Before we start…

**Lecturer:** Rui M. Castro  
**Office:** MF4.075  
**Phone:** (040 247) 2499  
**Email:** rmcastro@tue.nl (include the code 2DI90 in the subject)
Before we start...

Setup of the course:
• 9 weeks, consisting of lectures, informal quizzes, and tutorials/intermezzos
• Office hours (in lieu of instructions)

Evaluation:
• Four mandatory homework assignments
• Electronic test in week 9
• Final written exam

Prerequisites:
• TU/e calculus course or equivalent

All this information is available on the course webpage
http://www.win.tue.nl/~rmcastro/2DI90
Before we start...

**Study Materials:**
- (recommended) “Applied Statistics and Probability for Engineers” (Montgomery & Runger) (4th, 5th or 6th edition)
- Statistical Compendium (dikt. nr. 2218)
- Notes and assignments distributed during the course

**Announcements and other course materials:**
- I’ll post everything on the course webpage

http://www.win.tue.nl/~rmcastro/2DI90

CHECK THIS PAGE REGULARLY FOR NEWS, ANNOUNCEMENTS, ETC…
Introduction

Why Probability?

• Many systems are very complex (e.g. an operating system)
• It is not easy to model complex systems
• Sometimes we cannot predict how users are going to behave
A problem has been detected and Windows has been shut down to prevent damage to your computer.

The problem seems to be caused by the following file: SPCMDCON.SYS

PAGE_FAULT_IN_NONPAGED_AREA

If this is the first time you've seen this stop error screen, restart your computer. If this screen appears again, follow these steps:

Check to make sure any new hardware or software is properly installed. If this is a new installation, ask your hardware or software manufacturer for any Windows updates you might need.

If problems continue, disable or remove any newly installed hardware or software. Disable BIOS memory options such as caching or shadowing. If you need to use Safe Mode to remove or disable components, restart your computer, press F8 to select Advanced Startup Options, and then select Safe Mode.

Technical information:

*** STOP: 0x00000050 (0xFD3094C2,0x00000001,0xFBFE7617,0x00000000)

*** SPCMDCON.SYS - Address FBFE7617 base at FBFE5000, DateStamp 3d6dd67c
Introduction

Why Probability?

• Many systems are very complex (e.g. an operating system)
• It is not easy to model complex systems
• Sometimes we cannot predict how users are going to behave

THE OUTCOME OF SOME EXPERIMENTS IS “UNPREDICTIBLE” !!!

What can we do with probabilistic models?

• Use randomness to model various events that can happen.
• Use stochastic models to describe how users/clients interact with a system (e.g. TCP/IP, queuing models)
• Quantify the reliability of a system
Relation to Statistics

Probability and Statistics are NOT the same thing!!!

• Probability provides the foundation of statistics
• Statistics is concerned with making inference from data, assumed to be collected according to a probabilistic model

Population (World) → Probabilistic Model → Sample → Statistics - Inference about the population using the sample
What is this Course About?

After successfully completing this course you should be able to:

• Perform elementary probability calculations with stochastic models
• Identify situations where probabilistic models are adequate and useful, and be able to simulate them

• Perform elementary statistical analyses of data
• Use statistical software in a proper way to obtain both qualitative and quantitative data analysis results
• Recognize scenarios where linear regression is an adequate tool, and perform simple statistical analyses in such settings
An Example Out of the Street

Almost anyone is familiar with the three-card monte or shell game:

• There are three ‘shells’, and under one of these there is a little ball
• If you correctly identify the shell containing the ball you double your money, otherwise you lose.
A Variation of this Game

• I shuffle the ball under one of the shells without you seeing where it goes
• You choose one of the shells
• After you made your choice I lift one of the two other shells and show you that it is empty

• Then I give you two options:
  • You can stay with the initial shell (STAY - A)
  • You can pick any other shell (SWITCH - B)

• Finally I lift the shell you chose and you win if the ball is under it.

What is the best strategy???
Sample Spaces and Events (§ 2.1 MR)

**Definition:** Random Experiment

An experiment that can result in different outcomes, even if repeated in the same manner.

**Examples:**

- Time to reboot a server (in seconds)
- Measuring the clock frequency in a motherboard (in GHz)
- Age (in terms of number of birthdays) of a person
- Measuring the temperature evolution in a GPU while performing a benchmark task (in deg. Celsius)
Sample Space

**Definition:** Sample Space

The set of all possible outcomes of a random experiment is called the sample space of the experiment. We will frequently use $S$ to denote the sample space.

**Examples:**

- Time to reboot a server (in seconds) $S = [0, \infty) = \mathbb{R}_0^+$
- Measuring the clock frequency in a motherboard (in GHz) $S = \{0\} \cup [2.5, 4.0]$
- Age (in terms of number of birthdays) of a person $S = \{0, 1, 2, 3, \ldots\}$
- Measuring the temperature evolution in a GPU while performing a benchmark task (in deg. Celsius) $S = \{\text{functions taking positive values}\}$
Discrete and Continuous Sample Spaces

A sample space is **discrete** if it consists of a finite or countable infinite set of outcomes. A sample space is **continuous** if it contains an interval (either finite or infinite) of real numbers.

Examples:

- Time to reboot a server
  \[ S = [0, \infty) = \mathbb{R}^+ \]

- Measuring the clock frequency in a motherboard
  \[ S = \{0\} \cup [2, \infty) \]

- Age (in terms of number of birthdays) of a person
  \[ S = \{0, 1, 2, 3\} \]

- Measuring the temperature evolution in a GPU while performing a benchmark task
  \[ S = \{ \text{functions taking positive values} \} \]
More Examples

The choice between discrete or continuous sample spaces might depend on the particular objective of the study...

- Messaging in a communication system:
  - Three time sensitive messages are sent. These will either be on time, or delayed

What is the sample space?

\[ S \equiv \{000, 001, 010, 011, 100, 101, 110, 111\} \]


**Events**

**Definition: Events**

An event is a subset of the sample space of a random experiment.

Let $E_1$ and $E_2$ be two events, e.g.

$E_1 = \{\text{number of students is greater or equal than 100}\}$

$E_2 = \{\text{number of students is at least 50 and no more than 150}\}$

- **Union:** $E_1 \cup E_2$ ($E_1$ or $E_2$ occur) (or both occur simultaneously)

- **Intersection:** $E_1 \cap E_2$ ($E_1$ and $E_2$ occur)

- **Complement:** $E'_1 \equiv E_1^C = S \setminus E$ ($E_1$ does not occur)
Events

Example: Time it takes to deliver a message ($S = \mathbb{R}_0^+$)

$E_1 = [0, 10] \equiv \{\text{it takes less than 10ms to deliver the message}\}$

$E_2 = [5, 12] \equiv \{\text{it takes no less than 5ms and no more than 12ms to deliver the message}\}$

Events are just sets contained in the sample space!!!

- **Union:** $E_1 \cup E_2 = [0, 12]$ ($E_1$ or $E_2$ occur)

- **Intersection:** $E_1 \cap E_2 = [5, 10]$ ($E_1$ and $E_2$ occur)

- **Complement:** $E_1 = E_1' = [10, \infty)$ ($E_1$ does not occur)
Set Operations

Let $A$, $B$, and $C$ be arbitrary sets (events). Then

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

$$(A')' = A$$

De Morgan's laws (1806 – 1871)

$$(A \cup B)' = A' \cap B'$$

$$(A \cap B)' = A' \cup B'$$

**Definition:** Mutually Exclusive Events (Disjoint Events)

Two events $E_1$ and $E_2$ are **mutually exclusive** if $E_1 \cap E_2 = \emptyset$. This means the two events cannot occur simultaneously!
Venn Diagrams

\[ A \setminus B = A \cap B' \]

\[ A \cap B \]

\[ B \setminus A = B \cap A' \]

\[ S \]

B and C are mutually exclusive
Counting Techniques

Sometimes we need to count the number of possible outcomes. So far we seen only easy cases, but things can get complicated pretty quickly...

Example 1: Options in a car
- Manual or automatic transmission
- With or without air conditioning
- Three different stereo systems
- Four possible exterior colors

How many possibilities are there?
- A - 11
- B - 32
- C - 48
- D - 54

\[ S \text{ has 48 elements} \]
\[ (2 \times 2 \times 3 \times 4) \]
Permutations

**Example:** 4 friends (Alice, Bob, Eve and Robert) place 4 pieces of paper with their names in a hat. After shuffling them each one of them takes one piece.

What are the possible outcomes of this experiment?

The number of permutations of \( n \) different elements is \( n! \), where

\[
  n! = n \times (n - 1) \times (n - 2) \times \cdots \times 2 \times 1
\]

In this case there are \( 4 \times 3 \times 2 \times 1 = 24 \) possibilities

**Sinterklaas Challenge:** If you have \( n \) friends how many possible outcomes exist such that no one takes their own name out of the hat???

(if \( n=4 \) the answer is 9)
Permutations

Example: A motherboard has 8 slots, where we want to place 4 different cards. How many possibilities are there?

The number of permutations of subsets of $r$ elements selected from a set of $n$ different elements is

$$P^n_r = \frac{n!}{(n-r)!}$$

Why?
• Take card one and place it in one of the 8 positions (8 possibilities).
• Take card 2 and place it in one of the 7 remaining positions
• ... and so on

We have $P^8_4 = \frac{8!}{4!} = 8 \times 7 \times 6 \times 5 = 1680$ possibilities
Combinations

Example: A motherboard has 8 memory slots, and we want to put 4 identical memory cards in it.

The number of subsets of $r$ elements out of a set of $n$ is called the number of combinations and it is given by

$$C_r^m = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Why?
- Same as before, but now the order of the cards doesn’t matter. So, there are $4!$ possible ordering of the 4 cards all yield the same outcome.

We have $C_4^8 = \frac{8 \times 7 \times 6 \times 5}{4 \times 3 \times 2 \times 1} = \frac{1680}{24} = 70$ possibilities
Sampling With or Without Replacement

**Example:** I ask you to take five cards out of a deck, one by one without putting them back (sampling *without* replacement)

**Example:** I ask you to take a card out of the deck and look at it, then replace it, I shuffle the deck and ask you to take a card again, and so on, five times (sampling *with* replacement)

Which type of sampling is more adequate depends on the context (but typically sampling with replacement is easier to study)
Interpretation of Probability (§ 2.2 MR)

The concept of probability is a very abstract one:
  • Idea: quantify the “likelihood” or “chance” of the outcome of an experiment...

  • One can have a **subjective** interpretation of probability:

The probability of event $E$, is denoted by $P(E)$, and expressed the degree of belief we assign for the occurrence of event $E$.

It takes a number between 0 and 1, where 0 indicates the event does not occur, and 1 indicates the event certainly occurs (one can also consider percentages).

This might not be very satisfying, as different people will assign different values for the probability of an event...
Interpretation of Probability

- A frequentist interpretation of probability:

Suppose we can repeat a random experiment $n$ times. The possible outcome is either in $E$ or not in $E$. Now let $N(n, E)$ be the number of times the outcome of the experiment is in $E$. The frequentist interpretation of probability defines $P(E)$ as

$$P(E) \equiv \lim_{n \to \infty} \frac{N(n, E)}{n}.$$ 

This is simply the fraction of times $E$ happens when we repeat the experiment a large number of times!

Example:
- Take the first card out of a shuffled deck
- Consider the event \{the card selected is an ace\}

What do we exactly mean by the repetition of an experiment ???
Axioms of Probability

We will restrict ourselves to discrete sample spaces for now, to avoid some technical difficulties...

By taking an axiomatic approach we can make sure everything is well defined, if we take the axioms for granted...

Let \( S = \{s_1, s_2, \ldots\} \) be the sample space. Let \( E \subseteq S \) be an arbitrary event.

(i) \( P(S) = 1 \)

(ii) \( 0 \leq P(E) \leq 1 \)

(iii) For two events \( E_1 \) and \( E_2 \) such that \( E_1 \cap E_2 = \emptyset \)

\[
P(E_1 \cup E_2) = P(E_1) + P(E_2)
\]
Axioms of Probability

Example: Random experiment - Flipping a coin \( S = \{ \text{heads, tails} \} \)

\[
H = \{ \text{heads} \} \quad T = \{ \text{tails} \}
\]

\[
P(S) = P(\{ \text{heads, tails} \}) = 1
\]

\[
P(S') = P(\emptyset) = P(\text{neither heads or tails}) = 0
\]

\[
P(S) = P(H \cup T) = P(H) + P(T) \quad \text{(as } H \cap T = \emptyset)\]

If we have a fair coin then \( P(H) = P(T) = 1/2 \)

In general, if you have \( N \) possible outcomes:

If \( N \) outcomes are equally likely then the probability of each outcome is \( 1/N \)
Basic Properties

Just from the axioms we can deduce a number of simple, but useful properties:

Lemma:

(i) \( P(A') = 1 - P(A) \)

(ii) \( P(\emptyset) = 0 \)

Proof: Note that \( S = A \cup A' \), and that \( A \) and \( A' \) are mutually exclusive. Therefore

\[
P(S) = P(A) + P(A') \Leftrightarrow 1 = P(A) + P(A')
\]

Applying (i) to the event \( S \) immediately shows part (ii).
Addition Rules (§ 2.3 MR)

Actually, the probability of an event is just a way to measure it!!
- you can think of it as the generalized volume of the set.

\[ A \setminus B = A \cap B' \]

\[ B \setminus A = B \cap A' \]

**Lemma:**
\[ P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

If the events are mutually exclusive then (and only then!):
\[ P(A \cup B) = P(A) + P(B) \]
Proof: \( P(A \cup B) =? \)

\[ P(A) \]

\[ P(B) \]

\[ P(A) + P(B) \text{ counts} \]

\[ \text{twice, therefore} \]

\[ P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

We can obviously generalize all this to multiple events...
Conditional Probability (§ 2.4 MR)

This is a very important concept. In most cases events “share” information, and so we want to see how can we take this into account.

Let $A$ and $B$ be two arbitrary events. Then the **conditional probability** of $B$ given $A$ is denoted by

$$P(B|A)$$

You should read the above expression as “the probability of $B$ given $A$”.

**Example:** taking a card out of a shuffled deck

$A = \{\text{the selected card is a red queen}\}$

$B = \{\text{the selected card is of hearts}\}$

$P(A) = \frac{2}{52} \approx 0.0384$

$P(B) = \frac{13}{52} = 0.25$

$P(B|A) = \frac{1}{2}$

$P(A|B) = \frac{1}{13} \approx 0.0769$
**Conditional Probability**

**Definition:** Conditional Probability

Let $A$ and $B$ be two events, and assume $P(A) > 0$. The **conditional probability** of $B$ given $A$ is defined as

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$
An Illustrative Toyish Example

Sammy has only four types of pairs of socks in his drawer:

- Red or Blue socks
- Striped or plain

<table>
<thead>
<tr>
<th>Striped \ Plain</th>
<th>Red</th>
<th>Blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>Blue</td>
<td>16</td>
<td>8</td>
</tr>
</tbody>
</table>
An Illustrative Toyish Example

Sammy has only four types of pairs of socks in his drawer:

- Red or Blue socks
- Striped or plain

<table>
<thead>
<tr>
<th></th>
<th>Striped</th>
<th>Plain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>Blue</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>24</td>
</tr>
</tbody>
</table>

```
Red          Blue          Red          Blue
42
18
24
```
An Illustrative Toyish Example

Sammy takes a pair of socks from his drawer at random (he is equally likely to take any of the pairs):

• What is the probability that it is a red pair?
• Given he took a pair of red socks, what is the probability these have stripes?

$A = \{\text{taking red socks}\}$

$B = \{\text{taking striped socks}\}$

$P(A) = \frac{29}{42} \approx 0.69$

$P(B|A) = \frac{13/29}{42} \approx 0.45$
An Illustrative Toyish Example

\[ A = \{ \text{taking red socks} \} \]
\[ B = \{ \text{taking striped socks} \} \]

\[ P(B) = \frac{18}{42} \approx 0.429 \]

\[ P(A|B) = \frac{13}{18} \approx 0.722 \]
Multiplication Rules ($\S$ 2.5 MR)

**Multiplication Rule:**

Let $A$ and $B$ be two events

\[ P(A \cap B) = P(B|A)P(A) = P(A|B)P(B) \]

**Total Probability Rule:**

Let $A$ and $B$ be two events

\[ P(B) = P(B \cap A) + P(B \cap A') = P(B|A)P(A) + P(B|A')P(A') \]
Total Probability Rule for Multiple Events

Let $A$ be an arbitrary event, and let $E_1, \ldots, E_k$ be $k$ mutually exclusive and exhaustive events. That is,

$$\bigcup_{i=1}^{k} E_i = S, \quad \text{and} \quad \forall i \neq j \ E_i \cap E_j = \emptyset .$$

Then

$$P(A) = P(A \cap E_1) + P(A \cap E_2) + \cdots + P(A \cap E_k)$$

$$= P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \cdots + P(A|E_k)P(E_k) .$$
Independence of Events (§ 2.6 MR)

Sometimes two events might not be “related”:

• Given that the outcome of the experiment is in \( A \) might not affect the probability that this same outcome is also in \( B \).

**Example:** Take a card out of a shuffled deck

\[
A = \{ \text{the selected card is a queen} \} \\
B = \{ \text{the selected card is of hearts} \}
\]

\[
P(A) = \frac{4}{52} = \frac{1}{13}, \quad P(B) = \frac{13}{52} = \frac{1}{4}, \quad P(A \cap B) = \frac{1}{52}
\]

\[
P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/52}{13/52} = \frac{1}{13} = P(A)
\]

This means that \( B \) doesn’t give any probabilistic information about \( A \) !!!
Definition of Independence

**Definition:** Independence

Two events $A$ and $B$ are independent are equivalent if any of the following equivalent statements is true:

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

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

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

Independence is a truly probabilistic concept, and not a set relationship !!!

The events $E_1, E_2, \ldots, E_k$ are jointly independent if

$$P(E_1 \cap E_2 \cap \cdots \cap E_k) = P(E_1)P(E_2)\cdots P(E_k)$$

**Challenge:** Can you construct a probability model and three events that are pairwise independent but not jointly independent?
Properties

Lemma: The following statements are equivalent:

(i) $A$ and $B$ are independent

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

(iii) $P(A' \cap B) = P(A')P(B)$

(iv) $P(A \cap B') = P(A)P(B')$

(v) $P(A' \cap B') = P(A')P(B')$

Proof: left as exercise...
Independence is IMPORTANT !!!

Independence is a key assumption in most probability models, e.g.:

- We assume each memory chip from a certain manufacturer fails independently from one another
- Processes arriving to a server from different users are assume to arrive independently from one another

**Example:** There are three elevators in the MetaForum, but one is currently broken. The other two fail independently with probability 0.2 and 0.1 respectively.

What is the probability I cannot take one of the elevators to get to the 7\textsuperscript{th} floor?

A – 0.1  
B – 0.2  
C – 0.3  
D – 0.02
A Simple Example

We can think of this setting as a circuit:

\[
P(E_1) = P(\text{Elevator 1 works}) = 1 - 0.2 = 0.8
\]

\[
P(E_2) = P(\text{Elevator 2 works}) = 1 - 0.1 = 0.9
\]

\[
P(\text{at least one of the elevators works}) = P(E_1 \cup E_2)
\]
\[
= 1 - P((E_1 \cup E_2)')
\]
\[
= 1 - P(E_1' \cap E_2')
\]
\[
= 1 - P(E_1')P(E_2') = 1 - (1 - 0.8)(1 - 0.9) = 0.98
\]
Another Example

Example: A harddrive backup unit uses redundancy to ensure reliable operation. It consists of two redundant data storage units (each with probability of failure of 1/100) and two independent power supply units each with probabilities of failure respectively 1/50 (electric grid unit) and 1/120 (battery unit).

What is the overall probability of failure of such system?

Overall probability success = 0.9999 * 0.9998 = 0.9997
Bayes’ s Rule (§ 2.7 MR)

In many situations in practice what we can measure (estimate) are conditional probabilities. It is rather useful to be able to relate various conditional probabilities to one another.

Reverend Thomas Bayes (1702-1761):

Let $A$ and $B$ be two events, and $P(B) > 0$. Then

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

Proof:

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

Despite its simplicity, this is a very powerful result, and forms the basis of many inference procedures used nowadays (e.g. both the GPS system and the communication encoding used in cellphones rely on Bayes’ s rule).
Bayes’ s Rule

Let $A$ be an arbitrary event such that $P(A) > 0$, and let $E_1, \ldots, E_k$ be $k$ mutually exclusive and exhaustive events. That is,

$$\bigcup_{i=1}^{k} E_i = S, \quad \text{and} \quad \forall i \neq j \ E_i \cap E_j = \emptyset.$$  

Then

$$P(E_1|A) = \frac{P(A|E_1)P(E_1)}{P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \cdots + P(A|E_k)P(E_k)}$$
**A Real Life Example**

Gerd Gigerenzer is a cognitive psychologist who studies how people miscalculate risk.

In one study, Gigerenzer and his colleagues asked doctors in Germany and the United States to estimate the probability that a woman with a positive mammogram actually has breast cancer, even though she’s in a low-risk group: 40 to 50 years old, with no symptoms or family history of breast cancer. To make the question specific, the doctors were told to assume the following statistics — couched in terms of percentages and probabilities — about the prevalence of breast cancer among women in this cohort, and also about the mammogram’s sensitivity and rate of false positives:

The probability that one of these women has breast cancer is 0.8 percent. If a woman has breast cancer, the probability is 90 percent that she will have a positive mammogram. If a woman does not have breast cancer, the probability is 7 percent that she will still have a positive mammogram.

Imagine a woman who has a positive mammogram. What is the probability that she actually has breast cancer?

The doctors answer for this question ranged from 1 to 99%. What is the correct answer???

Enter your answer as a percentage (e.g. 47% corresponds to 47)
A Real Life Example

The probability that one of these women has breast cancer is 0.8 percent. If a woman has breast cancer, the probability is 90 percent that she will have a positive mammogram. If a woman does not have breast cancer, the probability is 7 percent that she will still have a positive mammogram.

Imagine a woman who has a positive mammogram. What is the probability that she actually has breast cancer?

\[ M = \{ \text{woman has positive mammogram} \} \]
\[ C = \{ \text{woman has breast cancer} \} \]

\[ P(C) = 0.008 \]
\[ P(M|C) = 0.9 \]
\[ P(M|C') = 0.07 \]

We desire to know what \( P(C|M) \) is...
A Real Life Example

\[ M = \{ \text{woman has positive mammogram} \} \]
\[ C = \{ \text{woman has breast cancer} \} \]

\[ P(C|M) = \frac{P(M|C)P(C)}{P(M)} = ??? \]

Total Probability Rule:

\[ P(M) = P(M|C)P(C) + P(M|C')P(C') \]
\[ = 0.9 \times 0.008 + 0.07 \times (1 - 0.008) = 0.07664 \]

Therefore

\[ P(C|M) = \frac{P(M|C)P(C)}{P(M)} = \frac{0.9 \times 0.0008}{0.070664} = 0.09395 \]

So a woman between 40 and 50 with a positive mammogram has only about 9.5% chance of actually having breast cancer...
**A Paradoxical Example**

**Example:** Real experimental data for two different treatments of kidney stones is shown below. Their success was studied when patients had both small and large stones in four groups of patients:

<table>
<thead>
<tr>
<th>Stone size</th>
<th>Treatment A</th>
<th></th>
<th>Treatment B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Successful</td>
<td></td>
<td>Successful</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Small</td>
<td>81</td>
<td>6</td>
<td>234</td>
<td>36</td>
</tr>
<tr>
<td>Large</td>
<td>192</td>
<td>71</td>
<td>55</td>
<td>25</td>
</tr>
</tbody>
</table>

Which treatment is better?
Treatment A

$A = \{\text{treatment A is successful}\}$

$S = \{\text{patient has small stones}\}$

$L = \{\text{patient has large stones}\}$

Let’s compute the overall success probability of treatment A:

$$P(A) = \frac{81 + 192}{87 + 263} = \frac{273}{350} = 0.780$$

Let’s also compute conditional success probabilities:

$$P(A|S) = \frac{81}{87} = 0.931$$

$$P(A|L) = \frac{192}{263} = 0.730$$
Treatment B

\[ B = \{ \text{treatment B is successful} \} \]
\[ S = \{ \text{patient has small stones} \} \]
\[ L = \{ \text{patient has large stones} \} \]

Let’s compute the overall success probability of treatment B:

\[
P(B) = \frac{234 + 55}{270 + 80} = \frac{289}{350} = 0.826
\]

Let’s also compute conditional success probabilities:

\[
P(B|S) = \frac{234}{270} = 0.867
\]
\[
P(B|L) = \frac{192}{263} = 0.688
\]
A Paradoxical Example

In terms of the overall success rate we have

\[ P(A) = 0.780 \quad \text{and} \quad P(B) = 0.826 \]

So treatment B seems to be better overall...

But, For each class of patients we have

\[ P(A|S) = \frac{81}{87} = 0.931 \quad \quad P(A|L) = \frac{192}{263} = 0.730 \]
\[ P(B|S) = \frac{234}{270} = 0.867 \quad \quad P(B|L) = \frac{192}{263} = 0.688 \]

So for patients with small stones treatment A is better, and for patients with large stones treatment A is also better...

Where’s the catch?
The Two Studies

<table>
<thead>
<tr>
<th>Stone size</th>
<th>Successful</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>Yes</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>6</td>
</tr>
<tr>
<td>Large</td>
<td>Yes</td>
<td>192</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>71</td>
</tr>
</tbody>
</table>

Few patients with small stones
Many patients with large stones

\[
P(A) = P(A|S)P(S) + P(A|L)P(L)
\]
\[
= 0.931 \times 0.249 + 0.730 \times 0.751
\]
\[
= 0.780
\]

This is known as Simpson’s paradox, and continues to confuse many people, even nowadays...

<table>
<thead>
<tr>
<th>Stone size</th>
<th>Successful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Large</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>

\[
P(B) = P(B|S)P(S) + P(B|L)P(L)
\]
\[
= 0.867 \times 0.771 + 0.688 \times 0.229
\]
\[
= 0.826
\]
Random Variables (§ 2.8 MR)

Often it is useful to consider random variables. These map the outcomes of a random experiment to a number.

A random variable is a function that assigns a real number to each outcome in the sample space of a random experiment.

Remark on notation: We will use capital letters (e.g. $X$) to denote random variables. After the random experiment is performed the observed value of the random variable is denoted by a lower case (e.g. $x$).

Example:
- $X$ is the random variable corresponding to the temperature of the room at time $t$.
- $x$ is the measured temperature of the room at time $t$. 
Random Variables

A **discrete** random variable is a random variable with a finite or countable range. For example: number of scratches in the surface of a disk; number of TCP connections to a server; number or errors in a transmission.

A **continuous** random variable is a random variable whose range is an interval of real numbers. For example: electric current; GPU temperature; clock frequency; roundtrip time.

Next in the course:

• We are going to study both discrete and continuous random variables, and learn about probabilistic settings that are useful to model real-world situations...