Crash Course in Statistics for Neuroscience Center Zurich University of Zurich

Dr. C.J. Luchsinger

1 Probability

Nice to have read: Chapters 1, 2 and 3 in Stahel or Chapters 1 and 2 in Cartoon Guide

Further readings: Chapter 4 in Stahel or Chapter 3 in Cartoon Guide

In this chapter...

we formalize **Events**, introduce the **P**robability of an Event ($P[A]$) and get to know the manipulations we may make with Events and Probabilities. **Conditional Probabilities** ($P[A|B]$) allow us to look at Probabilities of an event A given a second event B occurs; for example given some mouse has a disease, how high is probability it has a higher temperature than control group? In above question you might see that this probability is independent of the mouse having a certain disease. Then we say these two events are independent. Independence is in general an important assumption when analysing data. To end this chapter we look at **Bayes Theorem** which is important in elementary medical statistics.

1.1. What do you associate with "Probability" and/or "Statistics"?

1.2 Probability and Statistics is widely used in:

Games, Biology, Medicine, Pharmacy, Sociology, Demographics, Official Statistics, Finance and Insurance, Engineering, Physics, Quality Control

For more information see "www.math-jobs.ch/beruf.html" and "www.math-jobs.com"

1.3 History

17th century: I am winning in Gambling; Police; stop immediately; how much do I get?
19th century: Agriculture: Variability growth of crop, fertilizer; more in chapter 7
20th century: crusade of Statistics, all fields of science.
21th century is going to be the century of Biology and Medicine,

given Mathematics and Statistics will provide the necessary tools.

Famous people: Jakob Bernoulli, Moivre, Laplace, Gauss, Poisson, Tschebyschew, Markov, Kolmogoroff, **F**isher (**F**-Test) and many more.

1.4 Little bit of combinatorics

Factorial

$$n! := n(n-1)(n-2)\cdots 2\cdot 1 \qquad \text{called ”Factorial } n\text{”}$$

By convention: $0! := 1$.

Travelling Salesman Problem: Coming from Europe, have to visit 50 Cities in the U.S. (once each) and return to the first one at the end. How many possibilities?

Binomial coefficient

$$\binom{n}{k} := \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k!} \qquad \text{called ”}n\text{ choose }k\text{”}$$

In a study with 10 mice (numbered 1 to 10), 6 die and 4 survive (I did *not* say number 1 to 6 and 7 to 10 respectively). How many possibilities are there for such an event (we are not yet looking at the *probability*, only how many possibilities)?

Algebraically correct, but why must

$$\binom{n}{k} = \binom{n}{n-k}$$

hold intuitively?

Compute $\binom{21}{3}$

Compute $\binom{21}{18}$

Here's a tip for not getting invited again to some strange party: Ask the question: "There are 50 participants: How many times do pairs of 2 shake hands with each other to greet each other?"

1.5 Probability P

For the next few steps, it is easiest to think about games, gambling (we will return to more serious problems later): Dice / Coin Flipping.

* Ω is called the **sample space**, set (German: Menge) of possible outcomes of an experiment. In each experiment we have precisely one **elementary outcome**: $\omega_1 \in \Omega$ or $\omega_2 \in \Omega$ etc.

* Dice: *we choose* $\Omega := \{1, 2, 3, 4, 5, 6\}$.

* Coin: *we choose* $\Omega := \{h, t\}$.

* Pairs of Dice, red and blue (Monopoly): *we choose* $\Omega := \{(1, 1), (1, 2), \ldots, (6, 5), (6, 6)\}$; the sample space has 36 elements in this experiment.

* Return to single Dice: $A := \{2, 4, 6\}$ set of even numbers, $B := \{4, 5, 6\}$ set of numbers larger than 3. A and B are called **events**. In general: events are subsets of Ω. If the elementary outcome of the experiment "throw a Dice" is 2, then event A occurred: we *have* an even number. Event B did not occur: $2 < 4$. By the way: elementary outcomes are events too: $\{2\} \subset \Omega$.

* Small exercise: Play Monopoly, need 8 to get to "Start". Describe C, the event that we have "8".

We use set theory to formalize events:

Symbol	**set theory / what does it mean for probability theory**
Ω	Set / Sample Space, all possible outcomes of an experiment
ω	Element of Ω / elementary outcome of experiment
A	Subset of Ω / Event; if $\omega \in A$, Event A occurred
A^c	Complement of A / no elementary outcome of A occurred
$A \cap B$	Intersection of A and B / elementary outcome occurred that lies in A **and** B
$A \cup B$	Union of A and B / elementary outcome lies in A **or** B (or in both, A and B).
$A \backslash B$	A but not B / elementary outcome lies in A but not in B
$A \subset B$	A is subset of B / if we have event A, we always have event B too.
ϕ	empty set / impossible event
Ω	entire set / certain event, something must happen

Important examples for

* A^c (if A very complex, A^c might be simple to describe)

* $A \subset B$ (Zurich is Part of Switzerland!)

We now want to introduce the "P", the **probability** of an event. If you've grown up in a "normal" family with "normal" siblings, you have numbers such as 0.5 (Coin), 1/6 (Dice) and 1/36 (Pair of Dice) in mind. But this must not be so. All we ask for is

Definition 1.1 [Probability P] *Probability P is a function from the subsets of Ω into the interval $[0,1]$. P must satisfy:*

a) $0 \leq P[A] \leq 1$ *for all* A

b) $P[\Omega] = 1,\ P[\phi] = 0$

c) A_1, A_2 *disjoint [German: elementfremd] sets - that is* $A_1 \cap A_2 = \phi$*, then:*

$$P[A_1 \cup A_2] = P[A_1] + P[A_2].$$

Dice: $\Omega := \{1, 2, 3, 4, 5, 6\}$. A fair *choice* (this is a model) is $P[\{i\}] = 1/6$ for all $i \in \Omega$.

Let us look at c), and then go back to a) and b).

Returning to your childhood and your big brother: we (or better he) *can* choose $P[\{1\}] = 0.2, P[\{2\}] = 0.2, P[\{3\}] = 0.2, P[\{4\}] = 0.2, P[\{5\}] = 0.1, P[\{6\}] = 0.1$ and there are such dice. But they are not fair.

Probability and proportions: 8 million inhabitants in Switzerland. Choose one with equal probability for each of the 8 million inhabitants. Probability that I choose a millionaire(s) is the proportion of people of Switzerland who are millionaires. If proportion of millionaires is 10 %, then the probability to choose a millionaire(s) is 0.1.

Lemma 1.2 [useful stuff about P] *P satisfies*

a) $P[A] = 1 - P[A^c]$ *[in German: "Prinzip der Gegenwahrscheinlichkeit"]*

b) A_1, A_2 *sets, then:*

$$P[A_1 \cup A_2] \leq P[A_1] + P[A_2].$$

c) $A \subset B \Rightarrow P[B] = P[A] + P[B \backslash A]$

d) $A \subset B \Rightarrow P[A] \leq P[B]$

e) $P[A \cup B] = P[A] + P[B] - P[A \cap B]$.

Exercises: p. 36-39 in the cartoon guide

1.6 Conditional Probability of A given B: $P[A|B]$

Think of a very simple example, fair Dice: Probability that we have an even number (event A), given we have a number smaller than 4 (event B). The answer is:

How did you compute this number? It was clear, that we only have to look at the numbers $\{1, 2, 3\}$ "given smaller than 4". This is our *new sample space*. Each elementary outcome has probability 1/3 (originally 1/6). The answer is 1/3, because 2 is the only even number smaller than 4. What did we do? This is actually

$$\frac{1}{3} = 2 * \frac{1}{6} = \frac{1/6}{1/2} = \frac{P[A \cap B]}{P[B]}.$$

We have to divide by the "$P[B]$" (given B), because we want a probability again (must sum up to 1). That's why you thought automatically that the probability for an elementary outcome is 1/3 and not 1/6 in the new sample space. Therefore we define

Definition 1.3 [Conditional Probability of A given B: $P[A|B]$] *Suppose $P[B] > 0$; we define*

$$P[A|B] := \frac{P[A \cap B]}{P[B]}.$$

So called: "Conditional Probability of A given B".

We have:

$$P[A|B]P[B] = P[A \cap B] = P[B|A]P[A].$$

It can be shown that $P[.|B]$ is a probability too (check Definition 1.1).

Let $B \subset A$. Show: $P[A|B] = 1$. What is the intuitive interpretation of this result?

Bridge to 1.7; examine $P[A|B]$, if $P[A \cap B] = P[A]P[B]$.

1.7 Independent Events

Definition 1.4 [Independent Events] *Events A and B are independent of each other, if*

$$P[A \cap B] = P[A]P[B].$$

Given $P[B] > 0$ (or $P[A] > 0$ respectively), independence means

$$P[A|B] = P[A]$$

(or $P[B|A] = P[B]$ respectively).

$P[A \cap B] = P[A]P[B]$ is most widely used, but $P[A|B] = P[A]$ can be interpreted better than $P[A \cap B] = P[A]P[B]$:

Compute probability that first 4 children born are boys:

Modelling of experiments: you must decide yourself, looking at your experiment, whether you want to and can assume, that events are independent of each other. For example, you have 10 mice. If first one dies, does that change the probability that the second dies too? This is not to be mixed up with low survival probability in general. If you give all mice a lethal dosis, they all die. But number 2 does *not* die because number 1 died. They die independently of one an other.

1.8 Bayes theorem - an important medical application

Example from Cartoon Guide, p 46-50:

* A: patient has disease
* B: test is positive
* $P[A] = 0.001$ [1 patient in 1000 has the disease]
* $P[B|A] = 0.99$ [positive test given infection is 99 %]
* $P[B|A^c] = 0.02$ [positive test although not infected is 2 %]

$$P[A|B] = ? \quad \text{[Probability I have the disease given I tested positive]}$$

can we compute this last expression?

Bayes theorem actually reads:

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{P[B|A]P[A]}{P[B|A]P[A] + P[B|A^c]P[A^c]}.$$

1.9 Exercises (more in course)

Combinatorics:

1.1 Verify:

$$n\binom{n+k}{k} = (k+1)\binom{n+k}{k+1}$$

1.2 Verify:

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$$

1.3 How many guest are present if pairs of two shake hands 253 times (assume all shake hand with everybody once).

1.4 If you have 30 strait lines, how many points of intersections can you have at the maximum?

Probability:

1.5 Probability for 40 y old man to die in next y is 0.002. 20 men of age 40 meet. How high is the probability that at least one of the 20 dies in the coming y?

1.6 Events A, B and C. Show:

$$P[A \cap B \cap C] = P[A]P[B|A]P[C|A \cap B].$$

1.7 Looking at families with 3 children, how high is the probability that we choose a family with 2 girls and 1 boy if we choose randomly with equal probabilities amongst families with 3 children?

1.8 There are 6 red, 4 green and 2 yellow balls in a box. You take two out of the box randomly. What is the probability that the 2 balls have the same color?

Independent Events:

1.9 Playing Monopoly with pair of dice (red and blue). C is the event that red comes up "1", D is the event that blue comes up "2". Show that C and D are independent of each other.

1.10 Playing Monopoly with pair of dice (red and blue). C is the event that red comes up "1", D is the event the sum is 3. Show that C and D are not independent of each other.

1.11 Given there are 300 students in a class. 200 are studying biology, 80 geography and 20 an other subject. You select a person randomly. How high is the probability that this person is studying biology and is born on a Sunday?