Informal Inferential Reasoning: a Computer-based Training Environment

Joachim Engel[1,3], Tim Erickson[2]
[1]Ludwigsburg University of Education, Ludwigsburg, Germany
[2]Epistemological Engineering Oakland, California, USA
[3]Corresponding author: Joachim Engel, e-mail: engel@ph-ludwigsburg.de

Abstract

The logic behind statistical inference is difficult for students to understand. Recent research in statistics education focuses on learners' informal and intuitive ideas of inferential reasoning rather than on the mastery of formal mathematical procedures. We introduce a computer-based training environment (a "data game") to shape intuition for inference in the context of a particular type of statistical decision problem. In change-point detection tasks, one must decide if a process is running smoothly or if it is out of control. In our data game a mechanism produces data sequentially with a likely built-in shift in location at random time. The task for the students is to detect *if* and *when* the mechanism has changed the level of produced data as early as possible but without raising false alarms. The data game is embedded in a data analysis environment. We discuss the relation between change point detection and informal inference, the game itself as well as its relation to inferential reasoning.

Keywords: Statistics education, intuitive reasoning, statistical literacy, technological learning environments

1. Introduction

Most of our knowledge about the empirical world is based on careful generalizations from observations to a wider universe. Methods of statistical inference have a central role in evidence-based knowledge acquisition; they are used in all of the empirical sciences to ensure the legitimacy of new knowledge. The mathematical methods created for that purpose are based on advanced concepts of probability in combination with different epistemological positions grown in the history of scientific reasoning.
For several decades, as Zieffler, Garfield & delMas (2008) explicate, psychologists and education researchers have studied and documented the difficulties people have making inferences about uncertain outcomes. Researchers have pointed to various reasons for these difficulties including: the logic of statistical inference, students' intolerance for ambiguity (Carver 2006), and students' inability to recognize the underlying structure of a problem. Other research has suggested that students' incomplete understanding of statistics such as distribution, variation, sampling, and sampling distributions may also play a role in these difficulties. Some recent studies (e.g., Haller & Krauss 2002) indicate that even empirical researchers do not always understand these methods or use them correctly. As far as teaching in school is concerned, the issue is not to teach complex formal methods of inferential statistics but to initiate a way of thinking that helps to make sound decisions in a world of uncertainty where progress in many areas is characterized by the analysis of empirical data.

After summarizing some major aspects of the recent discussion of informal inferential reasoning (IIR) in Section 2, we argue that change point detection tasks constitute an ideal environment to explore students' IIR (Section 3). We then present the data game environment *Epidemic,* discuss its design and how it provides informal experiences to strengthen sound intuition in situations of decision-making under uncertainty.

2. Informal Inferential Reasoning

Inferential statistics is the scientific method for evidence-based knowledge acquisition. However, the logic behind statistical inference is difficult for students to understand. Statistical inference requires going beyond the data at hand, either by generalizing the observed results to a wider universe or by drawing a more profound conclusion about the relationship between the variables, e.g., determining whether a pattern in the data can be attributed to a real effect in a causal relationship. The recent literature (e.g., Pratt & Ainley 2008; Pfannkuch et al. 2012) discusses different concepts of an introduction to inductive reasoning, which precedes a formal treatment. They are based on the assumption that as early as in the middle grade classes (Biehler 2007; Watson 2008), basic concepts of statistical inference can be made accessible when students reflect on and evaluate arguments that are based on an analysis of data. The implied focus is on considerations of variability in data and an assessment of whether different characteristics are due to systematic effects or to random variability. Variation is the reason why complex methods of statistical inference were devised. A helpful perspective in understanding statistical inference problems is based on the signal-noise-metaphor (Konold and Pollatsek 2002). Data are perceived as a composite of a signal and noise. The challenge of any data analysis is to search or reconstruct the signal from the data corrupted by noise. In a data-drenched world statistical literacy—particularly the capacity to identify such signals or messages—is essential. Inference can be seen as the conclusion of reasoning from the data to the signal.

Given the importance of understanding and reasoning about statistical inference, and the consistent difficulties students have with this type of reasoning, there have been attempts to expose students to situations that allow them to use informal methods of making statistical inferences (e.g., comparing two groups based on boxplots of sample data). Research in this field begins to offer insight into the learners' inferential reasoning and how that thinking might be shaped more effectively by well-designed tasks. Rubin, Hammerman, and Konold (2006) define informal inferential reasoning as reasoning that involves the related ideas of properties of aggregates (e.g., signal and noise, and types of variability), sample size, and control for bias. Pfannkuch (2006) defines informal inferential reasoning as the ability to interconnect ideas of distribution, sampling, and center, within an empirical reasoning cycle. Rossman (2008) has described informal inference as "going beyond the data at hand" and "seeking to eliminate or quantify chance as an explanation for the observed data" through a reasoned argument that employs no formal method, technique, or calculation.

Combining these perspectives, we follow Zieffler's et al. (2008) working definition of informal inferential reasoning as the way in which students use their informal statistical knowledge to make arguments to support inferences about unknown populations based on observed samples. Furthermore, Zieffler et al. (2008) propose a framework for informal inferential reasoning with the following components:

- Making judgments, claims, or predictions about populations based on samples, but not using formal statistical procedures and methods such as *p*-value, *t*-tests;
- Drawing on, utilizing, and integrating prior knowledge (e.g., formal knowledge about foundational concepts, such as distribution or average; informal knowledge about inference such as recognition that a sample may be surprising given a particular claim; use of statistical language), to the extent that this knowledge is available; and
- Articulating evidence-based arguments for judgments, claims, or predictions about populations based on samples

In addition to these ideas Reading (2007) points out that informal inferential reasoning also includes ideas of choosing between competing models, expressing a degree of uncertainty in making inference, and making connections between the results and the problem context

3. Change point detection and informal inference

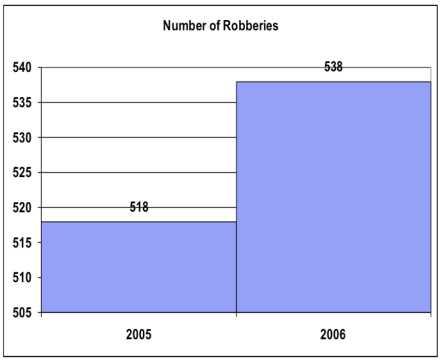
Change point detection problems can be characterized as follows: Suppose one accumulates independent observations from a process that is in a certain state. Observations vary around a certain mean. After a while some of the observations seem to be a bit unusual, too high or too low. Has something occurred that altered the state of the system (a "breakdown"), or are these observations within the range of the expected when acknowledging random variation? Should one declare that a change took place ("raise an alarm") as soon as possible? A false alarm costs resources, credibility etc., but NOT raising alarm when, for example, in fact a new health hazard occurs, may be even more harmful. When we want to detect the change quickly, any detection policy gives rise to the possibility of frequent false alarms when there is no real change. On the other hand, attempting to avoid false alarms too strenuously leads to long delays between the time of occurrence of a real change and its detection.

Practical examples of this problem arise in areas such as health, quality control and environmental monitoring. For instance, consider surveillance for congenital malformations in newborn infants. Under normal circumstances, the percentage of babies born with a certain type of malformation has a more or less known value p_0. Should something occur (such as an environmental change, the introduction of a new drug to the market, etc.) the percentage may increase (e.g., the thalidomide episode of the 1960s). One would want to raise an alarm as quickly as possible after a change takes place, while trying to control the risk of a false alarm. Generally, this type of problem arises whenever surveillance is being done.

In most practical situations, the question about a possible change point is a sequential decision problem: while the data become sequentially available, a decision has to make a trade-off between the risks of false alarm, misdetection and a detection delay. The input is usually given "online", i.e., in piece-by-piece or serial fashion without having the entire data available from the start as in contrast to an offline decision problem where all the data from the beginning to the end are available and the decision regarding a change point is done in retrospect. From a formal mathematical and decision-theoretic perspective, extensive research has been done in this field during the last few decades to derive optimal decision rules (see e.g., Pollack 1985 or Tartakovsky and Moustakides 2010 and the literature therein).

We consider change point detection tasks as an ideal scenario to investigate learners' IIR. These types of problems have been used previously by Rubin, Hammerman and Konold (2006). Engel, Sedlmeier and Woern (2008) investigated responses of 179 pre-service teachers in a context that required them to determine whether a particular change in a process occurred over time or not. One example of these tasks is presented in Figure 1. At a first glance, students usually look at the graphical display and may be tempted to agree with the reporters' claim due to the fact that the bar's height more than doubled from 2005 to 2006. But the statistical issue here is an evaluation of the increase between 2005 and 2006 in light of the fluctuations of the number of robberies over a certain time period. Here – and this may be in contrast to formal procedures of classical hypothesis testing – context knowledge about what may cause or prevent robberies is also an important ingredient in the decision process.

Applied to the change point detection problem the signal is the shift in the mean of the data distribution while the noise is the variation in the data among a mean that is fixed, either before or after the change point. To evaluate the reporters claim in the above example it is not enough just to focus on the sudden increase between 2005 and 2006 but to evaluate that increase in the light of the natural variability of past data. This type of problem fits well into above framework for IIR because the task requires (1) making judgments and claims on the basis of the sample of observations over 7 years, (2) considering variation and the distribution of observed data (integrating prior knowledge including the recognition that a sample may be surprising) and (3) articulating evidence-based arguments for the claim. Further, the task requires a choice between competing models, connects the statistical aspect with the context and elicits multiple and partially contradictory arguments and also encourages expressing the degree of uncertainty in the decision process.



A reporter showed the following graph representing the number of robberies over the last years and commented: “The graph shows that there is a huge increase in the number of robberies between the year 2005 and 2006.” Inquiring about the number of robberies over the last seven years you obtained the following information. Do you agree with the reporters’ statement? Why or why not?

Year	2001	2002	2003	2004	2005	2006	2007
No of Robberies	528	525	499	523	518	538	533

Fig. 1: Change point detection task “Robberies” to elicit students’ informal inferential reasoning

4. A Data Game as a learning environment for Informal Inferential Reasoning

“Data Games” is an innovative project created in cooperation between KCP Technologies in San Francisco, CA and the Scientific Reasoning Research Institute at the University of Massachusetts at Amherst. These two teams developed a set of web-based games and supporting materials that engage students in developing mathematics and data analysis skills. The underlying concept is that students play short games that stream data to a surrounding data analysis environment. To be successful in the game students have to conceive a strategy that requires analyzing the data produced in the process of playing the game. The games are freely available on the web[1]. Erickson (2012) describes the challenges faced in designing and teaching with these games. The purpose of the game *Epidemic*[2] is to provide a learning environment for informal inferential reasoning in the context of change point detection tasks. On each consecutive day of a future month (January 2022), the incidence rate of newly diseased people is shown. These counts have some inherent natural variation, some days the numbers are below and other days they are above average. At some unknown day, the mean shifts to a larger value: This is the day of the outbreak - the epidemic. From that day on the actual number of sick people varies around the new mean. The challenge in the game is to identify the day of the outbreak of the epidemic with as little delay as possible and without raising false alarm (see Figure 2). With data arriving sequentially “on-line”, any observed increase in incidence rate may be a temptation to raise an alarm. But considerations of variation may calm down the excitement and prevent jumping the gun by raising alarm too early. A sound decision is always based on a negotiation of random and systematic influences. Full score is given for clicking STOP (raising the alarm) right on the day of the outbreak of the epidemic; slight penalties are subtracted for each day of delay. Raising a false alarm results in a large penalty.

---

[1] http://www.kcptech.com/datagames/ resp. http://www.srri.umass.edu/datagames

[2] http://www.eeps.com/changepoint/

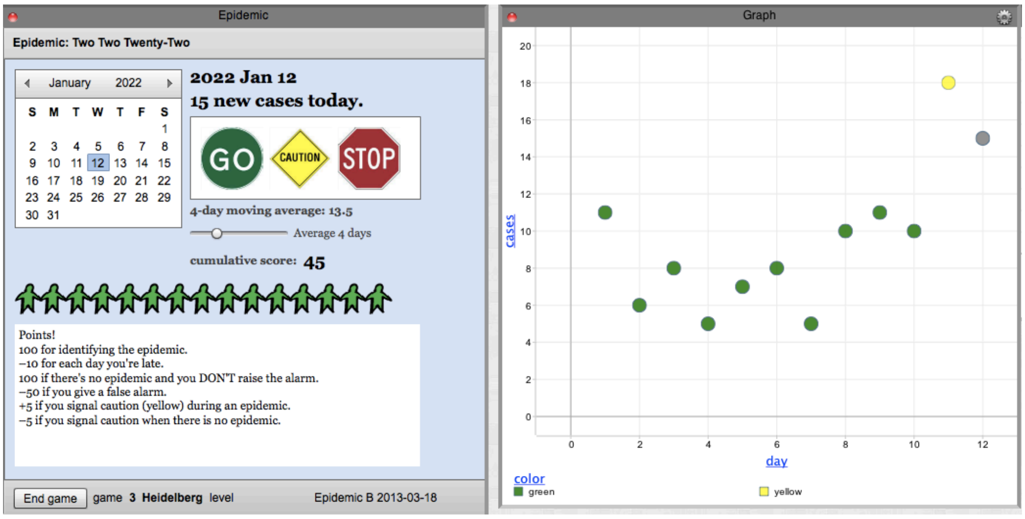


Fig. 2: *Screenshot from day 12 in the data game Epidemic: Can the incidence of 15 new cases be attributed to random variability or is it an indication of the outbreak of an epidemic?*

An attractive context that students can relate to (e.g., poisoned fish served at the school cafeteria as reason for the disease outbreak) together with some animation are important motivators to spend time playing a game. However, from classroom experience with other games of the Data Games project (Erickson 2012), it became obvious that it is equally important not to overload the context with too many details and to keep things as simple as possible. Therefore, in *Epidemic* we did not consider an infectious disease (dependent data) or a change in distribution beyond a simple shift in mean. The degree of difficulty in *Epidemic* depends largely on two variables: the magnitude of the shift of the mean (severity of the epidemic) and the noise level, expressed as variability of the data and measured in standard deviations. The quotient of these two quantities defines a reasonable measure for the signal-to-noise ratio SNR which determines the difficulty of the game. The smaller SNR, i.e. the smaller the signal relative to the noise, the more difficult it is to detect a change point. The game can be played on different levels.

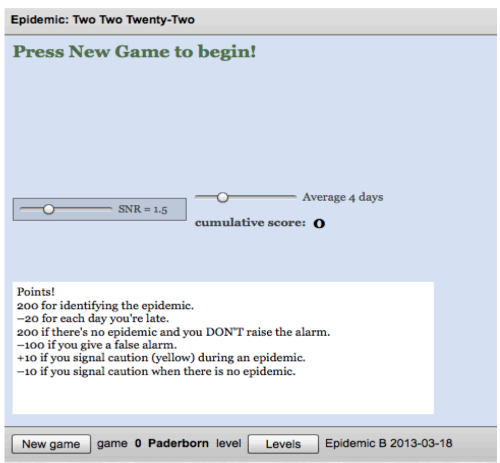


Fig. 3: *At the highest level, Epidemic allows the user to choose the difficulty of the game by selecting a value for SNR.*

The lowest level of the game is intended to familiarize the user with the environment and gaining some first experience with playing the game. Some users may realize here for the first time that not every increase is due to a new epidemic. At this level SNR is fairly large and there will definitely be an outbreak of the epidemic during the time period of a month. At the medium level the SNR is set lower implying that one may easier be tempted to commit a type-I error (misjudging a random deviation as a systematic change, false alarm) or a type-II error (misjudging a systematic deviation as a random change, failing to detect the epidemic). Also, at the medium level there isn't always an epidemic. The highest level of the game lets the user freely choose the level of difficulty by selecting a value for SNR. The attainable score will be adjusted accordingly, i.e. a low SNR implies the chance of earning more points, but also the risk of higher penalties for false alarms (see Figure 3). After getting to know the game at lower levels and having gained experience in detecting change points, it is an important design feature to turn around and experiment with altering game parameters. Learners are guided. The design concept behind freely choosing the SNR level is twofold: it illustrates directly the importance of the signal-to-noise ratio for statistical inference and hence illustrates the concept of data analysis as search for signals in noisy processes. Furthermore, rooted in the psychology of Hans Aebli (1976) this design follows the principle of operative concept formation which intends to stimulate thinking in the framework of acting and constructing a system of operations that will finally serve the acting and concept formation itself.

References

Aebli, H. (1976) *Grundformen des Lehrens. Eine Allgemeine Didaktik auf kognitionspsychologischer Grundlage.,*Klett, Stuttgart.

Biehler, R. (2007) ”Denken in Verteilungen – Vergleichen von Verteilungen.” *Der Mathematikunterricht,* 53(3), 3-11.

Engel, J., Sedlmeier, P., and Woern, C. (2008) ”Modeling Scatterplot Data and the Signal-Noise Metaphor,” In: C. Batanero et al. (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education.* Monterrey, Mexico: ICMI and IASE.,

Erickson, T. (2012) ”Designing games for understanding in a data analysis environment,“ In: *Proceedings of the International Association of Statistical Education (IASE) Roundtable Conference*, Cebu, Philippines, Voorburg ISI.

Haller, H. and Krauss, S. (2002) “Misinterpretations of significance: A problem students share with their teachers?” *Methods of Psycholog. Research*, 7(1), 1–20.

Konold, C., and Pollatsek, A. (2002) “Data analysis as the search for signals in noisy processes,” *Journal for Research in Mathematics Education* 33(4), 259-289.

Pfannkuch, M. (2006) “Informal inferential reasoning,” In: A. Rossman and B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*. Salvador, Brazil, Voorburg, ISI

Pfannkuch, M., Wild, C., and Parsonage, R. (2012),”A conceptual pathway to confidence intervals,” *Zentralblatt Didaktik der Mathematik,* 44 (7), 899 – 911.

Pollak, M. (1985) “Optimal Detection of a Change in Distribution,” *Annals of Statistics*, 13, 206–227.

Pratt, D., Ainley, J. (2008) ”Introducing the special issue on informal inferential reasoning,” *Statistics Education Research Journal* 7 (2), 3-4.

Reading, C. (2007) ”Cognitive development of reasoning about inference”, Discussant reaction presented at the Fifth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-5), University of Warwick, UK.

Rossman, A. (2008) ”Reasoning about Informal Statistical Inference: One Statistician's View,” *Statistics Education Research Journal* 7(2), 5-19.

Rubin, A., Hammerman, J., and Konold, C. (2006) ”Exploring informal inference with interactive visualization software,” In: A. Rossman and B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. Voorburg, ISI.

Tartakovsky, A. and Moustakides, G. (2010) ”State-of-the-Art in Bayesian Changepoint Detection,” *Sequential Analysis*, 29: 125–145, 2010.

Watson, J.M. (2008) ”Exploring beginning inference with novice grade 7 students,” *Statistics Education Research Journal*, 7(2), 59-82.

Zieffler, A., Garfield, J.and delMas, R. (2008) ”A framework to support research on informal inferential reasoning,” *Statistics Education Research Journal*. 7(2), 40-58.