Abstract Title Page

Not included in page count.

Title: Bayesian Decision Theory Guiding Educational Decision-making: Theories, Models and Application

Authors and Affiliations:
Yilin Pan
Teachers College, Columbia University

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context.

Given the importance of education and the growing public demand for improving education quality under tight budget constraints, there has been an emerging movement to call for research-informed decisions in educational resource allocation (Honig and Coburn, 2007; Lai and Schildkamp, 2013; Slavin, 2002). Despite the abundance of rigorous studies on the effectiveness, cost, and implementation of educational interventions, school- and district-level practitioners find that the applicability of these studies to their decision- making problems in the real educational context can be limited (Biesta, 2007; Slavin, 2002).

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

The purpose of this paper is to address three methodological challenges that hamper the influence of educational research on educational decision-making. These three methodological issues include how to generalize or extrapolate effectiveness and cost information from the evaluation site(s) to a specific context, how to incorporate information from multiple sources, and how to aggregate multiple consequences of an intervention into one framework. In particular, I introduce Bayesian decision theory, a method that processes research findings and subjective judgments using a transparent mechanism, to address these three challenges so as to provide clear, localized direction for decision makers.

The first part of this paper introduces the background and the outline of the research. The second part provides a brief description of the three components of Bayesian decision theory: maximization of expected utility, Bayes' theorem, and Multi-Attribute Utility Theory; and explains how Bayesian decision theory addresses the three methodological challenges. The third section transfers the idea of Bayesian decision theory to a statistical model under certain assumptions. The fourth section applies the proposed statistical model to a typical decision problem of choosing between two reading programs, and demonstrates the estimation, presentation and interpretation of the results. The last part summarizes the advantages of Bayesian decision theory. In all, the ultimate goal of this paper is to explore some methodological improvements to the current evaluation framework that may contribute to lowering the barriers between educational research and educational decision-making and bringing the two worlds into greater alignment.

Significance / Novelty of study:

Description of what is missing in previous work and the contribution the study makes.

Bayesian decision theory, also referred to as decision analysis, is a combination of multiple analytical approaches based on the subjective expected utility model (French, 2008). The underlying assumption of the theory is the rationality of belief under uncertainty, i.e., among all the alternatives, an economic person would choose the one that maximizes her expected utility (Hausman and McPherson, 2006). Based on this assumption, decision analysis utilizes

Bayesian inference to update people's belief on the uncertainty, and uses Multi-attribute Utility Theory to model decision maker's preference (French, 2008). Edwards (1998) collectively defines these three components, Bayes' theorem, Multi-Attribute Utility Theory and maximization of expected utility, as Bayesian decision theory, and asserts that "the 21st century will be the Century of Bayes", in which "the explicit use of these three formal decision models by decision makers will be as commonplace as use of spreadsheets is now" (p.416). In the paper, the three components and the relationship between the three components will be explained in detail.

While Bayesian decision theory has been widely adopted in many fields to solve choice problems extending from nuclear emergency response (French, 1996) and health-care evaluations (Spiegelhalter, 2004) to selection of manual wheelchair (Delcroix et al., 2013) and design of Six Sigma (Rajagopal and Castillo, 2007), its application to educational decision-making has been relatively scant. Girshick (1954) introduced the basic idea of statistical decision theory to the field of education, and described how the combination of Bayes' theorem and the loss function offers a statistical framework to model rational behavior under uncertainty by accounting for both costs and consequences. However, the theory is illustrated based on a coin-toss example, so it does not provide a concrete example on how to apply it to the specific educational decision-making context. Duff and Lynch (1977) combined Bayesian and cost-benefit analysis and demonstrated how to identify the optimal cutoff in GRE scores to minimize the opportunity loss based on a bivariate normal model. Probably the first educational study that formally adopted the framework of Bayesian decision theory, Saar (1980) specified a step-by-step approach to identify the utility function and update the expected utility in the light of data using Bayes' theorem. These pioneering studies have opened the door to explore the potential of Bayesian decision theory in the educational context; however, due to underdevelopment of Bayesian computation before 1990s, the models demonstrated are still very primitive.

Winkler (2001) pointed out that one of the primary reasons that the application of Bayesian decision theory in health care evaluations falls behind the burgeoning development of Bayesian computation is the lack of demonstrative models of analysis that are readily available; the same argument may also apply to the field of education. Realizing the gap in the extant literature, more research is needed in at least three aspects. First, it is necessary to explore how to construct the prior distributions by summarizing existing research evidence and eliciting subjective judgments. Second, provided that prediction is heavily involved in decision-making, more exploration is needed in model checking, model comparison and model extension with the purpose of increasing the out-of-sample prediction power. Third, how to construct the utility function is also an essential component in the application. To fill in these gaps, this paper is going to propose a demonstrative statistical model that may help boost the application of Bayesian decision theory in real educational decision-making practice.

Statistical, Measurement, or Econometric Model:
Description of the proposed new methods or novel applications of existing methods.

Mathematically, Bayesian decision theory is described as Formula (1) and (2) (revised based on French, 2008; Gelman et al., 2013; Leonelli and Smith, 2013).

$$\begin{aligned} max_{a\in A}E(U(c,\theta|X,a)) &= max_{a\in A}\int_{\Omega} U(c,\theta|X,a)df(U(c,\theta|X,a)) && (1) \\ &= max_{a\in A}\int_{\Omega} U(c,\theta|X,a)f(\theta|X)d\theta && (2) \end{aligned}$$

where θ is a vector of all unknown parameters with the sample space of Ω; a is a vector of alternatives with the sample space of A; X represents a matrix of observed data, such as the effectiveness and cost data for a reading program in the example; c is a vector of measurable quantities that map the consequences of each alternative into numbers; U (.) represents the (multi-attribute) utility as a function of the consequences c and θ; and f (.) is the probability density function. As shown in Formula (2), the expected utility can be decomposed into two parts: a utility function *U* (.) that represents desirability or preference, and a probability distribution *f* (.) that models uncertainty. *U* (.) is usually constructed based on Multi-Attributed Utility Theory, and *f* (.) is updated using Bayes' theorem in the light of data *X*.

1. Uncertainty model *f* (.)

Identification of the Likelihood: The first step of Bayesian inference is to set up a joint probability distribution for all the observable (i.e., X) and unobservable quantities (i.e., θ) in a problem (Gelman et al., 2013). To achieve this in the Bayesian decision theory model, it is necessary to identify the variable types of the consequence measures and explore how they are jointly distributed. Since modeling the joint distribution of a likelihood that combines different variable types is demanding, the consequence measures in the following models are restricted to continuous or dichotomous variables that are expressed in effect size, which are also the dominating variable types of consequence measures in education. In addition, it is also necessary to make a reasonable parametric assumption on the joint distribution of the likelihood. In the paper, I will explain when and why it is reasonable to assume that the consequences of interest are distributed as a multivariate normal or a multivariate t distribution. Tests to check multivariate normality and multivariate t will also be introduced. Hierarchical models will be expressed mathematically under the assumption of multivariate normal and multivariate *t* joint distributions of the likelihood.

Identification of the Priors: Under the Bayesian framework, both quantitative and qualitative information is expected to be transformed into mathematical prior distributions. Prior distributions based on quantitative empirical studies can be derived via meta-analysis, while priors from other sources (e.g., expert opinions, testimonies, implementation reports, case studies, contextual information, etc.) rely on the elicitation of the decision maker's subjective judgments (Congdon, 2008). Priors can be very flexible. When the decision maker prefers not to incorporate any prior information and relies on the data to tell the story, one can use a non-informative prior or a weakly-informative prior to model her "ignorance" (Kaplan and Park, 2013). Sensitivity analysis may also be necessary to check how the final conclusion responds to a change of priors (Spiegelhalter et al., 2004).

2. Desirability model *U* (.)

One Attribute Utility Function: The construction of the utility function can be divided into two steps. The first step is to map each consequence onto a one-attribute utility function. The function form of the one-attribute utility function is determined by the decision maker's attitude towards risks and uncertainty (Keeney and Raiffa, 1993). In the paper, I will propose a monotonically increasing utility function for consequences that represent gains (e.g., the impact of a program on the fluency scores or comprehension scores), indicating that the decision-maker is decreasingly risk-averse for these gains. As to consequences that represent loss (e.g., cost), I will propose a monotonically decreasing utility function which implies that the decision-maker is increasingly risk-averse for these losses.

Multi-attribute Utility Function: After identifying the utility function for each attribute, the second step is to combine them together based on one's assumptions regarding the dependencies of these attributes. A linear function and a bilinear function will be proposed based on Edwards and Newman (1982).

3. Combination of the Uncertainty Model and the Desirability Model

I will explain several additional steps to plug the posterior predictive distributions of the consequences generated from the uncertainty model into the desirability model so that the posterior predictive distribution of utility can be displayed.

Usefulness / Applicability of Method:
Demonstration of the usefulness of the proposed methods using hypothetical or real data.

The model is applied to a choice problem of after-school reading programs: Reading Partners and Fast ForWord. Given the small sample size of the two dataset (e.g., 6 sites for Reading Partners and 3 sites for Fast ForWord), I will simulate two large datasets with n = 50 based on the mean vectors and the variance-covariance matrix of all the consequence variables in the observed datasets. The purpose of enlarging the sample size is to make the parameters in the models estimable. The example will illustrate how to estimate the model and present the results.

Conclusions:
Description of conclusions, recommendations, and limitations based on findings.

In summary, Bayesian decision theory is a mechanism to process research evidence as well as subjective judgments in a decision-making context. It may help bridge the gap between research and practice by localizing the evidence and making it more context- and user- dependent in three ways. First, the extrapolation of effect and cost information indicates what is likely to occur at the specific site of interest. Second, by constructing subjective priors, the decision maker can decide what prior information is taken into account for her decision-making, and how much each kind of information contributes to the decision-making process. Third, the subjective utility function enables the decision maker to illustrate how much she values each consequence. Since Bayesian decision theory is aimed at generating a solution that works best for a specific decision maker in a specific context, its power lies in guiding the decision maker to "get her head straightened out", rather than persuading other people from an objective perspective. The localization of research evidence is expected to formalize the process of communication and interaction between researchers and practitioners, pushing the two worlds into greater alignment.

Appendices

Not included in page count.

Appendix A. References

Biesta, G. (2007). Why "what works" won't work: evidence-based practice and the democratic deficit in educational research. *Educational Theory*, 57(1):1–22.

Congdon, P. (2008). Bayesian statistics in quantitative risk assessment. In *Encyclopedia of quantitative risk analysis and assessment*. John Wiley & Sons, Ltd, London; Hoboken, NJ.

Delcroix, V., Sedki, K., and Lepoutre, F.-X. (2013). A Bayesian network for recurrent multi-criteria and multi-attribute decision problems: choosing a manual wheelchair. *Expert Systems with Applications*, 40(7):2541–2551.

Duff, W. L. and Lynch, R. M. (1977). Graduate admission policy: a Bayesian analysis. *The Journal of Experimental Education*, 45(3):38–41.

Edwards, W. (1998). Hailfinder: tools for and experiences with Bayesian normative modeling. *American Psychologist*, 53(4):416–428.

Edwards, W. and Newman, J. R. (1982). *Multiattribute evaluation*. Sage Publications, Beverly Hills.

French, S. (1996). Multi-attribute decision support in the event of a nuclear accident. *Journal of Multi-Criteria Decision Analysis*, 5(1):39–57.

French, S. (2003). Modelling, making inferences and making decisions: the roles of sensitivity analysis. *Top*, 11(2):229–251.

French, S. (2008). Decision analysis. In Everitt, B. S. and Melnick, E. L., editors, *Encyclopedia of quantitative risk analysis and assessment*. John Wiley & Sons, Ltd, London; Hoboken, NJ.

Gelman, A., Hwang, J., and Vehtari, A. (2013b). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6):997–1016.

Girshick, M. A. (1954). Chapter VII: An elementary survey of statistical decision theory. *Review of Educational Research*, 24(5):448–466.

Hausman, D. M. and McPherson, M. S. (2006). *Economic analysis, moral philosophy, and public policy*. Cambridge University Press, Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo, Delhi, 2nd edition.

Honig, M. I. and Coburn, C. (2007). Evidence-based decision making in school district central offices: toward a policy and research agenda. *Educational Policy*, 22(4):578–608.

Kaplan, D. and Park, S. (2013). Analyzing international large-scale assessment data within a Bayesian framework. In Rutkowski, D., editor, *Handbook of international large-scale assessment: background, technical issues, and methods of data analysis*, pages 547–582. Chapman and Hall/CRC, London, UK; San Diego, USA; Walham, USA; Oxford, UK.

Keeney, R. L. and Raiffa, H. (1993). *Decisions with multiple objectives: preferences and value tradeoffs*. Cambridge University Press, Cambridge, New York.

Lai, M. K. and Schildkamp, K. (2013). Data-based decision making: an overview. In Schildkamp, K., Lai, M. K., and Earl, L., editors, *Data-based decision making in education*, number 17 in Studies in Educational Leadership, pages 9–21. Springer, Dordrecht, Heidel- berg, London, New York.

Leonelli, M. and Smith, J. Q. (2013). *Using graphical models and multi-attribute utility theory for probabilistic uncertainty handling in large systems: with application to the nuclear emergency management*, pages 181–192. IEEE.

Rajagopal, R. and Castillo, E. d. (2007). A Bayesian approach for multiple criteria decision making with applications in design for Six Sigma. *The Journal of the Operational Research Society*, 58(6):779–790.

Saar, S. S. (1980). The role of decision-analysis and Bayesian approach in planning and evaluating education programs. *Educational Evaluation and Policy Analysis,* 2(3):63–72.

Slavin, R. E. (2002). Evidence-based education policies: transforming educational practice and research. *Educational Researcher,* 31(7):15–21.

Spiegelhalter, D. J. (2004). Incorporating Bayesian ideas into health-care evaluation. *Statistical Science,* 19(1):156–174.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation*. John Wiley & Sons, Chichester.

Winkler, R. L. (2001). Why Bayesian analysis hasn't caught on in healthcare decision making. *International journal of technology assessment in health care,* 17(1):56–66.

Appendix B. Tables and Figures

Not included in page count.