PREFACE

Special issue: Bayesian applications to educational and psychological testing

Kazuo Shigemasu[1] · Wim J. van der Linden[2]

Bayesian approaches have become the standard of industry in nearly every branch of statistics. In fact, they have already penetrated many aspects of our daily lives. For instance, they are now used to filter our email messages for malicious content, suggest the purchase of items based on past shopping behavior, run surveillance video searches, and support both speech recognition and machine translation of natural language. These applications are not universally good yet and, frankly, sometimes even bothersome. But where they do confuse us as statisticians, it is through the choice of probabilistic model or source of data rather than their underlying Bayesian methodology.

For the more substantive sciences, Bayesian statistics has provided sound directions to every aspect of the process of research planning, including such key activities as optimal research design, parameter inference, model evaluation, hypothesis testing, and decision making. It does so without nearly any practical limit on model complexity. The main reason for its success resides in the choice of probability calculus as the only valid logic of statistical inference while consequently accounting for model uncertainty both badly missed in the rather ad hoc collection of inferential arguments that used to populate the statistical literature.

The field of test theory has become important to modern-day society. In fact, it has grown into the statistical underpinning of major events at nearly every intersection in our daily lives, whether it be admission to a school, a career decision, certification for a job, evaluation of educational progress, work placement, or choice of treatment. Test theory has acquired this role because of its response models with parameters explicitly representing the effects of abilities, task difficulty, rater strictness, time pressure, or any other relevant determinant of human performances.

This special Behaviormetrika issue demonstrates cutting-edge research by authors who rely on Bayesian approaches to solve real-world problems associated with educational and psychological testing. More specifically, the first article

✉ Kazuo Shigemasu
   kshigemasu@gmail.com

[1] Keio University, Tokyo, Japan

[2] University of Twente, Enschede, Netherlands

by (Sinharay and Johnson 2020) demonstrates the use of Bayes factors to decide on the score difference between a subset of educational test items suspected to be compromised and the regular part of the test. The article offers results from a simulation study demonstrating the power of the factor for the detection of test fraud. It also illustrates its practical applicability using a real-data example.

The next article by (Luby et al. 2020) opens up an entire new application of item response theory: fingerprint identification in forensic analysis. Though item response models were developed for the analysis of educational and psychological test, they fit this new type of application beautifully. The question whether an examiner is able to match a new fingerprint with a print of known origin on file has a response distribution that also is a function of parameters characterizing both the examiner and features of the print. The authors present an informative set of analyses based on a range of standard response models as well as new modifications introduced by them.

The question whether item parameters, and hence the ability scale, remain invariant when moving from one group of examinees to another is one of the core problems of measurement in the behavioral sciences. In the third article, (Fox et al. 2020) address the question through modeling of the covariance structure of multigroup data. Their model includes random group-specific parameters the variation of which points at violation of the assumption of measurement invariance. The authors use fractional Bayes methodology to check how the data provide support to the hypothesis of no violation.

The fourth article by (Mayekawa and Yamashita 2020) presents regression analysis also developed for a multigroup setup. The interesting feature of their main model is separation of regression parameters common to all groups from parameters representing unique features of each group. The model is claimed to have easier parameter interpretation as well as improved predictive efficacy. The authors show how the claim is supported by both simulation study and real-data analysis.

To date, the Bayesian approach to adaptive testing has been empirical Bayes with point estimates substituted for all the item parameters. The main reason for the restriction was the computational demands involved in real-time updates of all model parameters. In the fifth article, (Ren et al. 2020) investigate an algorithm for Bayesian adaptive testing with polytomous test items that does produce updates of all model parameters and selects the next item for administration in milliseconds. The core of the algorithm is a Gibbs sampler with resampling of the posterior distributions of the incidental parameters and an MH step for the structural parameters. The sequential nature of adaptive testing is exploited to efficiently implement the MH step.

The sixth article by (Shigemasu et al. 2020) investigates the factor structure of one of the best-known intelligence tests, i.e., the Wechsler Intelligence Scale for Children (WISC). Their research tries to answer the century-old question whether the higher-order structure of intelligence is essentially unidimensional (g-factor) or plural. The authors show how a Bayesian approach to determining the factor structure of intelligence allows us to integrate theoretical considerations with empirical information and, in doing so, provides insightful hypotheses about the organization of intelligence.

The evaluation of test performances in the domains of problem solving, critical reasoning, and creative thinking is typically done using human raters. The article by (Uto and Ueno 2020) offers a new IRT method to obtain more accurate ability estimates from these rating tasks. Specifically, the authors propose a generalized many-facet Rasch model which, in addition to examinee and item parameters, incorporates three important rater characteristics. To demonstrate the practical feasibility of the model, the authors use a Bayesian estimation method with No-U-Turn Hamiltonian Monte Carlo for their analyses of both simulated and real-world data.

The last article by (Yamaguchi and Okada 2020) is a good example of how Bayesian hierarchical modeling allows for the integration of different models as a mixture. The two models used by these authors are a conjunctive and disjunctive latent-class model in use as base models in cognitive diagnostic testing. Both a simulation study and real-data analysis were conducted to demonstrate the benefit of the model integration.

The editors of this special issue hope it will contribute to the further extension of the use of Bayesian approaches both in the more fundamental research in educational and psychological testing and throughout its many applications.

References

Fox JP, Koops J, Feskens R, Beinhauer L (2020) Bayesian covariance structure modeling for measurement invariance testing. Behaviormetrika. https://doi.org/10.1007/s41237-020-00119-3

Luby A, Mazumder A, Junker B (2020) Psychometric analysis of forensic examiner behavior. Behaviormetrika. https://doi.org/10.1007/s41237-020-00116-6

Mayekawa S, Yamashita N (2020) Bayesian reduced rank multigroup regression analysis: a new model for multigroup data with hybrid parameter sharing. Behaviormetrika. https://doi.org/10.1007/s41237-020-00112-w

Ren H, Choi SW, van der Linden WJ (2020) Bayesian adaptive testing with polytomous items. Behaviormetrika. https://doi.org/10.1007/s41237-020-00114-8

Shigemasu K, Kono M, Ueno K (2020) Bayesian confirmatory factor analysis of Wechsler Intelligence Scale for children data. Behaviormetrika. https://doi.org/10.1007/s41237-020-00108-6

Sinharay S, Johnson MS (2020) Detecting test fraud using Bayes factors. Behaviormetrika. https://doi.org/10.1007/s41237-020-00113-9

Uto M, Ueno M (2020) A generalized many-facet Rasch model and its Bayesian estimation using Hamiltonian Monte Carlo. Behaviormetrika. https://doi.org/10.1007/s41237-020-00115-7

Yamaguchi K, Okada K (2020) Hybrid cognitive diagnostic model. Behaviormetrika. https://doi.org/10.1007/s41237-020-00111-x