Inference as a dynamic concept map

Sytse Knypstra *

Abstract

Many students find it hard to understand the fundamental concepts in inferential statistics. This becomes apparent when students are mixing up population parameters and their estimates. Reasons why inferential statistics is so hard to understand are given and concept maps are introduced as a way to facilitate understanding. Two computer programs are described: 'Sila', which supports teaching of inferential statistics by means of a dynamic concept map and 'PQRS', a probability calculator that can help studying distributions.

Introduction

Many students have difficulties in understanding concepts in classical inferential statistics. This becomes evident when they write down expressions like:

- $H_0 : \bar{x}_1 = \bar{x}_2$,
- a 95% confidence interval (for $\bar{x}$) is $[\mu - 1.23; \mu + 1.23]$,
- the probability that μ lies in the interval [1.23; 4.56] is 95%,
- (if σ is unknown): the test statistic is $\dfrac{\bar{X}}{\sigma/\sqrt{n}}$.

In these examples the student seems to be unable to distinguish between *parameters* (in non-Bayesian statistics considered as non-random quantities whose values are not known) and sample *statistics* (random variables whose outcome may vary from one sample to another, but which can be evaluated for the particular sample at hand). But also the interpretation of other concepts like 'p-value', 'power' and 'confidence interval' provide serious problems for students. The fact that students find it hard to understand these concepts is widely acknowledged.

* Faculty of Economics, University of Groningen, PostBox 800, 9700 AV Groningen, The Netherlands. Tel. ++31 50 3633808. e-mail: `S.Knypstra@eco.rug.nl`

Lipson (1994) notes that the idea of a sampling distribution is generally poorly understood. She ascribes this to the fact that it is "introduced using a deductive approach based on probability theory". Furthermore, "it is a theoretical development which is difficult to relate to the physical process of drawing a sample from a population". She argues that a theoretical explanation should be accompanied by an empirical argument.

Hawkins et al. (1992) claim that the words 'population' and 'sample' in the framework of statistics have a very specific meaning that differs from the everyday usage of the words. This will cause confusion in students. The authors continue:

> Almost before the student has come to terms with that mental step, the realm of the random variable has been entered, where populations (and samples) are defined in terms of distributions, the population being described by its *parameters* and the sample by its *statistics*. Close on the heels of these concepts comes the idea of *sampling distributions*, being the derived distributions of statistics from repeated sampling from population distributions. It is hardly surprising that students regularly confuse these three kinds of distribution, especially when their descriptors have so much in common, μ and $\bar{X}$ to describe location, both being referred to as *means*, and σ^2 and s^2 the *variances* of population and sample, respectively. The nightmare really begins, however, with the introduction of the term standard *error* to represent the standard *deviation* of the sampling distribution. There is much to be said for a presentation that colour-codes population, sample and sampling distributions, and which emphasizes the distinctive use of the terms 'parameter' and 'statistic'.

In our opinion a key role is played by the concepts 'distribution' and 'random variable' which should be fully understood before advancing to inferential statistics. We completely agree with Hawkins et al. (1992) that the use of the word 'mean' for the expectation of a (population) distribution, the (stochastic) mean $\bar{X}$ of a sample and its outcome $\bar{x}$ is most confusing. Similarly terms like variance, standard deviation and median are used in three different ways. Although textbooks tend to distinguish between 'estimator' and 'estimate', generally the same word is used for a confidence interval as a stochastic interval and for its outcome. We recommend to avoid expressions in which a random variable is linked with its outcome by means of the equal sign. Instead of 'The test statistic is: $\dfrac{\bar{X}-\mu_0}{S/\sqrt{n}} = 1.23$.' we prefer: 'The test statistic is: $\dfrac{\bar{X}-\mu_0}{S/\sqrt{n}}$. Its outcome is 1.23.'

Concept maps

One way to deal with the complexity of inferential statistics is to use a concept map. A concept map consists of boxes representing the concepts and arrows representing their relations. A concept map shows the structure of a complex concept in a schematic, graphical way. A concept map is not uniquely defined, different authors may or may not agree on the (number of) relevant concepts and the links between them.

In a typical introductory course in statistics we consider the following concepts to be the most fundamental for inferential statistics: random variable, probability distribution, parameter, population, random sample and its realization, statistic and its realization. Other concepts, more specifically concerned with estimation, confidence intervals and testing hypotheses, are: (un)biasedness, null hypothesis, alternative hypothesis, level of significance α, critical value and critical region, p-value, power(-function). For the moment however we limit ourselves to the first mentioned generic inferential concepts and to the situation that one or more independent populations are involved (the treatment of e.g. linear regression needs some adaptations). Figure 1 shows a first, tentative concept map describing them.

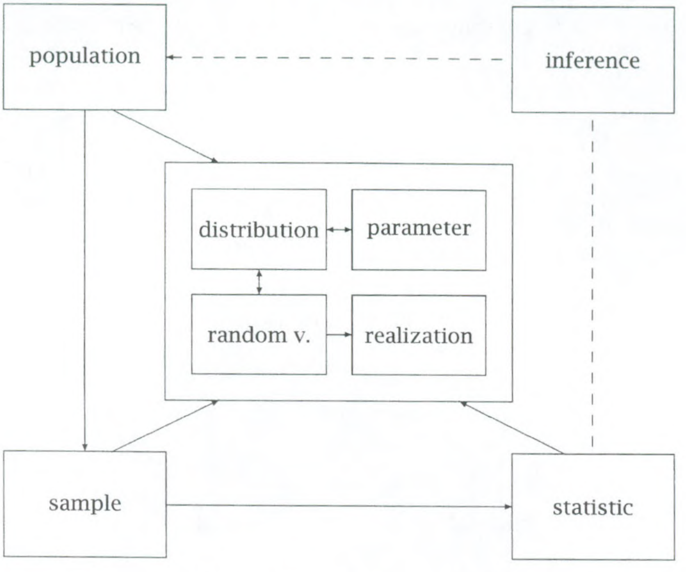

Figure 1: Concept map of inferential statistics.

Central to all other concepts in figure 1 is what we will call the distribution cluster. It consists of $\boxed{\text{distribution}}$, $\boxed{\text{parameter}}$, $\boxed{\text{random v.}}$ which stands for both random variable and random vector, depending on the univariate or multivariate character of the associated distribution, and $\boxed{\text{realization}}$ (outcome). The first two concepts are linked by a two-way arrow: a given probability distribution determines the value of its parameter(s) and, given the type of distribution, its parameter values completely specify the distribution. For each $\boxed{\text{distribution}}$ a $\boxed{\text{random v.}}$ can be defined having this distribution, therefore also the link between these two concepts is two-way. The last two concepts in the distribution cluster are linked by a one-way arrow: from each $\boxed{\text{random v.}}$ a $\boxed{\text{realization}}$ can be obtained.

Each of the three concepts $\boxed{\text{population}}$, $\boxed{\text{sample}}$ and $\boxed{\text{statistic}}$ is closely connected with the distribution cluster: each of them is linked with a distribution determined up to one or more parameters. The leftmost vertical arrow in figure 1 represents the process of drawing a (random) sample from a population. The lowermost horizontal arrow indicates the summarizing of the sample into a statistic. The dashed lines, together with $\boxed{\text{inference}}$ indicate that a statement about a population is made on the basis of the statistic's outcome.

How can concept maps help in teaching? They help to clarify the concepts and their (obscured) links. In inferential statistics a concept map can be used to show that a $\boxed{\text{statistic}}$ is an instance of a $\boxed{\text{random variable}}$, which has a $\boxed{\text{distribution}}$ depending on a $\boxed{\text{parameter}}$ and hence that its outcome can be seen as drawn from this distribution. Concept maps also show in which context a word like 'mean' is used.

The concept map of figure 1 is rather complex and therefore not suitable for educational purposes. In a simplified version we consider the $\boxed{\text{distribution-cluster}}$ as a well-known basic concept which may be omitted. Then only four boxes remain. In figure 2 this reduced concept map is shown, but some essential parts of the distribution cluster are transferred to the remaining boxes: $\boxed{\text{population}}$ now contains the parameter θ and $\boxed{\text{sample}}$ and $\boxed{\text{statistic}}$ are each divided into two parts, a stochastic part and a realization part.

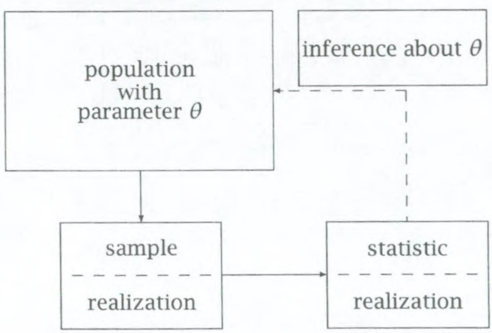

Figure 2: Reduced concept map of inferential statistics.

When dealing with one or more independent populations this reduced concept map is generic for estimation, testing hypotheses, and constructing confidence intervals. Its elements may be refined and further specified:

- When referring to 'a population' we usually have one distribution in mind, e.g. a normal distribution with parameters μ and σ^2. But also more complicated cases fit into this framework. When discussing nonparametric statistics, F can be considered as the parameter of interest, where F represents the cumulative distribution function of some continuous or discrete distribution. When making inference on the basis of k independent populations each $\boxed{\text{population}}$

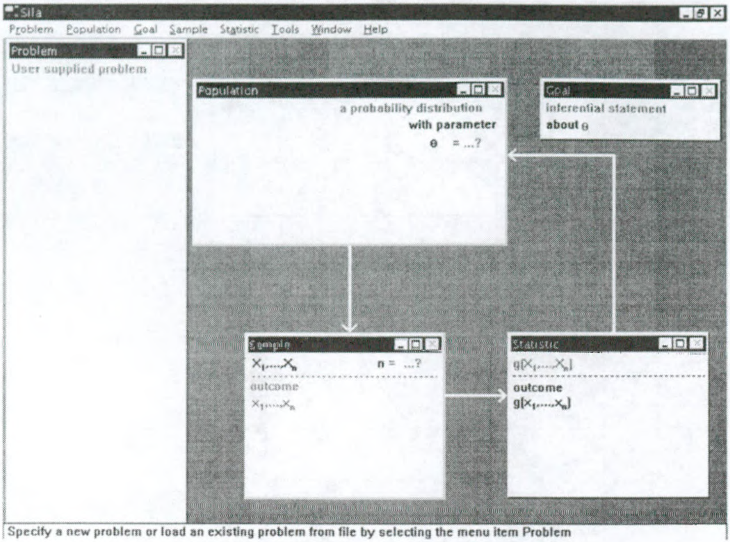

Figure 3: Initial screen after starting Sila.

may be replaced by a number of boxes, one for each population from which samples are drawn. In each box the type of distribution is displayed, together with the parameter(s) about which inference is to be made.

- Sample should be similarly replaced by the same number of boxes. In the random variable section each of the observed variables are displayed, e.g. $X_1, \ldots, X_9$. In the realization section their outcomes are shown after taking a sample.

- In the random variable section of statistic the summarizing function, e.g. $\bar{X} = g(X_1, \ldots, X_9)$ can be written. Its realization section will contain the outcome after taking a sample. In the case of confidence intervals *two* summarizing functions and *two* outcomes have to be specified, one for each endpoint.

- Inference may contain the sort of inference (estimation, testing hypotheses, constructing confidence intervals) and the parameter(function) or hypotheses of interest.

Specifying or changing the contents of one box immediately changes the contents of other boxes. Especially these changes show how the various components of a concept map interact.

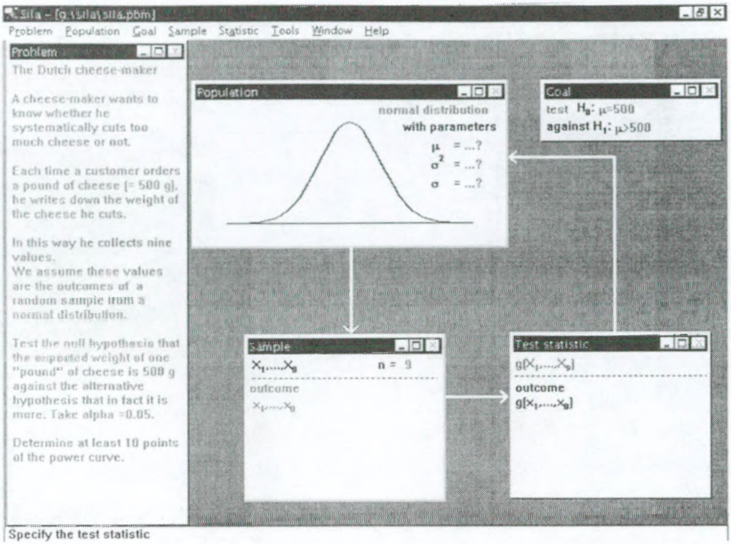

Figure 4: Sila after specifying population, goal and sample size.

Sila, a dynamical concept map

A concept map, if drawn on paper, is static in its nature. It shows neither the effects of drawing samples, nor the effects of specification or change in its boxes. A dynamical concept map should show how the contents of each box are updated if some element is changed or if a sample is drawn. A dynamical concept map should enhance the student's understanding of the inferential process because the consequences of a change in certain characteristics and of drawing a random sample are made visible.

The computer program 'Sila' (Statistical inference laboratory) was developed with the aim of offering students such a dynamic concept map. The concept boxes were implemented as windows, thus allowing for sizing/zooming, dragging, putting them on top or in the background. A copy of the initial screen, which is in fact a direct translation of figure 2, is shown in figure 3.

A window's contents can be specified or changed through the menu or by clicking on 'sensitive' portions of the text. Sample outcomes can be entered simply by typing them in or by letting the computer draw them at random from the specified population(s). There are in fact two modes of operation: either an inferential problem is supplied by the user or the problem is chosen from a file of predefined problems. Such a file can be created by the teacher using Sila. In the first case the user is free to choose all its characteristics whereas in the second case each choice has to match the predefined characteristics. Context sensitive help is supplied as well as feedback when erroneous choices are made.

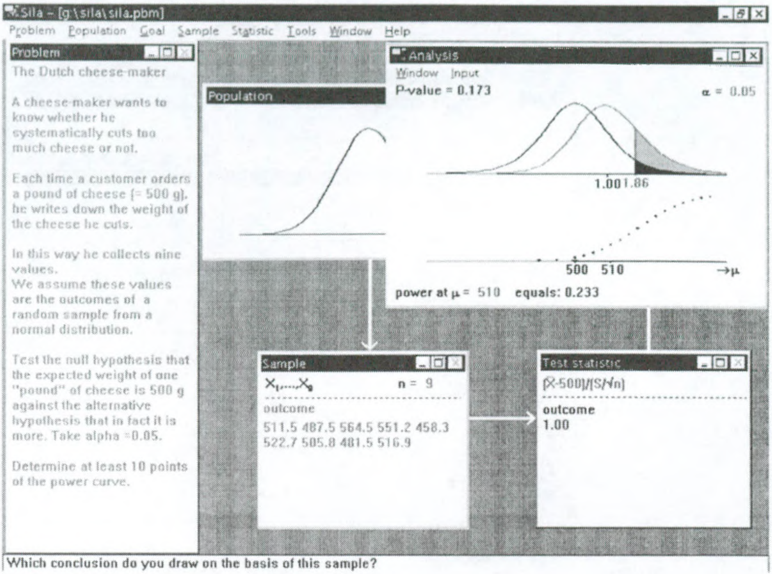

Figure 5: Power(function) in Sila.

Figure 4 shows the screen after a problem has been loaded from a file and the user has entered the type of population distribution (normal), the type of inference (testing hypotheses), the null- and alternative hypothesis and the sample size.

Figure 5 shows the same problem when the user has drawn a sample and additionally specified the test statistic and the level of significance. In the 'Analysis' window the graph of the test statistic's distribution is displayed under the null-hypothesis and under a specified parameter value of the alternative hypothesis, thus showing the probability of a type II error and the power. After entering successively several parameter values the contour of the power function becomes visible. Students especially appreciate this part of Sila.

Figure 6 shows a problem involving two populations and thus two samples. It shows the way parameter functions and statistics are edited using a formula editor. The problem in this figure was taken from Moore & McCabe (1989).

PQRS

The program PQRS (Probabilities, Quantiles and Random Samples) was developed as a tool with Sila that would make statistical tables obsolete. It can also be used independently from Sila. For some 25 discrete and continuous distributions (a.o. the non-central chi-square, *t* and *F*- distributions) PQRS renders:

- a graph of the probability density function or probability mass function,

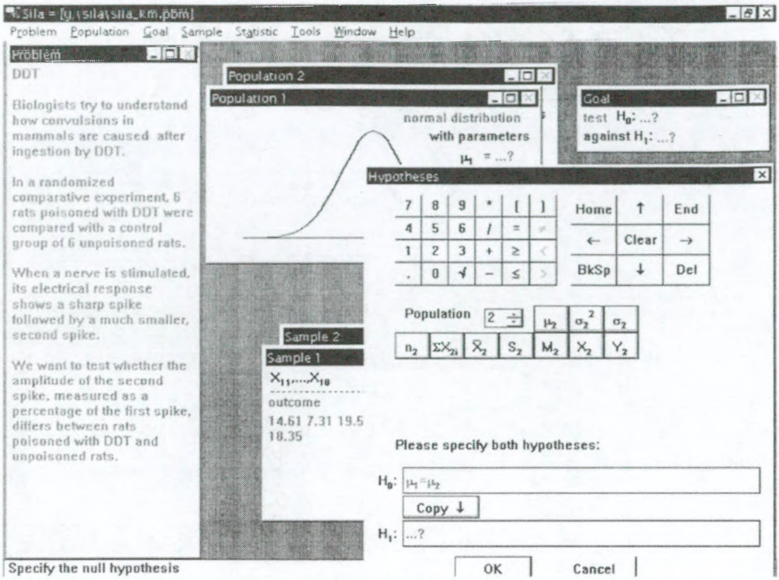

Figure 6: Formula editor in Sila.

- a graph of the cumulative distribution function,
- probabilities $P(X < \text{value})$, $P(X > \text{value})$ and – if nonzero – $P(X = \text{value})$, all in their natural position relative to the graph,
- quantiles (the application of the inverse cumulative distribution function) for a given value between 0 and 1, also in their natural position relative to the graph,
- random samples which can be saved to a file,
- formulas for the density function, the expectation and variance and the moment generating function.

Great care has been given to a clear design of the user interface. PQRS not only makes tables obsolete, but it also gives the students a good idea of the shape and other characteristics of probability distributions. Figure 7 shows a window with PQRS running.

Evaluation

Sila and PQRS are not meant to replace traditional teaching of statistics, but should be used to illustrate certain concepts in the classroom or to let the students work out problems in a hands-on computer session. Sila and PQRS are used by students

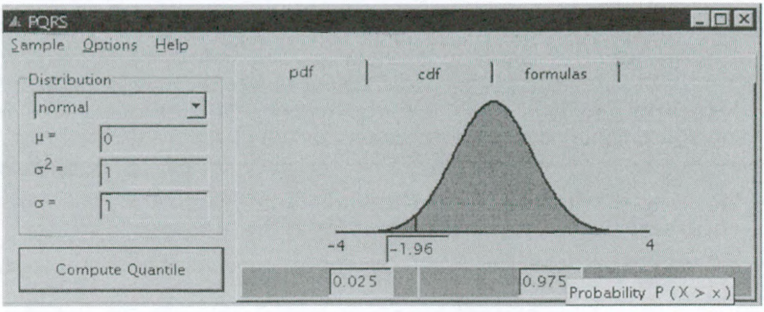

Figure 7: PQRS running.

in econometrics at the University of Groningen in their first year's course of statistics. During four weeks a one hour session is held each week in which the students have to solve a number of problems with the aid of Sila and PQRS. Sila appears to be a valuable tool in teaching the fundamental concepts of statistics when accompanied by special instructions and questions. These are necessary to make the students aware of otherwise unnoticed aspects. Our positive experience with Sila is confirmed by the results of evaluations. The most recent evaluation states that:

- Sila is easy to use (93% of the students),
- Sila has enhanced the students' understanding of inference (67%),
- Some students (10%) feel that feedback comes too soon; in that case they are not forced to think but can use a trial and error strategy in handling a problem.

Some remarks and technical details

Both Sila and PQRS run under Windows 95 and Windows NT 4.0. Sila also runs under Windows 3.1. Sila was programmed using Turbo (Borland) Pascal for Windows whereas PQRS was programmed using Delphi 3.0. A conversion of Sila using Delphi is planned. The distributions in both programs are laid down in a system of Turbo Pascal/Delphi objects (cf. Knypstra (1997)). Some algorithms were used from Press et al. (1989); the random number generator was taken from L'Ecuyer (1988) and many algorithms for drawing samples from non-uniform distributions were found in Devroye (1986).
One of the possible choices for a population distribution in Sila is a self-defined discrete distribution, which could in fact be a random sample from some other distribution. By drawing many samples with replacement from such a 'population' the principles of bootstrapping can be taught.

A student's version of 'Sila' and a complete version of PQRS can be downloaded from the WWW pages: `http://www.eco.rug.nl/medewerk/knypstra/sila.html` and `http://www.eco.rug.nl/medewerk/knypstra/pqrs.html`.

Recommendations

- Teachers, textbooks and software should use different words for different concepts, e.g. 'expected value' or 'expectation' when referring to a population or distribution, 'sample mean' for $\bar{X}$ and 'outcome of the sample mean' for $\bar{x}$. Random variables and their realizations should be clearly distinguished. This also applies to a confidence interval vs. the outcome of a confidence interval.
- The concepts of random variable and probability distribution should by thoroughly understood by students before advancing to inferential statistics.
- Sila and PQRS should be used to help students understand inferential statistics and probability distributions.

References

DEVROYE L. (1986) *Non-Uniform Random Variate Generation.* Springer, New York.

HAWKINS, A., F. JOLIFFE, L. GLICKMAN (1992) *Teaching Statistical Concepts.* Longman, London, New York.

KNYPSTRA S. (1997) Kansverdelingen en objecten. *Kwantitatieve Methoden 54.*

L'ECUYER P. (1988) Efficient and Portable Combined Random Number Generators. *Communications of the ACM, Vol. 31, Nr 6.*

LIPSON K. (1994) Understanding the role of computer based technology in developing fundamental concepts of statistical inference, *Proceedings of the Fourth International Conference on Teaching Statistics,* Marrakech, Morocco.

MOORE, D.S., G.P. MCCABE (1989) *Introduction to the practice of statistics.* W.H. Freeman and Company, New York, Oxford.

PRESS, W.H., B.P. FLANNERY, S.A. TEUKOLSKY, W.T. VETTERLING (1989) *Numerical Recipes in Pascal: the art of scientific programming.* Cambridge University Press, Cambridge.

Ontvangen: maart 1998
Geaccepteerd: januari 1999