
Twente Univ., Enschede (Netherlands). Faculty of Educational Science and Technology.

Nov 94

36p.

Bibliotheek, Faculty of Educational Science and Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.

Reports - Evaluative/Feasibility (142)

Bayesian Statistics; Decision Making; Foreign Countries; Intelligent Tutoring Systems; Models; Psychometrics; Scores; Test Results; Test Theory

Decision Theory; Minnesota Adaptive Instructional System

Some applications of Bayesian decision theory to intelligent tutoring systems are considered. How the problem of adapting the appropriate amount of instruction to the changing nature of a student's capabilities during the learning process can be situated in the general framework of Bayesian decision theory is discussed in the context of the Minnesota Adaptive Instructional System (MAIS). Two basic elements of this approach are used to improve instructional decision making in intelligent tutoring systems. First, it is argued that in many decision-making situations the linear loss model is a realistic representation of the losses actually incurred. Second, it is shown that the psychometric model relating observed test scores to the true level of functioning can be represented by Kelley's regression line from classical test theory. Optimal decision rules for the MAIS are derived using these two features. (Contains 3 tables, 1 figure, and 42 references.) (SLD)
Applications of Bayesian Decision Theory to Intelligent Tutoring Systems

Hans J. Vos

Faculty of Educational Science and Technology

University of Twente
Applications of Bayesian Decision Theory to Intelligent Tutoring Systems

Hans J. Vos

Applications of bayesian decision theory to intelligent tutoring systems, Hans J. Vos - Enschede: University of Twente, Faculty of Educational Science and Technology, November, 1994. - 29 pages.
Abstract

The purpose of this chapter is to consider some applications of Bayesian decision theory to intelligent tutoring systems. In particular, it will be indicated how the problem of adapting the appropriate amount of instruction to the changing nature of student's capabilities during the learning process can be situated within the general framework of Bayesian decision theory. Two basic elements of this approach will be used to improve instructional decision making in intelligent tutoring systems. First, it is argued that in many decision-making situations the linear loss model is a realistic representation of the losses actually incurred. Second, it is shown that the psychometric model relating observed test scores to the true level of functioning can be represented by Kelley's regression line from classical test theory. Optimal decision rules will be derived using these two features.
Applications of Bayesian Decision Theory

Introduction

During the last two decades, adaptive instructional systems have been studied by many researchers (e.g., Atkinson, 1976; De Diana & Vos, 1988; Gegg-Harrison, 1992; Hambleton, 1974; Hansen, Ross & Rakow, 1977; Holland, 1977; van der Linden & Vos, 1994; Vos, 1990, 1991, 1992, 1993, 1994a, 1994c, 1995; Vos & De Diana, 1987). Although different authors have defined the term "adaptive instruction" in a different way, most agree that it denotes the use of strategies to adapt instructional treatments to the changing nature of student abilities and characteristics during the learning process (see, e.g., Landa, 1976).

In the context of computer-based instruction (CBI), adaptive instructional programs are often qualified as intelligent tutoring systems (ITSs). Examples of such systems can be found in Capell and Dannenberg (1993) and De Haan and Oppenhuizen (1994). Tennyson, Christensen, and Park (1984) have described a computer-based adaptive instructional system denoted as the Minnesota Adaptive Instructional System (MAIS). The authors consider MAIS as an ITS, because it exhibits some machine intelligence, as demonstrated by its ability to improve decision making over the history of the system as a function of accumulated information about previous students. In the literature, successful research projects on MAIS have been reported (e.g., Park & Tennyson, 1980; Tennyson, Tennyson & Rothen, 1980).

Initial work on MAIS began as an attempt to design an adaptive instructional strategy for concept-learning (Tennyson, 1975). Concept-learning is the process in which subjects learn to categorize objects, processes or events. A
Applications of Bayesian Decision Theory

3

model for the instruction for the learning of concepts is described by Merrill and Tennyson (1977). These authors suppose that the learning of concepts consists of two phases. The first one is the formation of a prototype (i.e., formation of conceptual knowledge) and the second is the acquisition of classificatory skills (i.e., development of procedural knowledge). From this assumption, an instructional design model for the learning of concepts has been developed. This model has two basic components: content structure variables and instructional design variables. Furthermore, an important role in the model is played by expository examples (statement form), i.e. (non)examples, which organize the content in propositional format and interrogatory examples (question form), i.e. (non) examples, which organize the content in interrogatory format (see Tennyson and Cocchiarella, 1986, for a complete review of the theory of concept-learning).

In MAIS, eight basic instructional design variables directly related to specific learning processes are distinguished. In order to adapt instruction to individual learner differences (aptitudes, prior knowledge) and learning needs (amount and sequence of instruction), these variables are controlled by an ITS. Three out of these eight variables are directly managed by a computer-based decision strategy, namely, amount of instruction, instructional time control, and advisement on learning need. The functional operation of this strategy was related to guidelines described by Novick and Lewis (1974).

Four empirically based adaptive instructional models have been reviewed by Tennyson and Park (1984). The four models are Atkinson's mathematical model, Ross's trajectory model, Ferguson's testing and branching model, and the MAIS model. These four models vary in degree to which they use six characteristics (initial diagnosis, sequential character, amount of instruction,
Applications of Bayesian Decision Theory

sequence of instruction, instructional display time, and advisement on learning need) identified as essential in an effective adaptive instructional system. The authors conclude that MAIS provides for a complete adaptive instructional model, because all six defined characteristics of effective adaptive instruction are integrated into this model.

The purpose of this paper is to review the application of the MAIS decision procedure by Tennyson and his associates. First, it will be indicated how this procedure can be situated within the general framework of Bayesian decision theory (e.g., Ferguson, 1976; Lindgren, 1976), and what implicit assumptions have to be made in doing so. Next, it will be demonstrated how the decision component in MAIS can be improved by using other results from this decision-theoretic approach. In particular, it will be indicated how two features of the MAIS decision procedure can be improved by using other results from decision theory. The first feature is to replace the assumed threshold loss function in MAIS by a linear loss function. The second feature is Kelley’s regression line of classical test theory as the psychometric model relating observed test scores to the true level of functioning instead of the binomial model assumed in MAIS.

We shall confine ourselves in this paper only to one of the three instructional design variables directly managed by the decision component in MAIS, namely selecting the appropriate amount of instruction in concept or rule-learning situations. In MAIS, selecting the appropriate amount of instruction can be interpreted as determining the optimal number of interrogatory examples. Although the procedures advocated in this paper are demonstrated for instructional decision making in MAIS, it should be emphasized that these procedures are not limited to MAIS but, in principle, can be applied to decision components in any arbitrary ITS. In the next section, it will be indicated how the
problem of selecting the appropriate amount of instruction in MAIS can be situated within the general framework of Bayesian decision theory.

Adapting the Amount of Instruction

The derivation of an optimal strategy with respect to the number of interrogatory examples requires an instructional problem be stated in a form amenable to a Bayesian decision-theoretic analysis. In a Bayesian view of decision making, there are two basic elements to any decision problem: a loss function describing the loss \( l(a_i, t) \) incurred when action \( a_i \) is taken for the student whose true level of functioning is \( t \) (\( 0 \leq t \leq 1 \)), and a probability function or psychometric model, \( f(x|t) \), relating observed test scores \( x \) to student's true level of functioning \( t \).

These basic elements have been related to decision problems in educational testing by many authors (e.g., Atkinson, 1976; Huynh, 1980; Swaminathan, Hambleton, & Algina, 1975; van der Linden, 1990). As the use of the decision component in MAIS refers to mastery testing, we shall discuss here only the application of the basic elements to this problem.

It is assumed that, due to measurement and sampling errors, the true level of functioning \( t \) is unknown. All that is known is the student's observed test score \( x \) from a small sample of \( n \) interrogatory examples (\( x = 0, 1, ..., n \)). Furthermore, the following two actions are available to the decision-maker: advance a student \( (a_1) \) to the next concept if his/her test score \( x \) exceeds a certain cutting score \( x_c \) on the observed test score scale \( X \), and retain \( (a_0) \) him/her otherwise. Students with test scores \( x \) below the cutting score \( x_c \) are provided with additional expository examples. A new interrogatory example is then generated. This procedure is applied sequentially until either mastery is attained.
or the pool of test items is exhausted.

The mastery decision problem can now be stated as choosing a value of $x_c$ that, given the value of the criterion level $t_c$, is optimal in some sense. The criterion level $t_c \in [0,1]$ - the minimum degree of student's true level of functioning required - is set in advance by the decision maker. It is the unreliability of the test that opens the possibility of the mastery decision problem (Hambleton & Novick, 1973).

Generally speaking, a loss function specifies the total costs of all possible decision outcomes. These costs concern all relevant psychological, social, and economic consequences that the decision brings along. An example of economic consequences is extra computer time associated with presenting additional instructional materials. In MAIS, the loss function is supposed to be a threshold function. The implicit choice of this function implies that the "seriousness" of all possible consequences of the two available actions can be summarized by four constants, one for each of the four possible decision outcomes (see Table 1).

For convenience, and without loss of generality (e.g., Davis, Hickman & Novick, 1973), it is assumed in Table 1 that no losses occur for correct decisions. Therefore, the losses for correct advance and retain decisions, i.e., $1_{11}$ and $1_{00}$, can be set equal to zero.

In the decision component of MAIS, a loss ratio $R$ must be specified.
Applications of Bayesian Decision Theory

refers to the relative losses for advancing a learner whose true level of functioning is below $t_c$ and retaining one whose true level exceeds $t_c$, or, equivalently, the losses associated with a false advance compared to a false retain decision. From Table 1 it can be seen that the loss ratio $R$ equals $l_{10}/l_{01}$ for all values of $t$.

Finally, it is assumed that the psychometric model in MAIS, relating observed test scores $x$ to the true level of functioning $t$, can be represented by the well-known binomial model:

$$f(x|t) = \binom{n}{x} t^x (1-t)^{n-x}. \quad (1)$$

In a Bayesian procedure, a decision problem is solved by minimizing the Bayes risk, which is minimal if for each $x$ of $X$ an action with smallest posterior expected loss is chosen. The posterior expected loss is the expected loss taken with respect to the posterior distribution of $t$.

It can be seen from the loss table that a decision rule minimizing posterior expected loss is to advance a student whose test score $x$ is such that

$$l_{01} \text{Prob}(T \geq t_c|x,n) > l_{10} \text{Prob}(t < t_c|x,n). \quad (2)$$

and to retain him/her otherwise. Since $l_{01} > 0$, this is equivalent to advancing a student if

$$\text{Prob}(t \geq t_c|x,n) \geq R/(1+R). \quad (3)$$

and retaining him/her otherwise. $\text{Prob}(t \geq t_c|x,n)$ denotes the probability of the
student's true level of functioning $\alpha$; equal to $\alpha$; larger than $\alpha$; given a test score $x$ on a test of length $n$. In fact, this probability is one minus the cumulative posterior distribution of $t$. In MAIS, this quantity is called the "beta value" or "operating level" (Tennyson, Christensen, & Park, 1984).

It should be noted that, as can be seen from the optimal decision rule, the decision maker does not need to specify the values $l_{10}$ and $l_{01}$ completely. He needs only assess their ratio $l_{10}/l_{01}$. For assessing loss functions, most texts on decision theory propose lottery methods (see, for example, Novick & Lindley, 1979; Vos, 1994b). But, in principle, any psychological scaling method can be used.

In order to initiate the decision component in MAIS, three kinds of parameters must be specified in advance. Beside the parameters $t_c$ and $R$, a probability distribution representing the prior knowledge about $t$ must be available. In MAIS, a beta distribution, $B(\alpha, \beta)$, is used as a prior distribution, and a pretest score together with information about other students is used to specify its parameter values.

Keats and Lord (1962) have shown that simple moment estimators of $\alpha$ and $\beta$, respectively, are given as

$$\alpha = (-1 + 1/\rho_{pre}) \mu_{pre}$$
$$\beta = -\alpha + n/\rho_{pre} - n.$$  

where $\mu_{pre}$ and $\rho_{pre}$ denote the mean and KR-21 reliability coefficient of the test scores from the previous students, respectively, and $n$ represents the number of test items in the pretest. As an aside, it may be noted that if administering a pretest is not possible for any reason, the prior distribution of a student can be
characterized by a uniform distribution on the interval from zero to one. In that case, the parameters of the beta prior should be specified as $\alpha = \beta = 1$. Also, the prior distribution can be estimated on the initial period of instruction, for instance, on the first four or six interrogatory examples (Tennyson, Christensen, & Park, 1984).

From an application of Bayes' theorem, it follows that the posterior distribution of $t$ will again be a member of the beta family (the conjugacy property). In fact, if the prior distribution is $B(\alpha, \beta)$ and the student's test score is $x$ from a test of length $n$, then the posterior distribution is $B(\alpha + x, \beta + n - x)$. The beta distribution has been extensively tabulated (e.g., Pearson, 1930). Tennyson and Christensen (1986) use a nonlinear regression approach that fits the best polynomial as an approximation of the beta distribution. Normal approximations are also available (Johnson & Kotz, 1970, sect. 2.4.6). Using numerical procedures for computing the incomplete beta function, a computer program called BETA was developed in PASCAL to calculate the beta values for the purpose of this paper. The program is available on request from the author.

The MAIS decision procedure for adapting the number of interrogatory examples can now be summarized as follows: If a student's beta value exceeds the quantity $R/(1+R)$, (s)he is passed to the next concept. However, if his/her beta value is below this quantity, his/her posterior distribution is used as a prior distribution in a next cycle. A new interrogatory example is then generated. The procedure is applied iteratively until either the beta value exceeds the quantity $R/(1+R)$ or all interrogatory examples have been presented. Notice that the iterative updating of the beta values takes into account improvements in learning while a straight percentage per number of items weights all responses equally. Consequently, as the student makes increasingly correct answers in the latter part
Applications of Bayesian Decision Theory

of instruction, those answers become weighted more than in the initial period of instruction (Tennyson, Christensen, & Park, 1984).

In the MAIS decision procedure, it is assumed that the form of the loss structure involved is a threshold function. Therefore, only the loss ratio R has to be assessed empirically. In addition to the threshold loss function, however, more realistic functions have been adopted in decision theory. One such function will be considered below.

The Linear Loss Model

An obvious disadvantage of the threshold loss function is that it assumes constant loss for students to the left or to the right of $t_c$, no matter how large their distance from $t_c$. For instance, a misclassified "true master" (see Table 1) with a true level of functioning just above $t_c$ gives the same loss as a misclassified "true master" with a true level far above $t_c$. It seems more realistic to suppose that for misclassified "true masters" the loss is a monotonically decreasing function of t.

Moreover, as can be seen in Table 1, the threshold loss function shows a "threshold" at the point $t_c$, and this also seems unrealistic in many cases. In the neighborhood of this point, the losses for correct and incorrect decisions frequently change smoothly rather than abruptly.

In view of this, Mellenbergh and van der Linden (1981) proposed the following linear loss function:

$$l(a_i,t) = \begin{cases} b_0(t-t_c) + d_0 & \text{for } i = 0 \text{ (retain)} \\ b_1(t_c-t) + d_1 & \text{for } i = 1 \text{ (advance)} \end{cases} \tag{5}$$
where $b_0, b_1 > 0$. The above defined function consists of a constant term and a term proportional to the difference between the true level of functioning $t$ and the specified criterion level $t_c$. The constant amount of loss, $d_i$ ($i = 0,1$), can, for example, represent the costs of testing. The condition $b_0, b_1 > 0$ is equivalent to the statement that for actions $a_0$ and $a_1$, loss is a strictly increasing and decreasing function of the variable $t$, respectively. The parameters $b_i$ and $d_i$ have to be assessed empirically (e.g., Novick & Lindley, 1979; Vos, 1994b). Figure 1 displays an example of this function.

The linear loss function seems to be a realistic representation of the losses actually incurred in many decision making situations. In a recent study, for example, it was shown by van der Gaag, Mellenbergh, and van den Brink (1988) that many empirical loss structures could be approximated satisfactorily by linear functions.

Since this paper is only meant to give a flavor of the possible applications of Bayesian decision theory to ITSs, only the case $d_0 = d_1$ will be considered in the linear loss function of (5). In other words, it will be assumed that the amounts of constant loss, $d_i$, for both actions are equal, or there are no constant losses at all (i.e., no costs of testing are involved). Confining ourselves to this special case, the mathematical derivations given below will remain rather simple. For the more general and a bit more complicated case of $d_0 \neq d_1$, we refer to Vos (1994b). It should be noted, however, that no fundamentally new
ideas are encountered in this more general case.

For the case of $d_0 = d_1$, it can easily be verified from (5) that the decision rule that minimizes the posterior expected loss in case of a linear loss function is to advance a student with test score $x$ for which

$$E[tlx,n] \geq t_c,$$  

(6)

and to retain him/her otherwise. As can be seen from (6), under the assumption of $d_0 = d_1$, there is no need to assess the parameters $d_1$ and $b_1$ in adapting the number of interrogatory examples. In this case, the optimal decision rule takes the rather simple form of advancing a student if his/her expectation of the posterior distribution of $t$ is equal to or larger than the specified criterion level $t_c$, and to retain him/her otherwise. Following the same terminology as in the threshold loss model, the expectation of the posterior distribution of $t$ will be denoted as the "linear value". So, a student is advanced in the threshold loss model if his/her beta value exceeds the quantity $R/(1+R)$ and is advanced in the linear loss model if his/her linear value exceeds the criterion level $t_c$.

Using the fact that the expectation of a beta distribution $B(\alpha,\beta)$ is equal to $\alpha/(\alpha+\beta)$, and thus, the posterior expectation equals $(\alpha+x)/(\alpha+\beta+n)$, it follows from (6) that a student is advanced if his/her test score $x$ is such that

$$x \geq t_c(\alpha+\beta+n) - \alpha,$$  

(7)

and retained otherwise.

In MAIS, it is assumed that the form of the psychometric model relating observed test scores to student's true level of functioning can be represented by
the binomial model (Equation 1). In the next section, another psychometric model frequently used in criterion-referenced testing will be considered.

**Classical Test Model**

The expectation of the posterior distribution, \(E[t|x,n]\), represents the regression of \(t\) on \(x\). A possible regression function is the linear regression function of classical test theory (Lord & Novick, 1968):

\[
E[t|x,n] = \mu_X + (1-\rho_{XX'}) \mu_X / n,
\]

with \(\mu_X\) and \(\rho_{XX'}\) being the mean and KR-21 reliability coefficient of \(X\) (i.e., the group to which the student belongs), respectively. Equation 8 is known as Kelley's regression line. According to Lord and Novick (1968), Equation 8 is "an interesting equation in that it expresses the estimate of the true level of functioning as a weighted sum of two separate estimates - one based upon the student's observed score, \(x\), and, the other based upon the mean, \(\mu_X\) of the group to which s(he) belongs. If the test is highly reliable, much weight is given to the test score and little to the group mean, and vice versa." (p.65)

Substituting (8) into (6), and solving for \(x\) gives the following optimal decision rule

\[
x \geq \frac{\mu_X(\rho_{XX'}-1)+n_0}{\rho_{XX'}}.
\]

(9)

Since \(0 \leq \rho_{XX'} \leq 1\), and, thus \(-1 \leq \rho_{XX'}-1 \leq 0\), it follows from (9)
that $\mu_X$ and the optimal cutting score are related negatively. The higher the average performance, the lower the optimal cutting score. Hard-working students are rewarded by low cutting scores, while less hard-working students will just be penalized and confronted with high cutting scores. This effect is the opposite of what happens when norm-referenced standards are used (van der Linden, 1980). They vary up and down with the performances of the examinees. Van der Linden (1980) calls this effect a "regression from the mean".

It should be stressed that, as can be seen from (9), the optimal cutting score, i.e., the number of interrogatory examples to be administered to the student, depends upon $\mu_X$ and $\rho_{XX'}$. Hence, it follows that the decision component in MAIS allows for an updating after each response to an interrogatory example. This explains why, though the decisions for determining the optimal number of interrogatory examples are made with respect to an individual student, the rules for the decisions are based on data from all students taught by the system in the past and, in doing so, are improved continuously. In other words, instructional decision-making procedures for ITSs can be designed in this way; that is, a system of rules improving itself over the history of the system as a result of systematically using accumulated data from previous students. The parameters of the model, $\mu_X$ and $\rho_{XX'}$, are updated each time a student has finished his/her dialogue with the system.

Comparison of the Models

In this section, the threshold loss, linear loss, and classical test model will be compared with each other. First, both the threshold and linear loss model will be
Applications of Bayesian Decision Theory

compared with the classical test model. Next, the threshold and linear loss model will be compared with each other.

As noted earlier, both the threshold and linear loss model do not take test scores into account of the group to which the student belongs. Both models were primarily designed for instructional decision making on the level of the individual student. The classical test model, however, explicitly takes into account both the student’s observed test score and the mean of the group to which the student belongs, which is illustrated by the "regression from the mean" effect.

The "individual" models (i.e., the threshold and linear loss model), however, explicitly take into account information about other students (so-called "collateral" information) to specify the parameter values of a distribution function representing the prior knowledge about the true level of functioning. In the case of a beta distribution, as shown by Keats and Lord (1962), the estimates \( \alpha \) and \( \beta \) of the prior distribution are given by (4). Inserting (4) into (7) results into

\[
x \geq \left[ \mu_{\text{pre}}(\rho_{\text{pre}}^{-1}) + n_{\text{pre}} \right] / \rho_{\text{pre}}.
\]

Comparing (9) and (10) with each other, it follows immediately that the linear loss model and classical test model yield the same optimal cutting score if \( \rho_{\text{pre}} = \rho_{XX} \) and \( \rho_{\text{pre}} = \rho_{XX} \); that is, if the means and KR-21 reliability coefficients of the pretest scores and scores of the group to which the student belongs are the same. Under the (realistic) assumption \( \rho_{\text{pre}} = \rho_{XX} = \rho \), and using \( -1 \leq \rho \leq 0 \), it follows from (9) and (10) that the optimal cutting score in the classical test model can be set lower than in the linear loss model if \( \mu_X > \mu_{\text{pre}} \) and vice versa. This makes sense, because this implies that the student is rewarded for performing better than the average student from the
"collateral" group. Using a normal approximation for the beta distribution and applying a logistic transformation with scale parameter equal to 1.7 (e.g., Lord & Novick, 1968, sect. 17.2), the same conclusion can easily be derived for the threshold loss and classical test model (Vos, 1994b).

After having compared the threshold loss and linear loss model with the classical test model, these two "individual" models will now be compared with each other. Setting $t_c$ equal to 0.7, the beta values (left-hand side of Expression 3) and linear values (left-hand side of Expression 6) were computed using the program BETA. Since pretest information was available, $\alpha$ and $\beta$ were estimated from (4) with $n = 10$, $\mu_{pre} = 8$, and $\rho_{pre} = 0.8$. The results of the computations for the threshold and linear loss model are given in Tables 2 and 3, respectively, for 10 test items and different number correct scores.

---

Insert Table 2 about here

---

Insert Table 3 about here

---

As can be seen from (6), a student is advanced in the linear loss model if the number correct score of his/her linear value exceeds $t_c = 0.7$. In Table 3, these values are indicated by an asterisk. Similarly, as can be seen from (3), a student is advanced in the threshold loss model if his/her beta value exceeds the quantity $R/(1+R)$. Let us suppose that the relative losses associated with a false
Applications of Bayesian Decision Theory

17

advance compared to a false retain decision are considered equally worse (i.e., \( l_{10} = l_{21} \)). This implies that \( R = l_{10}/l_{01} = 1 \), and, thus, \( R/(1+R) \) equals 0.5. In Table 2, the values for which the number correct score exceeds the quantity \( R/(1+R) = 0.5 \) are also indicated by an asterisk. Using the program BETA, in Table 2 it is also indicated for which value of the loss ratio \( R \), say \( R_c(0.7) \), both models yield the same optimal cutting score \( x_c \) if \( t_c \) is set equal to 0.7. The optimal cutting score \( x_c \) for the linear loss model was derived from (7) for \( t_c = 0.7 \) and is depicted in Table 3.

Tables 2 and 3 indicate that with this choice of the loss ratio \( R \), the number correct score for which a student is granted mastery status does not differ much in both models. Only if the number of items is equal to 9 a student needs one more item correct in the linear loss model than in the threshold loss model for being advanced. So, the linear loss model is somewhat more severe than the threshold loss model in the case of \( R = 1 \).

This can also be concluded from examining the values of \( R_c(0.7) \), because all these values are larger than 1. Hence, if it is required that a student is advanced in both models with the same number correct score, then, the losses associated with a false advance decision should be considered more worse than the losses associated with a false retain decision. Since Table 2 shows that \( R_c(0.7) \) can be lowered with increasing number of items, however, both false decisions become more and more equally worse with increasing values of \( n \).

Of course, the values of \( R \) for which students are advanced with the same number of correct score in both models depend upon the value of \( t_c \). Therefore, in Table 2, these values of \( R \) are also displayed for \( t_c = 0.6 \) and \( t_c = 0.8 \) denoted as \( R_c(0.6) \) and \( R_c(0.8) \), respectively. As can be concluded from Table 2, the linear loss model becomes more and more severe than the threshold
Applications of Bayesian Decision Theory

loss model for increasing values of $t_c$, whereas for decreasing values of $t_c$ the opposite happens.

Finally, it should be noted that for any choice of the loss ratio $R$ and criterion level $t_c$, always a linear loss model can be found yielding the same optimal cutting score by choosing appropriate values for the linear loss parameters $b_i$ and $d_i$. Hence, the threshold loss model can be considered as a special case of the linear loss model. In other words, the linear loss model offers us a great deal of flexibility in designing the adaptive decision making procedure in MAIS. In the program BETA, the optimal cutting scores $x^*_c$ in the linear loss model and its associated $R_c$ values can also be computed for the general case of $d_0 \neq d_1$. For this general case of the linear loss model, it is shown in Vos (1994b) that a student is advanced to the next concept if his/her linear value exceeds the $t$-coordinate of the intersection point of both loss lines from (5), which is equal to $[t_c+(d_1-d_0)/(b_0+b_1)]$. All results reported in this paper, however, can be obtained by setting, in addition to $b_0$, $b_1 > 0$, $d_0$ and $d_1$ equal to each other in the computer program BETA.

Concluding Remarks

In this paper it was indicated how the MAIS decision procedure could be formalized within a Bayesian decision-theoretic framework. In fact, it turned out that this decision procedure could be considered as a sequential mastery decision.

Moreover, it was argued that in many situations the assumed threshold loss function in MAIS is an unrealistic representation of the losses actually incurred. Instead, a linear loss function was proposed to meet the objections to
threshold loss.

Further, Kelley's regression line of classical test theory was proposed as the psychometric model relating observed test scores to the true level of functioning. Using this psychometric model instead of the binomial model assumed in MAIS, ISSs can be designed in which the determination of the optimal number of interrogatory examples for an individual student is based on data from all students taught by the system in the past.

Integrating these two features into MAIS, it might be expected that the computer-based decision strategy in MAIS can be improved. Using computer simulation and deriving theoretical implications, a critical comparison of the models was carried out in order to validate these two extensions of MAIS. The results of the computer simulations and theoretical implications indicated that both extensions were realistic. That is, both extensions of MAIS are potentially valuable and feasible for current ITS applications. Whether or not the proposed linear loss model and the classical test model are, however, real improvements of the present decision component in MAIS (in terms of student performance on posttests, learning time, and amount of instruction) must be decided on the basis of empirical data.
Applications of Bayesian Decision Theory

References


Applications of Bayesian Decision Theory


Applications of Bayesian Decision Theory


Table 1

Twofold Table for Threshold Loss Function

<table>
<thead>
<tr>
<th>Decision</th>
<th>True level</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t \geq t_c$</td>
<td>$t &lt; t_c$</td>
<td></td>
</tr>
<tr>
<td>Advance</td>
<td>0</td>
<td>$l_{10}$</td>
<td></td>
</tr>
<tr>
<td>Retain</td>
<td>$l_{01}$</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Beta values Calculated by Item Number and Number Correct at a 0.7 Criterion Level and Rc values

<table>
<thead>
<tr>
<th>Item Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta value</td>
<td>0.84*</td>
<td>0.49*</td>
<td>0.24*</td>
<td>0.11</td>
<td>0.05</td>
<td>0.02</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Rc(0.7)</td>
<td>1.290</td>
<td>1.251</td>
<td>1.234</td>
<td>1.204</td>
<td>1.188</td>
<td>1.165</td>
<td>1.136</td>
<td>1.116</td>
<td>1.107</td>
<td>1.107</td>
</tr>
<tr>
<td>Rc(0.8)</td>
<td>1.471</td>
<td>1.471</td>
<td>1.471</td>
<td>1.471</td>
<td>1.471</td>
<td>1.471</td>
<td>1.471</td>
<td>1.471</td>
<td>1.471</td>
<td>1.471</td>
</tr>
</tbody>
</table>

Applications of Bayesian Decision Theory

26
<table>
<thead>
<tr>
<th>Item Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_c=0.45$</td>
<td>0.86*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.67</td>
<td>0.80*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_c=1.15$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.55</td>
<td>0.71*</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_c=1.45$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.46</td>
<td>0.62</td>
<td>0.77*</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_c=5.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.35</td>
<td>0.54</td>
<td>0.67</td>
<td>0.80*</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_c=5.25$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0.40</td>
<td>0.54</td>
<td>0.67</td>
<td>0.80*</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_c=4.04$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>0.42</td>
<td>0.53</td>
<td>0.63</td>
<td>0.74*</td>
<td>0.84</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_c=4.65$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.29</td>
<td>0.38</td>
<td>0.48</td>
<td>0.67</td>
<td>0.76*</td>
<td>0.86</td>
<td>0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_c=5.15$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>0.26</td>
<td>0.35</td>
<td>0.43</td>
<td>0.52</td>
<td>0.61</td>
<td>0.69</td>
<td>0.78*</td>
<td>0.87</td>
<td>0.96</td>
</tr>
<tr>
<td>Linear value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_c=6.05$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>0.24</td>
<td>0.32</td>
<td>0.40</td>
<td>0.48</td>
<td>0.56</td>
<td>0.64</td>
<td>0.72*</td>
<td>0.80</td>
<td>0.88</td>
</tr>
<tr>
<td>Linear value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_c=6.75$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.96</td>
</tr>
</tbody>
</table>

Table 3. Linear values and $x_c$ values calculated by item number and number correct at a 0.7 criterion level.
Applications of Bayesian Decision Theory

Figure Caption

Figure 1. Example of a Linear Loss Function.

\( b_0 \neq b_1, d_0 \neq d_1 \)
Titles of recent Research Reports from the Department of
Educational Measurement and Data Analysis.
University of Twente, Enschede,
The Netherlands.

RR-94-16 H.J. Vos, Applications of Bayesian decision theory to intelligent tutoring systems
RR-94-15 H.J. Vos, An intelligent tutoring system for classifying students into Instructional
treatments with mastery scores
RR-94-14 R.J.H. Engelen, W.J. van der Linden & S.J. Oosterloo, Fisher's information in the Rasch
model
RR-94-13 W.J.J. Veerkamp & M.P.F. Berger, A note on the attenuation paradox in item response
theory
RR-94-12 R.R. Meijer, Nonparametric and group-based person-fit statistics: A validity study and an
empirical example
RR-94-11 M.P.F. Berger, Optimal test designs for polytomously scored items
RR-94-10 W.J. van der Linden & M.A. Zwarts. Robustness of judgments in evaluation research
RR-94-9 L.M.W. Akkermans. Monte Carlo estimation of the conditional Rasch model
developments
RR-94-7 W.J. van der Linden & R.M. Luecht, An optimization model for test assembly to match
observed-score distributions
RR-94-6 W.J.J. Veerkamp & M.P.F. Berger. Some new item selection criteria for adaptive testing
RR-94-5 R.R. Meijer, K. Sijtsma & L.W. Molenaar, Reliability estimation for single dichotomous
items
RR-94-3 W.J. van der Linden. A conceptual analysis of standard setting in large-scale assessments
RR-94-2 W.J. van der Linden & H.J. Vos, A compensatory approach to optimal selection with
mastery scores
RR-94-1 R.R. Meijer. The influence of the presence of deviant item score patterns on the power of a
person-fit statistic
RR-93-1 P. Westers & H. Kelderman. Generalizations of the Solution-Error Response-Error Model
RR-91-1 H. Kelderman. Computing Maximum Likelihood Estimates of Loglinear Models from
Marginal Sums with Special Attention to Loglinear Item Response Theory

ERRATUM

Title Research Report 94-13:

RR-94-13 W.J.J. Veerkamp & M.P.F. Berger, A simple and fast item
selection procedure for adaptive testing
Research Reports can be obtained at costs from Bibliotheek, Faculty of Educational Science and Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.