we observe?

- Hearing alarm must changed our belief about whether burglary happened or not
- Test postive must changed our belief about whether the patient has the disease or not

Conditional Probability is the concept and the tool that address this fundamental question.

"Conditional Probability is the soul of Statistics."

1. How to incorporate evidence into our understanding of the world in a logical, coherent manner

   It's the language of many scientific/ medical / legal reasoning

2. Conditioning is a very powerful problem-solving strategy

   Often make it possible to solve a complicated problem by decomposing it into manageable pieces with case-by-case or condition-by-condition reasoning.

Conditional probability

If A and B are events with P(B)>0,

Then the conditional probability of A given B, denoted by P(A|B) is:

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

- To find the conditional probability that event A occurs given that event B occurs, divide the probability that both A and B occur by the probability that B occurs
- In general, A and B can be any events
- But it's important to interpret:
    - A is the event whose uncertainty we care about
    - B is the evidence we observe, such as burglary alarm, positive test

Suppose A is the event that a person is colorblind, B is the event that a person is men, our population composition is shown as:

	Men(B)	Women (B^C)	Total
Colorblind (A)	.04	.002	.042
Not Colorblind (A^C)	.47	.488	.958
Total	.51	.49	1.00

What is the risk of colorblind for men?

$$P(colorblind \mid man) = ?$$

What is the risk of colorblind for women?

$$P(colorblind \mid woman) = ?$$

=> Risk for men is higher for women!

Here Gender provides extra information about the risk of colorblindness,

therefore is a source of evidence, that explains why we condition on gender.

A standard deck of cards is well-shuffled.

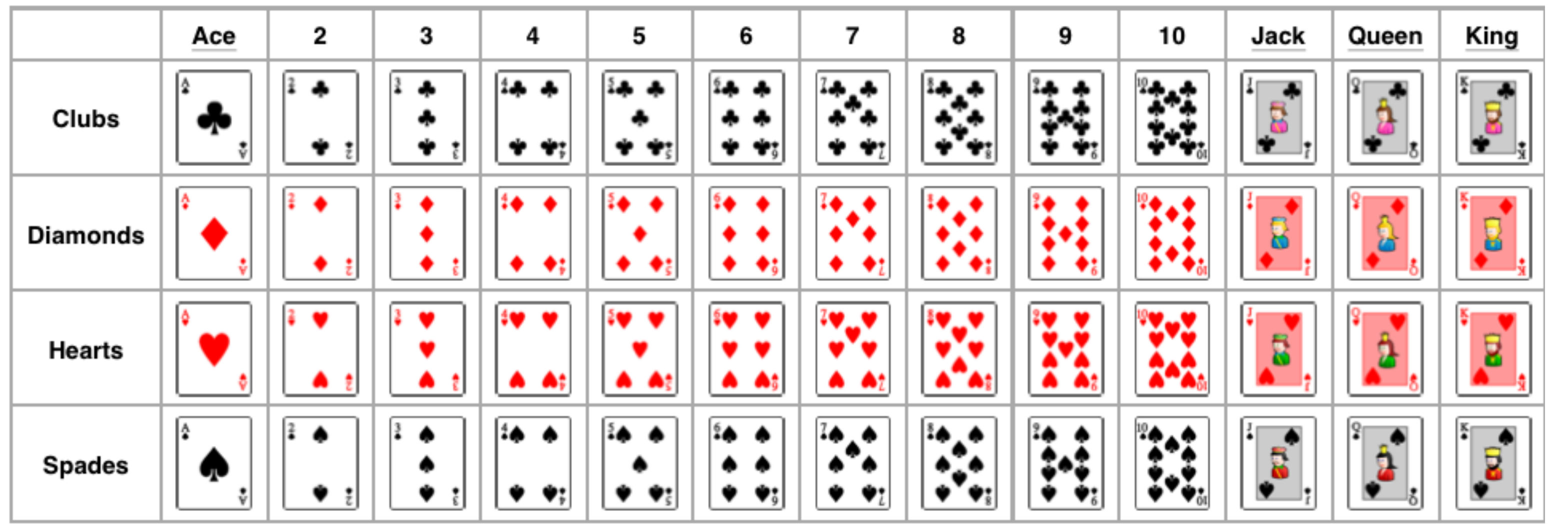
Two cards are drawn randomly, one at a time without replacement.

Let A be the event that 1st card is a heart,

B be the event that 2nd card is a red.

Find P(A|B) and P(B|A)

Example set of 52 playing cards; 13 of each suit: clubs, diamonds, hearts, and spades



(Two Children problem from Scientific American 1950s)

Mr. Jones has 2 children, the order child is a girl,

what is the probability that both are girls?

Mr. Smith has 2 children, at least one of them is a boy,

What is the probability that both children are boys?

Let's re-arrange the definition of conditional probability:

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

Probability of intersection of two events

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

The probability of both events happen can also be calculated by the multiply two probabilities:

- conditional probability of A given B
- Probability of B

We will use it as a tool for Bayes' rule and other rules.

Bayes' Rule

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

Proof: $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(B \mid A)P(A)}{P(B)}$

Why the trouble? Why substitute $P(A \cap B)$ by $P(B \mid A)P(A)$?

- In many applications, we are asked to calculate conditional probability P(A|B)
- So by definition we need $P(A \cap B)$
- But $P(A \cap B)$ is often very hard to calculate directly, unless you're all-knowing God
- It turns out the other way of conditional probability $P(B \mid A)$ is often very easy!

Bayes' rule is extremely famous, one of most useful result!

One more problem with Bayes' rule:

Often, P(B) in Bayes' rule is still hard to calculate directly

The second main theorem of this part:

The Law of Total Probability (LOTP) will complete the main goal of calculating any conditional probability

"Conditional Probability is the soul of Statistics."

1. How to incorporate evidence into our understanding of the world in a logical, coherent manner

   It's the language of many scientific/ medical / legal reasoning

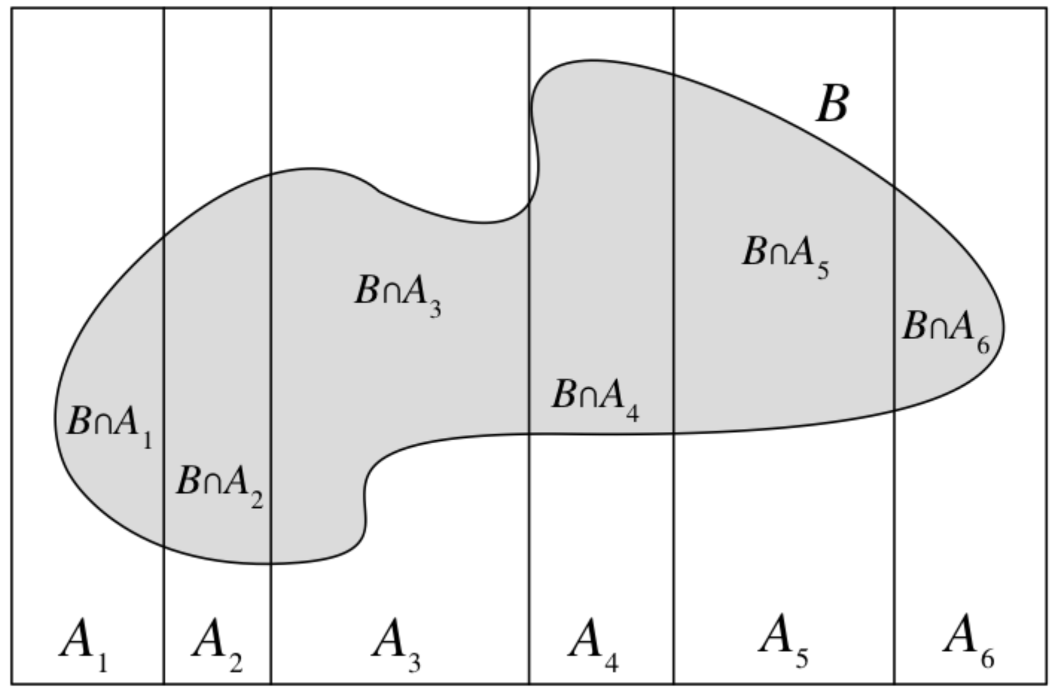
2. Conditioning is a very powerful problem-solving strategy

   Often make it possible to solve a complicated problem by decomposing it into manageable pieces with case-by-case or condition-by-condition reasoning.

The Law of Total Probability

Let n events $A_1 \ldots A_n$ be a partition of sample space S,

i.e. partition: $A_1 \ldots A_n$ are disjoint events, but their union is the whole sample space S



Then: $P(B) = \sum_{i=1}^{n} P\left(B \mid A_i\right) P\left(A_i\right) = = P\left(B \mid A_1\right) P\left(A_1\right) + P\left(B \mid A_2\right) P\left(A_2\right) + \cdots + P\left(B \mid A_n\right) P\left(A_n\right)$

When direct calculation of the probability of some event B is difficult,
we can divide the sample space into slices,
Find the conditional probability in each slices,
Then take a weighted sum of these conditional probabilities, weights being how large the slices are

The choice of how to slice S is crucial: a well-chosen partition will reduce a complicated problem into a simple one.

It's often used together with Bayes' rule, esp. for Bayes' denominator calculation.

Combining:

- the definition of conditional probability

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

- probability of intersection of two events

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

- Law of Total Probability

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

$\Rightarrow$

The final version/most useful version of Baye's rule:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{\sum_{i=1}^{n} P\left(B \mid A_i\right) P\left(A_i\right)}$$

Imagine being awakened one night by the sound of your burglar alarm.

What is the probability or your degree of belief that a burglary attempt has taken place, i.e. someone has broken into your place?

Data has been collected and some evidence provided as:

- There is a 95% chance that a burglary will trigger that alarm

P(alarm|Burglary)=0.95

(This can be easily obtained by testing the alarm system,

On the other hand, P(burglary|alarm) is not easily testable...)

- There is a slight 1% change that alarm will be triggered by a mechanism other than burglary, such as a bug

P(alarm|no burglary)=0.01

- Previous crime pattern in the region indicate that there is 0.0001 chance of burglary

P(burglary)=0.0001

A patient is tested for a disease that affects only 1% of the population.

The test result is positive, i.e. the test claims that the patient has the disease.

What is the probability that he indeed has the disease, given that the evidence provided by test result?

We know additional evidence that:

$$P(D) = 0.01$$

$$P(T \mid D) = 0.95$$

$$P(T^C \mid D^C) = 0.95$$