PSYCHOLOGY STUDENTS' UNDERSTANDING OF ELEMENTARY BAYESIAN INFERENCE

Carmen Díaz
University of Granada, Spain
mcdiaz@ugr.es

We explore the possibility of introducing basic ideas of Bayesian inference to undergraduate psychology students and report on the outcomes of our training. We present empirical results on how 78 psychology students learned the basics of Bayesian inference after a 12 hour teaching experience, which included Bayes' theorem, inference on proportions (discrete and continuous case) and means. Learning was assessed through a questionnaire that included multiple choice items and open-ended problems that were solved with the help of computers. In this paper we report part of our results, that show that a majority of students reached good intuitive understanding of most teaching goals, even with a limited time of teaching. We also remark that the main problems detected do not directly relate to Bayesian inference. Difficulties in distinguishing a conditional probability and its inverse that have been repeatedly pointed out in the literature arose in our students and had an influence in general performance.

INTRODUCTION

There is a tendency nowadays to recommend the teaching of Bayesian inference included in undergraduate statistics courses as an adequate and desirable complement to classical inference (Lecoutre, 2006). Situations where prior information can help to make an accurate decision and software that facilitates the application of these methods are becoming increasingly available. Moreover, top core statistical journals now include an important proportion of papers that use Bayesian analysis, but this does not yet translate into comparable changes in the teaching of statistical inference to undergraduates.

Some excellent textbooks whose understanding does not involve advanced mathematical knowledge and where basic elements of Bayesian inference are contextualized in interesting examples (e.g., Berry, 1995) can help introducing these ideas to students. This is however a controversial point. On the one hand, in the past 50 years, errors and difficulties in understanding and applying frequentist inference have been widely described (e.g., in Harlow, Mulaik & Steiger, 1997). These criticisms suggest researchers do not fully understand the logic of frequentist inference and give a (incorrect) Bayesian interpretation to p-values, statistical significance and confidence intervals. On the other hand, it is argued (Moore, 1997) that Bayesian inference relies too strongly on conditional probability, a topic hard for undergraduate students in non-mathematical majors to learn.

It is then possible that learning Bayesian inference is not as intuitive as assumed or at least that not all the concepts involved are equally easy for students. Moreover, empirical research that analyze the learning of students in natural teaching contexts is almost non-existent. Consequently, the aim of this research was to explore the extent to which different concepts involved in basic Bayesian inference are accessible to undergraduate psychology students.

TEACHING EXPERIMENT

The sample taking part in this research included 78 students (18-20 year-olds) in the first year of the Psychology degree at the University of Granada, Spain. These students were taking part in the introductory statistics course and volunteered to take part in the experiment. The sample was composed of 17.9% boys and 82.1% girls, which is the normal proportion of boys and girls in this Faculty. These students scored an average of 4.83 (in a scale 0-10) in the statistics course final examination with standard deviation of 2.07.

The students were organized into four groups of about 15-20 students each and attended a short 12 hour long course given by the same lecturer with the same material. The 12 hours were organized into 4 days. Each day there were two teaching sessions with a half hour break in between. The first session (2 hours) was dedicated to present the materials and examples, followed by a short series of multiple choice items that each student should complete, in order to reinforce

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

their understanding of the theoretical content of the lesson. In the second session (one hour), students worked in pairs in the computer lab with the following Excel programs that were provided by the lecturer to solve a set of inference problems:

1. *Program Bayes:* This program computes posterior probabilities from prior probabilities and likelihood (that should be identified by the students from the problem statement).
2. The *program Prodist* transforms a prior distribution $P(p=p_0)$ for a population proportion p in the posterior distribution $P(p=p_0|data)$, once the number of successes and failures in the sample are given. Prior and posterior distributions are drawn in a graph.
3. The *program Beta* computes probabilities and critical values for the Beta distribution $B(s,f)$, where s and f are the numbers of successes and failures in the sample. It can be used to perform inferences with population proportions for the continuous case.
4. The program *Mean* computes the mean and standard deviation in the posterior distribution for the mean of a normal population, when the mean and standard deviation in the sample and prior population are known.

In Table 1 we present a summary of the teaching content. Students were given printed didactic material that covered this content. Each lesson was organized in the following sections: a) Introduction, describing the lesson goals and introducing a real life situation; b) Progressive development of the theoretical content, in a constructive way and using the situation previously presented; c) Additional examples of other applications of the same procedures and concepts in other real situations; d) Some solved exercises, with description of main steps in the solving procedure; e) New problems that students should solve in the computer lab; and f) Self assessment items. All this material together with the Excel programs described above was also made available to the students on the Internet (http://www.ugr.es/~mcdiaz/bayes, see Fig. 1). We added a forum, so that students could consult the teacher or discuss themselves their difficulties, when needed.

Table 1. Teaching content and its organization

Lesson	Content	In classroom Session 1	Computer lab Session 2
1	Bayes theorem in the context of clinical diagnosis	Prior and posterior probabilities; likelihood; Bayes theorem; comparing subjective and frequentist probability; revision of beliefs; sequential application of Bayes theorem	Solving Bayes problems (Program Bayes)
2	Inference for proportion. Discrete case in the context of voting	Parameters as random variables; prior and posterior distribution; informative and non informative prior distribution; credible intervals; comparing Bayesian and frequentists approaches to inference	Computing credible intervals for proportion; assigning non informative and informative prior distributions (Program Prodist)
3	Inference for proportion. Continuous case in the context of production	Generalizing to continuous case; Beta distribution; its parameters and shape; credible intervals; Bayesian tests	Assigning non informative and informative prior distributions; computing credible intervals for proportion; testing simple hypotheses (Program Beta)
4	Inference for the mean of a normal population in the context of psychological assessment	Normal distribution and its parameters; credible intervals and tests for the mean of a normal distribution with known variance; non informative and informative prior distributions	Assigning non informative and informative prior distributions; computing credible intervals for means; testing simple hypotheses (Program Mean)

Figure 1. Web page with Bayesian inference materials

ASSESMENT

Along the teaching, the students were given several questionnaires to assess their understanding of the topic. Students prepared in advance for the assessment that was part of the workshop they were following. In this paper we will only analyse the learning of inference for the population proportion (continuous case), that is, learning of content introduced in the third lesson. The following group of concepts were assessed: a) Selecting an a-priori distribution; b) computing a-posteriori distribution; c) deciding the best estimator; d) computing credible intervals; e) hypotheses testing and properties of the credible interval These concepts, as well as the philosophical principles of Bayesian inference were assumed to be the core content of basic Bayesian inference in our a-priori analysis and might cause different types of difficulties to students in social sciences. We also assumed that learning of one of these concepts would not automatically assure the learning of the other ideas. The questionnaire included seven multiple choice and one open ended problem and was developed by the author with the specific aim to cover the most important contents in the teaching (see appendix).

In table 2 we show the results in the multiple choice items. In general, few students found difficulties to generalize from the discrete to the continuous case in the Bayesian estimation of proportions. Most of students were able to assign a reasonable prior distribution, in both informative (item 3, 85.9%) and non-informative case (item 1, 93.6%) and could easily handle probability tables and critical values in the Beta distribution to compute credible intervals (item 5, 91%; item 6, 38.5%).

Table 2. Results in multiple choice items (n=78)

Item	Content Assessed	% Correct	Credible interval 95%	
			Lim sup	Lim sup
1	Assigning non informative prior distribution	93.6	0.858	0.971
2	Best prior estimation for proportion	88.5	0.795	0.937
3	Assigning informative prior distribution	85.9	0.764	0.918
4	Best posterior estimation for proportion	52.6	0.416	0.632
5	Computing credible intervals from the posterior distribution	91.0	0.826	0.955
6	Testing hypotheses	55.1	0.441	0.656
7	Algebraic properties of the credible interval	39.7	0.296	0.508

They also understood the meaning of best estimator (88.5% in item 2), and its computation when estimating a proportion from the Beta distribution parameters. This implies the correct interpretation of the parameters of this distribution as the number of successes and failures in a sample and its application in a problem context. The problem was harder in the posterior distribution where 34.6% (item 4) of students did not take into account the previous information, giving an estimator based only in the sample data, this is, using a non-informative prior distribution in item 4.

It was hard for 45.5 % students to test some hypotheses (item 6), because students did not manage the reasoning by contradiction, and misunderstood the global logic of hypothesis testing. Students also had difficulties in relating the width of credible intervals with the credible interval and the sample size. These mistakes are in fact similar to other described in classical inference, although its incidence were not so extended as described in relation to frequentists inference (e.g. in Vallecillos, 1999). It is, however, necessary to consider the limited time of teaching, as well as the lack of other mistakes, described by Vallecillos (1999), for example, in the interpretation of the error risks.

In table 3 we present results of the open-ended problem, which was solved by the students with the help of computer (Beta programme). In the problem we posed questions related to inference for the proportion in the continuous case. Students were firstly asked to select a prior distibution, then to find a posterior distribution, the best estimator before and after the data collection. Finally they were asked to compute some posterior probabilities, credible interval and test an hypothesis.

Table 3. Results in the open-ended problem (*n*=78)

			Confidence interval 95%		Credible interval95%	
	Item and content	% Correct	Lim inf	Lim inf	Lim sup	Lim sup
a.	A-priori distribution	89.7	0.830	0.964	0.810	0.946
b.	A-posteriori distribution	89.7	0.830	0.964	0.810	0.946
c.	Best estimator before collecting data	28.2	0.182	0.382	0.184	0.372
d.	Best estimator after collecting the data	85.9	0.782	0.936	0.765	0.919
e.	Credible interval	92.3	0.864	0.982	0.842	0.963
f.	Posterior probability for a given value	76.9	0.675	0.863	0.664	0.848
g.	Testing hypotheses	85.9	0.782	0.936	0.765	0.919

89.7% of students assigned the prior distribution for non informative case and determined correctly the parameters of the beta posterior distribution, using the data. However, only 28.2% of students gave the correct best prior estimator (the most frequent error in this question was giving the best posterior estimator). Most of students correctly compute the credibility interval (92.3%). 76.9% got a correct posterior probability and 85,9% correctly carried out the hypothesis testing. Testing hypotheses in the open-ended problem was easier for the students because they only were given a possible hypothesis, while in item 6 students had to decide which hypothesis was more reasonable for some given data.

DISCUSSION

A comparative analysis of the undergraduate teaching of statistics shows a clear imbalance between what it is taught and what it is later needed; in particular, most statistics introductory courses present only frequentists inference and many students never get a chance to learn some Bayesian concepts which would improve their professional skills (Bolstad, 2002).

In this paper results from assessing students’ understanding of elementary Bayesian ideas after a short teaching experiment were presented. Our research shows that undergraduate students are able to acquire an intuitive understanding for a number of concepts related in elementary Bayesian inference in a short period of teaching.The high percentage of correct responses in the questionnaire (even in highly complex tasks, such as computing credible intervals and carrying out hypothesis tests) supports the claims for complementing the teaching of frequentists statistics with some ideas of Bayesian statistics in undergraduate statistics courses (e.g. Albert, 2002).

Moreover, the study also provides arguments to reinforce the study of the logic of hypothesis testing, and conditional probability in the teaching of data analysis to psychologists, not only because of the usefulness of these topics in clinical diagnosis, but as a base for future study of

Bayesian inference. We are conscious this research should continue with new samples of students. However we think we have provided arguments to introduce basic Bayesian statistics in undergraduate courses, whenever we emphasize the elements of statistical thinking; incorporate more data and concepts, and fewer recipes and derivations in the classroom, provide students with automate computations and graphics and foster active learning.

ACKNOWLEDGEMENT
This research was supported by the project SEJ2004-00789 MEC, Madrid, and FQM-126, Junta de Andalucía, Spain.

REFERENCES

Albert, J. (2002). Teaching Introductory Statistics from a Bayesian Perspective. In B. Philips (ed). *Proceedings of the Sixth International Conference on Teaching Statistics*. [CDROM]. International Statistical Institute: Cape Town.

Bernardo, J. M. (2006). A Bayesian Mathematical Statistics Primer. In Rossman, A., Chance, B. (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*. [CDROM]. International Association for Statistical Education, Salvador (Bahia), Brazil.

Berry, D. A. (1995). *Basic statistics: a Bayesian perspective.* Belmont, CA: Wadsworth.

Boldstad, W. (2002). Teaching Bayesian Statistics to Undergraduates: Who, What, Where, When, Why, and How. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics*. [CDROM]. International Association for Statistical Education, Cape Town

Bolstad, W. (2004). *Introduction fo Bayesian Statistics*. New York: Wiley.

Iversen, G. R. (1998). Student Perceptions of Bayesian Statistics. In Pereira-Mendoza (Ed.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 234-240). International Statistical Institute, Singapore.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.

Lecoutre B. (2006). Training Students and Researchers in Bayesian Methods for Experimental Data Analysis. *Journal of Data Science, 4*(2), 217-232.

Moore, D. S. (1997). Bayes for beginners? Some pedagogical questions. In S. Panchapakesan (Ed.), *Advances in Statistical Decision Theory* (pp. 3-17). Suttgart: Birkhäuser.

Vallecillos, A. (1999). Some empirical evidence on learning difficulties about testing hypotheses. *Bulletin of the International Statistical Institute: Proceedings of the Fifty-Second Session of the International Statistical Institute Tome 58, Book 2* (pp. 201-204). Helsinki: International Statistical Institute.

APPENDIX. QUESTIONNAIRE

1. When there is no available information about the possible values of a population proportion, the prior distribution for the proportion in the continuous case is given by:
a. The normal distribution N(0,1), with mean 0 and standard deviation 1
b. The discrete uniform distribution in the interval [0,1].
c. The Beta distribution *B* (0,1).
d. The beta Distribution *B* (1,1).

2. Let's assume that the proportion of emigrants in a Spanish city is described by a prior distribution B(3,97). This mean that the best prior estimation of the proportion of emigrants in the city is:
a. 3 %
b. 97%
c. Higher than 3%
d. 3/97 %

3. A medical research found a 10% incidence of depression in women of a given age. A possible prior distribution for defining this population is:
a. *B* (10, 100)
b. *B* (10, 90)
c. *B* (10,10)
d. *B* (90, 10)

4. In a research about social behaviour in preschool children a prior distribution of the proportion of children accepting their mistakes was described by a distribution B (40,60). In a sample of 100 new cases, 55 children accepted their mistakes. The best estimation for the proportion of children that accept their mistakes is:
a. 55%
b. 40%
c. 60%
d. 47.5%

5. The following table present the probabilities and the critical values for *B*(40,60). distribution (Students were provided with the table of this distribution). The 95% credible interval for the proportion in a population described by a posterior distribution B(40,60) is approximately:
a. $(0.35 < p < 0.5)$
b. $(0 < p < 0.516)$
c. $(0.310 < p < 0.497)$
d. $(0.25 < p < 0.8)$

6. Using the data in the previous table, decide which is the most reasonable hypothesis we should accept regarding the proportion of preschool children that accept their mistakes:
a. $H: p<0.3$
b. $H: p > 0.55$
c. $H: p> 0.35$
d. $H: p >0.45$

7. Assuming the same value for proportion in a sample and the same prior distribution, the r% credible intervals for the proportion in a population would be:
a. Wider if the sample size is bigger
b. Wider if the value of r is bigger
c. Narrower if the value of r is bigger
d. It depends on the posterior distribution

Problem. In a study about consume satisfaction, 69 satisfied consumers and 29 unsatisfied consumers were found. There was no previous information available about the consumer satisfaction. Please answer the following questions:

a. Which Beta distribution would you use to reflect your previous knowledge about the population before collecting the data?
b. Which Beta final distribution would reflect your modified knowledge?
c. Which is the best estimator for the proportion in the population before collecting the data?
d. Which is the best estimator for the proportion in the population after collecting the data?
e. Compute the 95% credible interval for the proportion in the population
f. What is your best inference for the probability that more than half the consumers in the population were satisfied?
g. Would you accept the hypothesis that more than 60% of consumers are satisfied?