PSYCHOMETRIKA—VOL. 82, NO. 4, 1182–1184
DECEMBER 2017
https://doi.org/10.1007/s11336-017-9584-7

BOOK REVIEW

ALMOND, R. G., MISLEVY, R. J., STEINBERG, L. S., YAN, D., & WILLIAMSON, D. M (2015). *Bayesian Networks in Educational Assessment. NY: Springer. ISBN: 1493921258. DOI:http://doi.org/10.1007/978-0-387-98138-3*

Bayesian networks found their applications in many fields: in medicine, for medical decision making (Lucas 2001), in Artificial Intelligence, for robots and object recognition (Schneiderman 2004), in ecology, for environmental modeling (Aguilera et al. 2011), and in educational assessment, for measuring students' abilities (Mislevy et al. 2000). The network method provides a logical and graphical structure for reasoning under uncertainty (Neapolitan 2003; Pearl 1988). For instance, Lucas previously applied these methods in medicine, where symptoms are entered in a network as observable variables. By formulating probability distributions and a network structure, the most likely disease and a fitting treatment for patients could be determined. In Artificial Intelligence, Schneiderman used Bayesian networks to automatically learn network structures with the goal to detect frontal faces, eyes, and the iris of the human eye. These applications motivate the use of Bayesian networks to assess uncertainty associated with (human) environmental interventions and classifications. Uncertainty around complex relationships is found in many fields and can also arise in educational assessment settings. Innovative educational assessments, such as interactive simulations, games or complex constructed responses, can benefit from Bayesian network methodology. The book "*Bayesian Networks in Educational Assessment—The State of the Field*" provides an exhaustive overview of applying Bayesian networks in educational measurement. A unique reference work has been written aimed at professionals, researchers and students in this field.

The book consists of three main parts. Part I introduces the building blocks of Bayesian Networks in Educational Assessment. Firstly, the evidence-centered design (ECD) approach (Mislevy et al. 2004) is outlined and explained. Then, Part I introduces Bayesian probability and graphical networks, after which methods for efficient calculations, examples, and inferences by the model are described. Part II focuses on learning and revising the network and student models when actual student data has been observed. Different model parameterizations are described. These parameterizations will definitely contribute to the readers' understanding of the choices that could be made in construction and updating of the models. Estimating distributions of parameters, and approaches to model fit or model evaluation are also introduced. This part concludes with a complex example that illustrates the concepts that were discussed in the previous chapters. After dealing with the technical and mathematical issues of Bayesian networks, Part III is about application of Bayesian networks in assessment design. It is discussed in detail how Bayesian networks can be integrated in the various steps of the evidence-based design approach.

The three parts of the book give an excellent overview of current developments and the status of a still developing and emerging field. In this way, it helps researchers in psychometrics and educational assessment to familiarize themselves with Bayesian networks, and their applications. It is mentioned by the writers that Bayesian networks have not been operationalized very often in educational assessment just yet, but this book gives good and extensive instructions. Besides, ideas on the methods and possible extensions are mentioned. The book combines both mathematical and applied issues. Therefore, it is a very practical and inspiring book, which gives the current status of research around Bayesian networks in education and a solid background in theory. The targeted audience of the book seems to focus on students at the graduate level. In this way, it is written for a large audience, and it suits the needs of many researchers. In our opinion, it might have been wise to add 'with an interest in quantitative methods' to the description of the intended audience. The level of the content in the book is quite advanced, in the sense that the reader has to understand probabilistic graphical models, basic probability theory, and the evidence-centered design framework thoroughly, before being able to understand the later chapters in the book.

We are therefore very positive about the three introductory chapters on the required topics (ECD, Bayesian probability, and graph theory) in Part I of the book, before the actual practical part of building, estimating and learning networks for educational assessment applications are discussed in Part II and Part III. Besides, every chapter ends with a set of exercises, and the readers are urged to complete them. The authors also mention that applying the theory to new applications is the best way to learn. In this way, students and researchers who are new to the area will be equipped to master the complicated topic of BNs in Educational Assessment.

Two co-authors of this book review used the book for their study in Bayesian network modeling (De Klerk et al. 2015). The goal of this study was to estimate student abilities by rating their performance in a simulation-based assessment. A simulation-based assessment is a computer simulation in which students have, to some extent, a free space to act in. For example, to answer questions and fulfill assignments. De Klerk et al. (2015) addressed two specific challenges: the evidence identification challenge (i.e., scoring the interactive performance of students in the simulation), and the evidence accumulation challenge (i.e., accumulating scores in a psychometric model). The used psychometric model was a Bayesian network. They built two Bayesian networks to estimate the state of students' latent variables after students had completed a simulation-based assessment.

The simulation-based assessment was used to measure students' practical skills for working in an industrial vocation. Students' actions in the computer simulation were regarded as the observable variable in the networks that informed the state of several latent variables. The dependencies between observable and latent variables were defined by a group of subject matter experts. The conditional relationships between these random variables were determined by experts' ranking of the difficulty and essence of all actions in the different stages of the simulation-based assessment. By combining those rankings, the authors could build two different Bayesian networks. The book Bayesian networks in Educational Assessment was very helpful during the design and development of the Bayesian networks. Especially the fact that the book provides such a nice balance between theory and practice makes this book a very worthwhile contribution to the scientific literature on this topic. Furthermore, it was helpful that the second part of the book provides information on parameters for Bayesian networks and finishes with an illustrative example.

The book gives several examples on building network models. The illustrated examples of building networks on "mixed number subtraction" of Tatsuoka (1983) and the Biomass Measurement model contribute to the theory and makes it easier to construct networks later on. It has to be mentioned that, even after reading the book, the process of network construction will still require creativity due to the nature of the technique. This seems to be the case, since many assumptions on the initial structure of a Bayesian network are necessary in future steps of constructing a model. In general, the process of constructing networks seems to rely on several knowledge domains. This causes the reader of the book to have many skills before being able to construct networks. These skills are for example; the availability of domain knowledge of the subject tested, possessing knowledge on the network construction process and to preferably having additional experts on the topic available to increase the validity of the network and its assumptions.

A number of software packages are used throughout the book, but most of the examples for the book are prepared in *StatShop*; an internal Bayesian network program developed at ETS for building, calibrating, scoring and experimenting with assessments. Other network construction software which are used are the programs *Netica* and *GeNie and Smile* for manipulation purposes. Manual construction of networks is also possible and can be done with the R packages CPTtools, RNetica, or SSX. Finally, estimating Bayesian network parameters with Markov Chain Monte Carlo (MCMC) can be done either with built-in algorithms in the previously mentioned construction software or with stand-alone software with MCMC algorithms such as OpenBUGS or JAGS. The authors make clear that there is a large amount of possibilities for software in Appendix of the book on ‘Bayesian Network Resources’ and keeping up with all software packages is challenging. Therefore, they direct the readers to their online web site with a Wiki page on tools for Bayesian networks and examples of networks. Step-by-step instructions on the above mentioned packages are not provided by the book or the website. This requires researchers willing to build their own network for assessments to dive into the possibilities and instructions in other resources.

A topic that did not receive that much attention in the book is the relation of Bayesian networks with other methods. Bayesian networks can be related to other graphical models like structural equation models (SEM), and they can even be applied to circumvent issues related to maximum likelihood estimation in SEM. This relationship is only briefly discussed. Also the relation of Bayesian networks with classic machine learning methods like hidden Markov models (HMMs), neural networks, and Kalman filters is not really discussed in much detail. Besides, a chapter on fitting and updating network model structures and parameters with no fixed structure might have made the book more complete. The book contains a chapter on “Learning in Models with Fixed Structure,” while a starting researcher in the field might also be interested in more flexible structure and parameters for the network model. The subjectivity of expert models and panels is assumed by the ECD framework, but the role of data and the exact difference it would make to either incorporate expert knowledge or data could have been made more explicit for didactic purposes.

In summary though, we feel that this book is a great contribution to the use of Bayesian networks in educational measurement. Bayesian networks are introduced in great detail and the reader is equipped with several tools to apply them in the practice. It has to be kept in mind that there are many directions in the research and application of Bayesian networks in educational measurement. The thorough description of the evidence-centered design approach and the application of BNs within this framework does further increase the reader’s understanding of both the complexity and the opportunities of the applications.

UNIVERSITY OF TWENTE / CITO — *Nikky van Buuren*
EX:PLAIN — *Sebastiaan de Klerk*
UNIVERSITY OF TWENTE — *Bernard P. Veldkamp*

References

Aguilera, P. A., Fernández, A., Fernández, R., Rumí, R., & Salmerón, A. (2011). Bayesian networks in environmental modelling. *Environmental Modelling & Software*, *26*, 1376–1388.

De Klerk, S., Veldkamp, B. P., & Eggen, T. J. H. M. (2015). Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example. *Computers & Education*, *85*, 23–34.

Lucas, P. (2001). Bayesian networks in medicine: A model-based approach to medical decision making. In *Proceedings of the EUNITE Workshop on Intelligent Systems in Patient Care, Vienna*, pp. 73–97.

Mislevy, R. J., Almond, R. G., & Lukas, J. (2004). A brief introduction to evidence-centered design. CSE Technical Report. Los Angeles: The National Center for Research on Evaluation, Standards, and Student Testing (CRESST). http://www.cse.ucla.edu/products/reports/r632.pdf.

Mislevy, R. J., Almond, R. G., Yan, D., & Steinberg, L. S. (2000). Bayes nets in educational assessment: Where do the numbers come from? CSE Technical Report, 518.

Neapolitan, R. E. (2003). *Learning Bayesian networks*. New York, NY: Prentice-Hall.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco, CA: Morgan Kaufmann.

Schneiderman, H. (2004). Learning a restricted Bayesian network for object detection. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on IEEE* (Vol. 2, pp. II–II).

Tatsuoka, K. K. (1983). Rule-space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*, 34–38.

Manuscript Received: 23 DEC 2016
Final Version Received: 16 MAY 2017
Published Online Date: 15 SEP 2017