

# Bayesian Modeling in an Adaptive On-Line Questionnaire for Education and Educational Research

Jaakko Kurhila<sup>1</sup>, Miikka Miettinen<sup>2</sup>, Markku Niemivirta<sup>3</sup>, Petri Nokelainen<sup>1</sup>, Tomi Silander<sup>1</sup>, Henry Tirri<sup>1</sup>  
firstname.lastname@helsinki.fi

University of Helsinki

<sup>1</sup>Department of Computer Science, <sup>2</sup>Department of Psychology, <sup>3</sup>Department of Education  
Complex Systems Computation Group, P.O. Box 26, FIN-00014 Univ. of Helsinki, Finland  
Tel: + 358 9 1911, Fax: + 358 9 191 44441

**Abstract.** Bayesian modeling can be used for providing adaptation in an on-line questionnaire. In our research, adaptation means selecting the questions presented to the user in such a way that the total amount of answers required for profiling the user is minimized. In the article, we present the motivation to use Bayesian modeling as a basis for the adaptation and introduce our adaptive on-line questionnaire EDUFORM which employs these modeling principles. Preliminary empirical study of EDUFORM proved 3 to 17 minutes (35 to 64 percent) time saving and 9 to 32 propositions (36 to 80 percent) less to answer per questionnaire. Since EDUFORM is an open system in the sense that the content is not fixed, we discuss a range of possible uses of EDUFORM, including learner self-evaluation and testing by quizzes to provide assessment information for teachers.

**Keywords:** adaptive educational system, finite mixture models, questionnaire optimisation, self-evaluation, assessment

## 1 Modeling

In many cases, a researcher is in a situation where the domain of the study has been decided, and some research questions have been formulated. In addition, some data related to this domain exist as “legacy data” or otherwise. What to do with the data? Albeit different in theory, in practice the standard answers to this question given in social science literature include such “analysis tools” as exploratory factor analysis (for finding interesting structures in the data), comparative analysis such as discriminant analysis (for comparing different groups with respect to properties of interest), or regression analysis (to predict properties of interest based on known properties). For our purposes here, let us step back from this toolbox- thinking level, and set a simple general framework we will use to explain the Bayesian modeling approach.

For any problem there are infinitely many models, thus we have to somehow restrict the set of models to be examined. In other words, regardless of the analysis method a model is always chosen from a set of models. For example, almost invariably the models used in classical statistical analysis are from a particular set of probability distributions called *normal distributions* (Schervish 1995), i.e., the problem model is described using the language of probability distributions. Whether this is a good set of models or not naturally depends on the problem at hand.

Even if we have restricted the set of models to be considered to a particular set of models properly, we still face the problem of how such models are to be compared with each other. Intuitively, we have a simple answer to this question; one model is better than another if it predicts better. However, things are not as straightforward as this. Although during the process of searching for a good model we do know how well a model predicts the observables at hand, we do not have access to future (unseen) observables, and thus do not know which one of the current models performs best for the future data as well. In classical statistics, this observation reveals itself in many methodological issues such as various estimates of so-called “out-of-sample” performance (Schervish 1995). Consequently, one is forced to use some measure, called a *score*, to compare the different models with each other. The issue of a good score is many-faceted, and in many cases additional criteria such as “Ockham’s razor” are used in scoring the alternative models.

Once a score to compare different models has been selected, one still has to address the problem of how to search among the usually infinite set of models. Naturally, it is possible to do this search manually by using for example classical statistical tests as scores, but in the general case automated search methods should be used to explore different alternatives. Although important in practice for those developing computer programs to implement modeling techniques, we omit further discussion on this topic here.

For the discussion above, we have not made any “Bayesian” assumptions, although the emphasis on predictive modeling is deeply related to Bayesian modeling, and this general framework can be seen as a starting point for many different inference methods. With respect to the topics addressed, *Bayesian inference requires the set of possible models to be stated explicitly in the modeling phase*, which hardly can be considered harmful, as such an assumption has to exist anyway. However, a more characteristic feature for Bayesian inference is the central role of probability, both as a model comparison criterion and a means to predict.

## 2 Bayesian approach to modeling

Due to the impreciseness of natural language, there is a lot of confusion about the term “probability” and “uncertainty”. *Uncertainty* is something we need to model when we create models of the domain. Uncertainty can be modeled in many different ways, probability being one of such methods, and so-called “fuzzy set” (Manton et al 1994) another one. *Probability* is a mathematical construct that behaves in accordance with certain rules (Bernardo and Smith 2000, Berry 1996) and can be used to represent uncertainty. In order to be able to perform inferences using the model, probability needs to be interpreted somehow. Depending on

this interpretation, we end up in different inference frameworks; the classical statistical inference is based on a long-run frequency interpretation of probability, and the Bayesian inference is based on the “degree of belief” interpretation.

The frequency interpretation of the probability of an (observable) event is the long run proportion of the time it happens compared with the total number of observations. Here, long-run means “in the limit” as the total number of observations tends to infinity. Alternatively probability can be defined as *a subjective assessment concerning whether the event in question will occur (or has occurred)*. Now the degree of belief depends on the person who has the belief, as well as on the event in question. In Bayesian inference, this person could be any experimenter or observer. There is not such a thing as *the probability*  $P(A)$  of an event  $A$ , as the probability will always depend on the state of knowledge of the one who believes. Of course, some opinions are based on more information than others.

A subjective degree of belief interpretation applies any time the subject in question has an opinion, and if one counts ignorance as an opinion, this includes every setting. More importantly, subjective information can change when new information arrives. It should be observed that subjectivity in this context does not mean arbitrariness, i.e., that since all probabilities are subjective, everybody has different probabilities. The degree of belief definition of probability says that with different information one may get different probabilities. However, *all subjects sharing the same information will always assign the same probability to the event*. Thus the state of knowledge determines the value of the probability. *Bayesian inference is based on this degree of belief interpretation of probability*. Since all Bayesian probabilities depend on the available information, they are actually mathematical concepts known as *conditional probabilities*, and are denoted  $P(A | I)$ , where  $I$  represents the information affecting the probability assignment.

Let us now suppose that we have some data, and denote these data by  $D$ . In addition, we have several unknown things which we denote by  $M$ . Examples of typical unknown things concerning questionnaires exist. First, there are the values related to the model structure we have chosen. These values are needed to uniquely specify the model, and they are called the *parameters* of the model. Another unknown thing is the missing information in the data, e.g. unanswered questions in the questionnaire. Third, there might be events that were observed neither directly nor exactly. In case of a questionnaire, some of the students may have had low motivation to fill in a questionnaire due to some personal reasons.

Bayesian inference uses conditional probabilities to represent uncertainty. Therefore, we are interested in  $P(M | D, I)$  — the probability of unknown things ( $M$ ) given the data ( $D$ ) and background information ( $I$ ). The initial uncertainty about  $M$  is also represented as a conditional probability  $P(M | I)$ . For example, we could have some initial belief that some answers are more likely than others. Now the essence of Bayesian inference is in the rule that tells us how to update our initial probabilities  $P(M | I)$  if we see data  $D$ , in order to find out  $P(M | D, I)$ . If we return to our example case this means that we could “update” our beliefs in the various alternative answers based on the answers the student has already given. This update rule is known as Bayes’ theorem and can be formally expressed as follows:

$$P(M | D, I) = \frac{P(D | M, I)P(M, I)}{\int P(D | M, I)P(M, I)dM}$$

Consequently Bayesian inference briefly comprises the following principal steps:

- Obtain the initial probabilities  $P(M | I)$  for the unknown things. These probabilities are called the *prior (distribution)*.
- Calculate the probabilities of the data  $D$  given different values for the unknown things, i.e.,  $P(D | M, I)$ . This function of the unknowns is called the *likelihood*.
- Finally the probability distribution of interest,  $P(M | D, I)$ , is calculated using Bayes’ theorem given above. This so called *posterior (distribution)* will then express what is known about  $M$  after observing the data.

Bayes’ theorem can be used sequentially. If we first receive some data  $D$ , and calculate the posterior  $P(M | D, I)$ , and at some later point in time receive more data  $D'$ , the calculated posterior can be used in the role of prior to calculate a new posterior  $P(M | D, D', I)$  and so on. The posterior  $P(M | D, I)$  expresses all the necessary information to perform predictions. The more data we get, the more certain we will become of the unknowns, until all but one value combination for the unknowns have probabilities so close to zero that they can be neglected.

### 3 EDUFORM: an overview

In a questionnaire for education and educational research, there are two separate areas where adaptation makes the questionnaire more useful. The first is to optimise questionnaire length by dropping out uninformative questions. The second is to find user profiles from the sample data so that future users can be classified according to those profiles. In our adaptive on-line questionnaire EDUFORM, we use Bayesian modeling to achieve these benefits.

Even though EDUFORM is an electronic questionnaire on-line, it has a resemblance to the traditional questionnaires on paper (Figure 1). It presents a few multiple-choice questions at a time, with the possibility of adding comments. The navigation bar is at the bottom. The arrows on the right allow the user to move to the next or previous set of questions. Clicking the button with the pie chart icon shows the current profile. If it is already known with sufficient certainty, the user can quit before all questions have been asked by clicking the button with a cross on it. On the left there is a progress indicator showing an estimate of the amount of questions left. Because of the simplicity of the interface, there is no need for a separate help screen. The meanings of the buttons are shown as tooltips (in Figure 1, the word “Profile” next to the pointer).

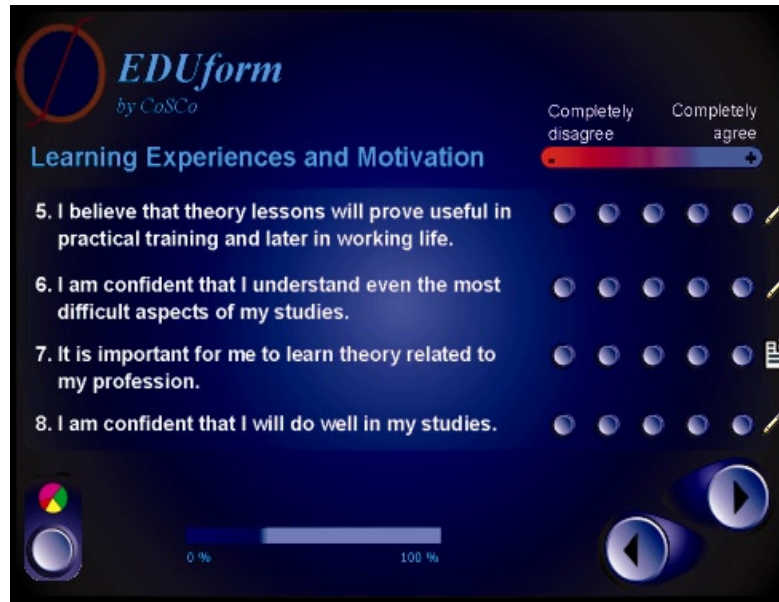


Figure 1: The user interface of EDUFORM.

## 4 Adaptation in EDUFORM

Before EDUFORM can function properly, a profile model to the questionnaire must be built. Although such profiles in principle can be derived in a theory-driven manner and coded manually, in EDUFORM we have adopted the data-driven viewpoint that such profiles are constructed from data gathered with similar questionnaires. Thus in the use of EDUFORM, we distinguish between a *Profile creation phase*, where the probabilistic profiles (clusters) from a sample data are created, and the *Query phase*, where the constructed profiles are used to predict the query responses of the user so that the amount of questions can be optimized. Although the design of EDUFORM is generic and allows the use of any type of predictive modules from neural networks to rule bases, we have adopted the Bayesian modeling approach to describe the profiles.

Possible choices for a model family in the Profile creation phase could be the family of Bayesian networks (Cowell et al. 1999) and family of finite mixtures (Titterton et al. 1985). Also Johnson's and Albert's (1999, 191) work in which they have estimated item response model parameters using Bayesian methods with prior distributions by assuming that the latent traits represent a random sample from a known population could have been a viable choice. The current version of EDUFORM relies on finite mixtures because of the criteria for terminating the questioning process in Query phase can be straightforward if the user is to be profiled into a cluster.

Construction of mixture models from a given data set  $D$  by using the Bayesian approach is described in articles by Kontkanen et al. (1996) and Tirri et al. (1996). In EDUFORM, we have adopted the Bayesian perspective as it allows us to use the prior information available (i.e., the theoretical framework of a questionnaire) and also helps us in the structure selection, i.e., selecting the proper number of profiles.

As a result of the Profile creation phase, a number of clusters have been identified in the sample data. A user answering the questions in EDUFORM eventually falls into one of these clusters. An attempt is made to reduce the required amount of answers significantly, while retaining the usefulness of the data acquired.

The order in which the questions appear in EDUFORM is based on maximising the amount of information gained for profiling for each question. *Kullback-Leibler distance* is used to measure the difference between the current distribution and the distribution that would be the result if the user gave a particular answer. For each of the remaining questions and their possible answers, the distance is calculated and weighted by the probability of the answer. As a result, questions with the maximum expected effect to the cluster distribution can be identified. At any moment, the finite mixture model knows the probability of the individual belonging to each of the clusters, as well as the probabilities of the alternative answers to the remaining questions.

Figure 2 demonstrates the log file of the user's actual answers and predicted answers. Every line represents a question in a questionnaire. The first column states the questionnaire name. The next column tells the number of a particular question in this "demo-1" questionnaire. The next five columns represent the probabilities of a given answer. If the number is 1.0, the user has actually answered to the question and chosen that particular option. The last to rows in Figure 2 show other probability distributions than 0.0 or 1.0. It means the questions corresponding those lines have not been presented to user. Instead, the potential answer of the user is predicted.

demo-1	33	0.0	1.0	0.0	0.0	0.0
demo-1	15	0.0	1.0	0.0	0.0	0.0
demo-1	10	0.0	0.0	0.0	1.0	0.0
demo-1	27	0.0	0.0	0.0	1.0	0.0
demo-1	5	0.0149	0.0292	0.1225	0.2392	0.5939
demo-1	11	0.0084	0.0086	0.0422	0.2451	0.6954

**Figure 2:** Probability distributions of the questions in a demo questionnaire. The text file and the probabilities are truncated to fit the layout.

In the current experimental version of EDUFORM, questions are presented in dynamic chunks of one to four until the probability distribution of the most likely cluster exceeds .80. Once this condition is met, the user is told he or she has provided the necessary information, and asked if the user would like to improve the accuracy of his or her profile by answering the remaining questions. An individual whose answering patterns are very different from the regularities captured by the model may have to answer all of the questions, i.e. the optimization of the amount of questions cannot be done. If the clusters have been named and explanations written for them, the profile can be used for providing immediate feedback to the users.

## 5 Empirical results

A sample of 66 students from Finnish Polytechnic Institute was collected with EDUFORM in February 2001. The educational questionnaire (Ruohotie 2001) consisting total of 116 propositions measured four dimensions: *Part A* “Learning and motivation” (28 propositions), *Part B* “Study habits” (40 propositions), *Part C* “The quality of teaching” (23 propositions) and *Part D* “The effects and outcomes of education” (25 propositions).

Once profiling information during answering process was clear, EDUFORM gave each respondent a chance to move on to next part and skip remaining propositions, or, alternatively, finish answering questions of the current part. Those respondents who skipped were categorized as members of “Group 1” (Adaptive) and those who wanted to give all answers by themselves were members of the “Group 2” (Non-adaptive). Table 1 shows that group 1 has only seven participants (10.6 %) in part A (versus 57, 86.4 %), but even 23 (34.8 %) in part B. Reader should notice that the first two parts of the questionnaire are more laborous to respondent, containing mostly abstract propositions, than the remaining two which measure more practical matters. It is interesting to see that the size of group 2 (All propositions answered) grows in the last two parts of the questionnaire (62.1 % in both). Only 22 students (33.3 %) answered all 116 questions.

**Table 1.** Descriptive statistics of Group 1 (Adaptive) and Group 2 (Non-adaptive) of the adaptive educational questionnaire.

Part A	Part B	Part C	Part D	All propositions
<i>Frequencies (n)<sup>a</sup></i>				
7 / 57 / 64 (2)	23 / 39 / 66 (4)	13 / 41 / 54 (12)	11 / 41 / 52 (14)	23 / 22 <sup>b</sup> / 45 (21)
<i>Percentages (%)<sup>a</sup></i>				
10.6 / 86.4 / 97.0 (3.0)	34.8 / 59.1 / 93.9 (6.1)	19.7 / 62.1 / 81.8 (18.2)	16.7 / 62.1 / 78.8 (21.2)	34.8 / 33.3 / 68.2 (31.8)

a Group 1 (Adaptive) / Group 2 (Non-adaptive) / total number of answers (missing data).

b Omitted case(s) due to total response time below 6 minutes and/or over 60 minutes.

We learn from Table 2 that the total number of propositions needed to complete the questionnaire averaged from 67 (58 %) to 114 (98 %). Time elapsed during answering process varied from 6.1 minutes to 23.8 minutes showing time saving of at least 3.2 minutes, compared to non-adaptive electronic questionnaire. We estimated that the traditional paper version of the same questionnaire should be finished within twenty minutes. The least time saving was obtained in part A (average 3.9 minutes versus 5.7 minutes) and the most in part C (average 1.7 minutes versus 3.1 minutes).

**Table 2.** Comparison of Group 1 and Group 2 by the number of propositions answered and the time elapsed.

Part A	Part B	Part C	Part D	All propositions
<b>Group 1 (Adaptive)</b>				
<i>Propositions answered (n)<sup>a</sup></i>				
7 / 26 / 17.8 / 8.6	17 / 38 / 32.9 / 5.7	8 / 20 / 14.3 / 4.3	12 / 24 / 17 / 4.4	67 / 114 / 100 / 14.3
<i>Time elapsed (s)<sup>a</sup></i>				
76 / 781 / 265.7 / 246.9	100 / 385 / 231.9 / 80.2	30 / 382 / 104.5 / 94.7	24 / 167 / 62.1 / 40.3	364 / 1430 / 760.9 / 342.1
<i>Time elapsed (min)<sup>a</sup></i>				
1.3 / 13.0 / 4.4 / 4.1	1.7 / 6.4 / 3.9 / 1.3	0.5 / 6.4 / 1.7 / 1.6	0.4 / 2.8 / 1.0 / 0.7	6.1 / 23.8 / 12.7 / 5.7
<b>Group 2 (Non-adaptive)</b>				
<i>Time elapsed (s)<sup>a</sup></i>				
87 / 578 / 227.8 / 91.3	213 / 650 / 341.4 / 124.3	77 / 696 / 187.8 / 125.2	63 / 161 / 94.5 / 27.1	556 / 1389 / 825.1 / 216.6
<i>Time elapsed (min)<sup>a</sup></i>				
1.5 / 9.6 / 3.8 / 1.5	3.6 / 10.8 / 5.7 / 2.1	1.3 / 11.6 / 3.1 / 2.1	1.1 / 2.7 / 1.6 / 0.5	9.3 / 23.1 / 13.8 / 3.6

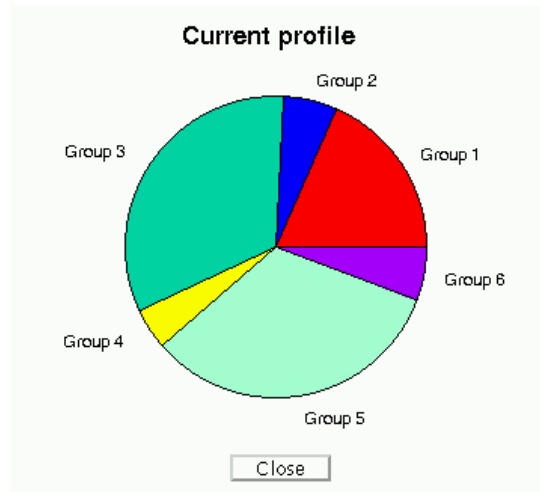
a Min / max / mean / S.D.

## 6 Future uses of EDUFORM

**Tool for self-evaluation.** Flexible tools such as EDUFORM can be used in assessing individual differences on-line. Questionnaires providing answers to questions like “What kind of a learner am I”, “How do I study efficiently” and “What is my motivational profile” can be used as a support material for learner self-evaluation as a part of virtual or campus university studies.

Another important advantage is the immediate feedback EDUFORM provides with an additional feature, namely the visualisation of the user’s profile. The user can see the estimate of his or her profile any time during the fill-in process. The users can be divided into different groups of learners based on their answers to the questionnaire according to the model created in the Profile creation-phase. In Figure 3, the data gathered from a particular user to this stage suggests that the user’s profile is likely to fall either into group three or group five, but more questions should be presented until a reasonably reliable prediction can be made.

It should be noted that the visualization of the user profile is dynamic, i.e. the user can always see his or her profile. This might affect the answering behaviour of an individual, an issue requiring further investigation.



**Figure 3:** Visualization of a user profile.

**Tool for giving feedback.** Preliminary testing implies that the obvious advantage with EDUFORM is that the questionnaires are usually significantly shorter compared to traditional non-adaptive questionnaires. This can help to raise the answering percentage if the questionnaire is seemingly long and tedious, such as course feedback questionnaires in the universities. It is possible that since the process of filling in the questionnaire becomes shorter, the answers can be more accurate because the user is not exhausted with the long list of questions.

In the context of course feedback from a web-based course, the model construction in the Profile creation phase can offer help for teachers to find differences among the various learner groups so that different versions of the web course can be prepared to suit the individual needs of the group.

Although there is a resemblance between EDUFORM and traditional paper-based questionnaires, EDUFORM offers a possibility to write an open comment for every question. This enables valuable feedback to the developers of questionnaires, i.e. to point out questions with poorly chosen wording etc.

**Tool for tests.** Teaching and testing the students knowledge based adaptive questioning is not a novel idea. However, all of the earlier systems starting from BUGGY (Brown and Burton 1978) adapted the questions to the knowledge (or lack of it) of the student, and the questions are tied to the same domain or problem at hand. EDUFORM can be used as a quiz in a different manner. When using EDUFORM as a test, the adaptation means optimizing the length of the test. In other words, the goal is to provide the teacher or the assessor enough information of the students’ progress with as few questions as possible. Of course, the learners can also use EDUFORM tests for self-evaluation. Optimization of the amount of questions opens new research issues. It will be interesting to see how the students evaluate their own actions, i.e. at which point they should stop answering since the prediction of a final score can drop if they start answering against the profile of a “good student”.

**Tool for teachers to evaluate the students.** As well as self-evaluation, EDUFORM can be used by teachers to evaluate the students. Every answer is stored in an easily accessible text-file, as seen in Figure 2. EDUFORM stores a full probability distribution for each variable for every predicted answer, in addition to the order in which the questions were presented. Comments and the amount of time used are also saved with the answers. A log of changes is created to help in identifying ambiguous questions and specific differences between groups of students.

The format for the answers is such that the data can be transferred to other statistical analysis tools. Various reporting tools to be used with EDUFORM can be developed in the future for further analysis of the data gathered in EDUFORM.

## References

- Bernardo, J.M & Smith, A.F. (2000). *Bayesian Theory*. John Wiley & Sons: New York. 2nd ed.
- Berry, D.A. (1996). *Statistics - A Bayesian perspective*. Duxbury Press.
- Brown, J.S. & Burton, R.R. (1978). A paradigmatic example of an artificially intelligent instructional system. *International Journal of Man-Machine Studies*, vol. 10, pages 323-339.
- Cowell, R., Dawid, P.A., Lauritzen, S. & Spiegelhalter, D. (1999). *Probabilistic Networks and Expert Systems*. Springer: New York.
- Johnson, V. & Albert, J. (1999). *Ordinal Data Modeling*. Springer: New York.
- Kontkanen, P., Myllymäki, P. & Tirri, H. (1996). Predictive Data Mining with Finite Mixtures. In *Proceedings of The Second International Conference on Knowledge Discovery and Data Mining*, pages 176-182. Portland, OR, August 1996.
- Manton, K.G., Woodbury & M.A., Tolley, H.D. (1994). *Statistical Applications Using Fuzzy Sets*. John Wiley & Sons: New York.
- Ruohotie, P. (2001). Motivation and Self-regulation in Learning. In P. Nokelainen, P. Ruohotie, T. Silander and H. Tirri (eds.) *Modern Modeling of Professional Growth*, vol. 2, pages 1 – 42. Research Centre for Vocational Education: Hämeenlinna. (In press.)
- Schervish, M. J. (1995). *Theory of Statistics*. Springer-Verlag: New York.
- Tirri, H., Kontkanen, P. & Myllymäki, P. (1996). Probabilistic Instance-Based Learning. In L. Saitta, *Machine Learning: Proceedings of the Thirteenth International Conference*, pages 507-515. Morgan Kaufmann Publishers: San Francisco.
- Titterton, D.M. and Smith, A.F.M. & Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons: New York.