

UC Irvine

UC Irvine Previously Published Works

Title

Comment

Permalink

<https://escholarship.org/uc/item/98q9r4c3>

Journal

The American Statistician, 68(2)

ISSN

0003-1305

Author

Stern, Hal

Publication Date

2014-04-03

DOI

10.1080/00031305.2014.897257

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



Published in final edited form as:

Am Stat. 2014 ; 68(2): 83–84. doi:10.1080/00031305.2014.897257.

Discussion of “The Need for More Emphasis on Prediction: A ‘Nondenominational’ Model-Based Approach”

Hal Stern

Department of Statistics, University of California, Irvine

David Harville has provided a compelling case for an increased focus on prediction in the teaching of statistics. I am very sympathetic to Harville’s plea. Indeed, it was the ability of statistical methods to address prediction problems (in sports and finance, the same fields that Harville mentions) that attracted me to the field of statistics more than 30 years ago. Even in application areas where the focus has been on parameter estimation, e.g., regression coefficients in economics or treatment effects in clinical trials in medicine, it seems quite natural to me to think of these parameter estimates in terms of the predictions that they imply. Given that I agree with Harville on the central role of prediction, my comments below concern his focus on models and the relevance of the “nondenominational” approach.

The model-based approach

The article presents a thorough treatment of model-based inference and the role it can play in predictive settings. I certainly can’t argue with the ability of standard statistical models to perform well in the applications that Harville describes; in fact, I’ve published work influenced by Harville applying the linear mixed model to ranking sports teams (Glickman and Stern, 1998) and to ranking animals in the breeding context (Reber et al., 2000). The results in the former case are very similar to those that are obtained by Harville here and these results strongly support the argument that such models can be extremely useful in predictive settings. Harville’s article though raises some important issues about the model-based approach. The role of parametric statistical models in traditional inferential settings is well known but the role of parametric statistical models when we are focused on prediction can be different. Indeed, this is the lesson that I take from Breiman (2001). Just as there are settings where standard components of the statistician’s tool kit (e.g., the linear mixed model) work well, there are also settings where statisticians should be applying newer algorithmic tools (e.g., regression trees). It is important to note that these newer tools often do correspond to models but definitely not the standard ones. Important problems like handwriting recognition (e.g., automatically identifying zip codes in post office operations) and computer vision are clearly prediction-based. Breakthroughs in these areas have required new types of models to address for example how shapes are characterized, and how to realistically generate objects and scenes. Bayesian nonparametric models are another example of the types of tools that analysts seeking flexible models have found useful. The models that have been used in these settings can still be thought of as parametric models but they typically use very large numbers of parameters to avoid assumptions regarding the appropriate distributional family. Care is required in the application of such “nonparametric” models to insure that practitioners don’t overfit to the training data - but this is true also for

standard models of course. I welcome Harville's reminder that we should not underestimate the role of parametric models in predictive settings. I hope he will agree that asking for more focus on predictive inference should also encourage practitioners to continue developing more flexible tools that can perform well.

Statistical denominations

Harville's treatment of the different approaches to statistical inference is somewhat subtle. On a first reading I did not completely appreciate the "nondenominational" argument that was being made. On rereading it carefully though I became a bit concerned. Although I agree with Harville that the vast and sometimes acrimonious literature contrasting Bayesian and frequentist methods has not been helpful to the field I'm less confident that the approach presented here addresses the question in convincing fashion.

I should first start by self-identifying as a Bayesian, one for whom performance in repeated applications is relevant. This includes both repetitions under the usual repeated-sampling framework but also repetitions in the sense of predicting new outcomes. This should make me more open to Harville's approach and there are elements of Harville's nondenominational argument that resonate with me. For example, it is general enough to easily handle the hierarchical model formulations which are a critical element of modern Bayesian inference (see, e.g., Gelman et al. 2013). The bulk of the discussion, however, in this nondenominational "chapel" does not seem particularly friendly to the Bayesians in attendance. Partly this is because of the language being used. Harville notes that the distinction between model and prior distribution is somewhat arbitrary. I don't believe this is a fair criticism. The distinction can appear arbitrary because the term "model" itself is somewhat ambiguous – it is frequently used, as it is here, as a catchall to denote all of the distributional assumptions being made. If so then it's true that the usual Bayesian prior distributions are included in the model. Bayesians make a distinction in practice by separating the data model which describes the distribution of the random observable y and the remaining elements of Harville's model which are then denoted as the prior distribution. This seems quite similar to the way Harville develops his hierarchical model thus perhaps the distinction is not quite so artificial. Indeed the choice to model the parameters as a draw from the population is an assumption, a form of additional information available before data collection that leads us to believe the parameters ought to be related in this way. When described in this way the assumption seems natural as part of the Bayesian's prior distribution.

Beyond questions of terminology my biggest concern is that Harville's discussion does not hit on two key benefits of modern Bayesian inference both of which are useful for the prediction context. First, I am surprised that simulation-based inference (e.g., Markov chain Monte Carlo methods) was not mentioned. Computational advances have made it possible to draw inferences for complex models without relying on standard distributional forms or large sample asymptotics. This is especially relevant at the end of Section 2. Simulations also provide for a wide range of inferential possibilities – thus it is possible via simulation to easily obtain posterior intervals for a team's ranking or to estimate the posterior probability that one team is ranked more highly than another. Second, Harville's nondenominational

approach does not directly address what to do with missing data, values of y that were intended to be collected but were not, or with additional information about the method used for data collection (e.g., censoring, stratified samples, etc.). This not to say these topics can't be addressed in his framework, my aim is merely to point out that I've found it conceptually straightforward to modify models to address such concerns utilizing Bayesian ideas (see Gelman et al. 2013).

My intent here is not to add to the divisive literature that both Harville and I dislike. I genuinely like Harville's approach to the ranking application in Section 3 and would trust his nondemoninational approach to perform well in a wide range of problems. I like to believe he would equally trust my ability to apply Bayesian methods in such settings. My experience teaching and practicing statistics has lead me to the conclusion that although Bayesian and frequentist methods are closely related, they are different enough that a nondenominational approach probably won't work. Instead I'd argue for pluralism; there are many appropriate and effective approaches to addressing applied problems and these should be welcomed and accepted.

Let me close with a brief note of what is usually referred to as "personal communication" in the references. I recall a conversation that I had with David Harville approximately 20 years ago after I presented a seminar on Bayesian methods to the Animal Breeding and Genetics Seminar at Iowa State University. At that time he agreed that some of the benefits of the Bayesian approach described above (simulation-based computation, flexible inference) were attractive but remarked that he was not prepared to specify the prior distribution required to reap those benefits. It is an opinion that I respected then and continue to respect today but an example of why he and I will probably remain in different denominations.

Acknowledgments

This work was supported in part by NSF grant AGS1025374 and NIH grant P50-MH096889.

References

- Gelman, A.; Carlin, JB.; Stern, HS.; Dunson, DB.; Vehtari, A.; Rubin, DB. Bayesian Data Analysis. London: CRC Press/Chapman and Hall; 2013.
- Reber DL, Stern HS, Berger PJ. Bayesian Inference for the Mixed Linear Model with Application to Selection in Animal Breeding. Journal of Agricultural, Biological, and Environmental Statistics. 2000; Vol. 5:240–256.