ED447200 2000-11-00 Bayes' Theorem: An Old Tool Applicable to Today's Classroom Measurement Needs. ERIC/AE Digest.

ERIC Development Team

www.eric.ed.gov

Table of Contents

If you're viewing this document online, you can click any of the topics below to link directly to that section.

ERIC Identifier: ED447200
Publication Date: 2000-11-00
Author: Rudner, Lawrence M.
Source: ERIC Clearinghouse on Assessment and Evaluation College Park MD.

Bayes' Theorem: An Old Tool Applicable to Today's Classroom Measurement Needs. ERIC/AE Digest.

Much of today's assessment research and development concentrates on norm-referenced tests which, by definition, are designed to rank-order students by placing them on broad continua representing unidimensional traits. While the summative information from norm-referenced assessment serves many purposes, there is a rising call for criterion--referenced information concerning what students know and can do relative to clearly defined desired outcomes of instruction. Although criterion-referenced interpretations of norm-referenced tests are commonplace, the literature on criterion-referenced tests from the 1970s and 80s can provide some insights to guide today's research and practice. As Hambleton and Sireci (1997) point out, the differences between the performance tests of today and the criterion-referenced tests of the 1970s are not fundamental. Both are focused on assessment of what students know and can do.

This Digest introduces ways of responding to today's rising call for criterion--referenced information using Bayes' Theorem--a method that was coupled with criterion-referenced testing in the early 1970s (see Hambleton and Novick,1973). After introducing Bayes' Theorum, I discuss how it can be applied to diagnostic testing, adaptive testing, and the scoring of performance items. The key advantages of using this model are that relatively small datasets are required and that the necessary computations are surprisingly simple.

BAYES' THEOREM

Rather than placing a student on an ability scale, the goal here is to identify the most likely classification for the examinee. This classification can be dichotomous (e.g.,master/non--master) or polychotomous (e.g., master/at--risk/non--master) or involve placement on a categorical or interval scale. To illustrate Bayes' Theorem, I will provide a simple example where the goal is to classify an examinee as being either a master or a non-master. We will use responses to previously piloted items to determine the probabilities of mastery P(M) and non--mastery P(N) and then classify the examinee based on those probabilities. Lacking any other information about the examinees, we will assume equal prior probabilities, i.e., P(M)=.50 and P(N)=.50. After each item is scored, we will update P(M)and P(N) based on the response to the item.
As givens, we will start with a collection of items for which we have determined the following four probabilities:

1. Probability of a correct response given that the examinee has mastered the material.

2. Probability of an incorrect response given that the examinee has mastered the material.

3. Probability of a correct response given that the examinee has not mastered the material.

4. Probability of an incorrect response given that the examinee has not mastered the material.

We will denote these as P(C|M), P(I|M), P(C|N), and P(I|N), respectively; note that we have different conditional probabilities for each item. These conditional probabilities can be determined from very small--scale, low--cost pilot testing; one approach is to use the proportions of examinees in each group responding correctly or incorrectly. Suppose that for item 1, 90% of the masters and 40% of the non-masters responded correctly. Since a person either responds correctly or incorrectly, P(C|M)=.90, P(I|M)=.10, P(C|N)=.40, and P(I|N)=.60.

The task then is to update P(M) and P(N) based on the item responses. The process for computing these updated probabilities is referred to as Bayesian updating, belief updating (probabilities being a statement of belief), or evaluating the Bayesian network. The updated values for P(M) and P(N) are referred to as the posterior probabilities. The algorithm for updating comes directly from a theorem published posthumously by Rev. Thomas Bayes in 1763:

$$P(M|C) * P(C) = P(C|M) * P(M)$$

Let us suppose our examinee responds correctly to item 1. The probability of a correct response, P(C), is thus 1.0 and by Bayes' Theorem, the new probability that the examinee is a master given a correct response is
P(M|C) = (.90 * .5) / 1.0 = .45

Similarly, P(N|C) = P(C|N) * P(N) = .40 * .5 = .20. We can then divide by the sum of these joint probabilities to obtain posterior probabilities, i.e.,

P'(M) = .45 / (.45+.20) = .692 and

P'(N) = .20 / (.45+.20) = .308.

We next use these posterior probabilities as the new prior probabilities, score the next item, and again update our estimates for P(M) and P(N) by computing new posterior probabilities. We iterate the process until all the items have been scored. Equivalently, we could have computed the product of the relevant probabilities (correct or incorrect) for masters and non--masters and then divided by the sum to obtain the last posterior probability.

The Bayesian network defined here is a simple diverging graph. The master/non--master state is causally connected to the set of item responses. When applied to decision support systems and other expert systems, Bayesian networks are typically much more complex, involving hundred of interconnected and cross--connected variables. Evaluating such networks is computationally complex. As we have shown here, however, the computations for basic applications are quite simple

CLASSROOM APPLICATIONS

The basic framework described above is applicable to a wide range of settings. For example, the framework can be used to score a diagnostic pretest. Here the pretest would cover a variety of skills. A pilot test would determine the probabilities of responding correctly for people who have mastered each skill and the probabilities for those who have not done so. After the test is given to an individual, the probabilities of mastery for each skill could be computed. The resultant list would identify which skills have been mastered and which are likely in need of attention. One could go further and model specific misconceptions (e.g., the examinee sums denominators when adding fractions). Here the relevant probability would be the likelihood of selecting a particular incorrect option (or generating a particular type of wrong answer), given that an examinee has a specific misconception. Such a test would not only provide mastery information, but identify specific areas to correct.

The framework is applicable to multi-dimensional items and tests. One could write items, for example, that require the application of mathematical skill to solve a science problem. A pilot test would need to be administered to compute the probability of responding correctly to each item given mastery of the mathematics skills and the probability of responding correctly to each item given mastery of the science skills. The one test with complex items could then be scored using the Bayes' Theorem and information about each skill area.

Bayesian networks have been used as the basis for computer adaptive tests. Welch and Frick (1993) provide a excellent and simple overview of the topic. Basically, the new posterior probabilities are computed after each item is administered. One stops administering items when the probability of mastery is sufficiently high or low. Items are selected from the pool of remaining items to maximize information or minimize a loss function.

The framework can be embedded in an intelligent tutoring system (ITS) to determine mastery after each instructional unit, tailor individualized instruction to characteristics of the student, and adapt that instruction as the student learns material. This would again require a collection of pre--tested items that assess the concepts covered by each instructional unit.

REFERENCES AND RESOURCES

One can easily experiment with simple Bayesian networks using any of a large variety of free, readily available software packages. A search on the Internet in September 2000 for 'Bayesian Network Software Packages' yielded more than 20 free packages that could potentially be applied. Two that I have tried are Hugin Lite and Genie.

Bayes, T. (1763). Essay towards solving a problem in the doctrine of chances.

Philosophical Transactions of the Royal Society of London, 53, 370--418.

Charniak, E. (1991). Bayesian networks without tears. AI Magazine, Winter 1991.

Frick, T. W. (1992) Computerized adaptive mastery tests as expert systems. Journal of Educational Computing Research, 8(2), 187--213.

Hambleton, R.K., & Novick, M. R. (1973). Toward an integration of theory and method for criterion- referenced tests. Journal of Educational Measurement,10(3), 159--170.

Hambleton, R.K., & Sireci, S.G. (1997). Future directions for norm-referenced and criterion-referenced achievement testing. International Journal of Educational Research, 27(5), 379--393.

Spray, J.A & Reckase, M.D. (1996). Comparison of SPRT and sequential Bayes procedures for classifying examinees into two categories using a computerized test. Journal of Educational and Behavioral Statistics, 21(4), 405--414.

Welch, R.E. and T. Frick (1993) Computerized adaptive testing in instructional settings. Educational Training Research and Development, 41(3), 47--62.

This publication was prepared with funding from the Office of Educational Research and Improvement, U.S. Department of Education, under contract ED99CO0032. The opinions expressed in this report do not necessarily reflect the positions or policies of OERI or the U.S. Department of Education. Permission is granted to copy and distribute this ERIC/AE digest.

Title: Bayes' Theorem: An Old Tool Applicable to Today's Classroom Measurement Needs. ERIC/AE Digest.
Document Type: Information Analyses---ERIC Information Analysis Products (IAPs) (071); Information Analyses---ERIC Digests (Selected) in Full Text (073);
Descriptors: Adaptive Testing, Bayesian Statistics, Criterion Referenced Tests, Test Construction
Identifiers: Bayes Theorem, ERIC Digests
###

▲

[Return to ERIC Digest Search Page]