Introduction to statistical inference 1

Lauri Mehtätalo
University of Eastern Finland
School of Computing

January 5, 2018

Contents

Chapter 1

Preliminaries

1.1 Introduction

- Field of statistics builds on probability theory

> *"You can, for example, never foretell what any one man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So says the statistician."* - Sherlock Holmes

- The paragraph includes the important ideas of the statistical model:
  - The percentage p = the model of the process or underlying population
  - The behavior of individuals = data
- Assuming a constat probability p may be a too simplistic or naive assuption, and may be replaced by more realistic one where p is a function of the propabilities of the individual and the context where s/he is.
- We also need to specify a model, for the variability of the individuals around the p to complete the model formulation.
  - A crude summary if the variance-covariance matrix of the observations.
  - A complete definition is done by specifying the joint distribution of all individuals.
- We also may want to estimate how accurately we finally estimated p writing the available data
- The theoretical process that generates the data is called
  - Statistical model or (tilastollinen malli)

  - Stochastic process or (stokastinen prosessi)
  - Random process or (satunnaisprosessi)

- The process is random/stochastic because the "man" do not behave exactly according to model.[1]
  - Probability calculus and the theory of random variables provide tools to formulate and understand such models.

- Once model has been formulated or specified (muotoiltu), observed data can be used to[2]
  - estimate model parameters
  - evaluate the model fit (mallin sopivuus)
  - evaluate the inaccuracy related to the estimated model parameters

- When talking about models, we can talk about
  - True model (Tosi malli)
  - Estimated model (Estimoitu malli)
  - True model always stays the same, but as data used to formulate the estimated model gets larger, the estimated model gets closer to true model.
  - See example R-scipt `regsimu.R`

1.2 Set theory

- Consider a statistical experiment (e.g. rolling a dice, measuring the diameter of a tree, tossing a coin, measuring the photosynthetic activity in plant etc.)

Definition 1.1. All possible outcomes of a particular experiment (koe) form a set (joukko) called sample space (otosavaruus), denoted by S. For example:

A Toss of a coin; $S = \{H, T\}$

B Reaction time, Waiting time; $S = [0, \infty)$

C Exercise score of this course; $S = \{0, 1, 2, \dots, 210\}$

D Number of points (events) within fixed area; $S = \{0, 1, 2, \dots$

E CO_2 uptake within 0.5 hours in fixed area plot; $S = (-\infty, \infty)$

[1] This is what is done on this part of the course (ISI1).
[2] This is what is done on second part of the course (ISI2)

F Waiting time up to one hour (in minutes); $S = [0, 60)$

Sample space can be countable (numeroituva) or uncountable (ylinumeroituva). If the elements of a sample space can be put into one-to-one correspondence with a finite subset of integers, the space is countable. Otherwise, it is uncountable.

- Note: Examples A and C before are countable, the others are uncountable
- Note: If the waiting time in G are rounded to the minute / second / millisecond / microsecond, the sample space becomes countable.

Definition 1.2. An event (tapaus) is any collection of possible outcomes of an experiment, meaning it is a subset of S. Event A is said to occur, if the outcome of the experiment is in set A.

Example 1.1. Draw a card from standard deck.

$$S = \{\heartsuit, \diamondsuit, \clubsuit, \spadesuit\}$$

One possible event is $A = \{\heartsuit, \diamondsuit\}$. Another possible event is $B = \{\diamondsuit, \clubsuit, \spadesuit\}$. The union (unioni) of the two events includes all elements of both

$$A \cup B = \{\heartsuit, \diamondsuit, \clubsuit, \spadesuit\}$$

The intersection (leikkaus) includes elements that are common to both events

$$A \cap B = \{\diamondsuit\}$$

The complement (komplementti) of a set includes al elements of S that are not included in A

$$A^c = \{\clubsuit, \spadesuit\}$$

Events A and B are said to be disjoint (erillisiä), if

$$A \cap B = \emptyset$$

Number of events $A_1, A_2, A_3, \ldots$ are said to be pairwise disjoint, if

$$A_i \cap A_j = \emptyset$$

for all pairs of i, j. In addition, if $\bigcup_{i=1}^{\infty} A_i = S$, then $A_1, A_2, A_3, \ldots$ defines a partition of the sample space.

Example 1.2. Events $A = \{\heartsuit, \diamondsuit\}$ and $B = \{\clubsuit, \spadesuit\}$ are disjoint since

$$A \cap B = \emptyset$$

Events $A_1 = \{\heartsuit, \diamondsuit\}$, $A_2 = \{\clubsuit\}$ and $A_3 = \{\spadesuit\}$ are pairwise disjoint. Also, since

$$\bigcup_{i=1}^{3} = A_1 \cup A_2 \cup A_3 = \{\heartsuit, \diamondsuit, \clubsuit, \spadesuit\} = S$$

they are also partiotion of S.

Definition 1.3. Probability (todennäköisyys)

If a certain experiment is performed number of times (or infinite number of times), it may lead to different outcome, which is an event of the sample space. This frequency of outcome of an event is called probability.

For an event $A \subset S$ in an experiment, notation $P(A)$ (or $Pr(A)$) specifies the probability of outcome / event A.

Theorem 1.1. Axioms of probability

1. For every event A, $P(A) \neq 0$ (meaning every event is possible)
2. $P(S) = 1$ (because something will be observed)
3. For a sequence of pairwise disjoint events,

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Example 1.3. Assume we have two events $A_1, X \in A_1$ and $A_2, X \in A_2$, which have probabilities $P(A_1) = 0.2$ and $P(A_2) = 0.3$.

If the events are disjoint ($A_1 \cap A_2 = \emptyset$), probability for the union of events is

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) = 0.2 + 0.3 = 0.5$$

If events are not disjoint ($A_1 \cap A_2 \neq \emptyset$), then

$$P(A_1 \cup A_2) \neq P(A_1) + P(A_2) = 0.2 + 0.3 = 0.5$$

Example 1.4. In a fair deck, define events

$$A_1 = \{\heartsuit\}, A_2 = \{\diamondsuit\}, A_3 = \{\clubsuit\}, A_4 = \{\spadesuit\}$$

which have probabilities

$$P(A_1) = P(A_2) = P(A_3) = P(A_4) = 1/4$$

Events $A_1 \ldots A_4$ are disjoint. Therefor,

$$B = A_1 \cup A_2 = \{\heartsuit, \diamondsuit\}$$

$$P(B) = P(A_1 \cup A_2) = P(A_1) + P(A_2) = 1/4 + 1/4 = 1/2$$

Theorem 1.2. Consider events A and B

1. $P(A^c) = 1 - P(A)$
2. $P(B \cap A^c) = P(B) - P(A \cap B)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. If $A \subset B$, then $P(A) \leq P(B)$

Note Consider case 3 of theorem 1.2

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

$$P(A \cup B) \leq 1$$

$$P(A) + P(B) - P(A \cap B) \leq 1$$

$$P(A \cap B) \geq P(A) + P(B) - 1$$

This equation is called the Bonferroni inequality. Idea is, that if we have intersection of two events (A and B), the probability of the intersection can be shown to be higher or equal than the right term.

$$P(A \cap B) \geq P(A) + P(B) - 1$$

Supose A and B are two events that occur with probability $P(A) = P(B) = 0.975$. Then $P(A \cap B) = 0.975 + 0.975 - 1 = 0.95\ldots$

Theorem 1.3. a) $P(A) = \sum_{i=1}^{\infty} P(A \cap C)$ for any partition $C_1, C_2, \ldots$

b) $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$, for any sets $A_1, A_2, \ldots$

Definition 1.4. IF A and B are events in a sample space S and $P(B) > 0$, then conditional probability (ehdollinen todennäkoisyys) of A given B is

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

Example 1.5. Clinical trial. Assume that we know the probabilities that are represented in table 1.1. Let event A be the event "the patient recovered" with sample space $S = \{\text{"}OK\text{"}, \text{"}NOT\ OK\text{"}\}$, and event B is the event "The patient was treated with placebo" with $S = \{\text{"}YES'', \text{"}NO''\}$.

Now $P(A \cap B)$ can be computed directly using the table 1.1, and it is

$$P(A \mid B) = P(\text{"}OK'' \mid \text{"}Placebo'') = \frac{P(A \cap B)}{P(B)} = \frac{0.160}{0.227} \approx 0.70$$

Table 1.1: Probabilities related to clinical trial. Four different drugs and the probabilities that patient got or did not got ok after using it.

	Drug1	Drug2	Drug3	Placebo	Total
OK	0.120	0.087	0.147	0.160	0.513
NOT OK	0.147	0.167	0.107	0.067	0.487
Total	0.267	0.253	0.253	0.227	1.000

Definition 1.5. Bayes rule

Assuming we have event A and B,

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

if $P(B) > 0$

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

And the other way around

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

if $P(A) > 0$

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

and by using the earlier $P(A \cap B) = P(B)P(A|B)$

$$P(B)P(A|B) = P(A)P(B|A) \;\; \| : P(B)$$

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

This is known as Bayes rule.