Discussion of "A Gibbs Sampler for a Class of Random Convex Polytopes"

Jonathan P Williams

Published online: 07 Sep 2021.

Submit your article to this journal ⍈

View related articles ⍈

View Crossmark data ⍈

Taylor & Francis
Taylor & Francis Group

Check for updates

Discussion of "A Gibbs Sampler for a Class of Random Convex Polytopes"

Jonathan P Williams

Department of Statistics, North Carolina State University, Raleigh, NC

1. Introduction

I commend the authors for their efforts in revisiting the DS framework for statistical inference, but from a computational perspective. It is perhaps untraditional for the Theory and Methods section of the *Journal of the American Statistical Association* to publish an article with contributions that are without methodological innovation, but purely computational in nature. This speaks to the importance of the problem the authors address; it seems that this has been an open problem for many decades.

I begin my comment piece with a reiteration of the authors' concluding remark that the field of statistics lacks a unifying foundation for parametric inference, and I add, for how to understand data more broadly. Much of foundations research is thought of as a novelty by the broader statistics community; philosophical discussions about how to think about data are dismissed as inconsequential and peripheral to contemporary data problems. This has largely been driven in modern decades by efforts to chase the algorithmic development success of more computationally oriented experts in the mathematical and computer sciences in their efforts to solve engineering data problems/tasks.

The significance of foundations questions have also been obscured by an obsession, in statistics communities, with appeals to asymptotic justification for methods research, and a lack of emphasis for studying finite-sample properties of new methods. Arguably, the distractions from addressing more foundational questions in statistical inference is in large part because these types of investigations do not lend themselves to be packaged as succinct or affirming mathematical statements/characterizations. For instance, how to analyze a model/variable selection method in the absence of a sparsity assumption? Surely, sparsity is *not* an assumption that is believed to be true as often as it is taken as an assumption in academic papers. This is an issue that I have attempted to address in the series of articles (Williams and Hannig 2019; Williams, Xie, and Hannig 2019; Koner and Williams 2021), as well as in ongoing work.

Nonetheless, rather than developing an argument, here, for the relevance and importance of foundations research, in what follows are comments organized into sections focused on important contributions, central questions, and prevailing matters of the article under discussion.

2. Important Contributions

The genius of the article under discussion is the insightful connection made between the expression for the data generating, or structural equation of a categorical random variable,

$$x_n = \sum_{k \in [K]} k \mathbf{1}\{u_n \in \Delta_k(\theta)\}, \tag{1}$$

and the framework of directed graphs. This connection is explicitly characterized by the authors' Proposition 3.2, where, conditional on a sample of auxiliary points, $u_1, \ldots, u_N$, a closed-form expression of some $\theta \in \Delta$ is constructed that is consistent with the observed data in Equation (1). That is, the θ value constructed via Proposition 3.2 is contained in the set,

$$\mathcal{F}(\mathbf{u}) := \{\theta \in \Delta : \forall n \in [N], u_n \in \Delta_{x_n}(\theta)\},$$

so that the set $\mathcal{R}_x := \{\mathbf{u} : \mathcal{F}(\mathbf{u}) \neq \emptyset\}$ is guaranteed to be nonempty. The elegantly constructed θ in Proposition 3.2 has components proportional to the inverse of the exponentiated directed graph path "value", minimized over ratios of components $\frac{u_{n,\ell}}{u_{n,k}}$ over all $n \in \{1, \ldots, N\}$.

Thus, given a sample of auxiliary points, $u_1, \ldots, u_N$, the authors provide a mechanism to update θ consistent with the data-generating Equation (1) so that it is possible to uniformly sample a new set of auxiliary points via an original algorithm proposed in Dempster (1966) (this algorithm is re-stated as Algorithm 1 in the article under discussion). The contribution is summarized as Algorithm 2 in the article under discussion.

Algorithm 1, restated from Dempster (1966), provided a very simple mechanism for sampling a value u on the sub-simplex $\Delta_k(\theta)$, but did not resolve the issue of updating θ. So it seems that the article under discussion has resolved this critical link to construct a sampling mechanism for uniform sampling of the auxiliary components of the data-generating Equation (1). Certainly this is a remarkable accomplishment.

Second, the authors do a great job in the article of highlighting the role of the structural equation in a fiducial approach to

inference in the categorical model. In particular, the mapping $\mathcal{F} : \mathbf{u} \to \theta$ is the inverse mapping of the data-generating Equation (1). Such a notion/construction characterizes a whole class of contemporary fiducial approaches. Two notable examples are the inferential models (IM) approach (Martin and Liu 2015b) and the generalized fiducial inference (GFI) approach (Hannig et al. 2016). The IM approach is especially relevant because in the IM framework belief and plausibility functions also drive the statistical inference, where questions of interest are formulated as measurable sets in the parameter space. Accordingly, just as described for the DS approach, the IM approach admits a "don't know" probability, denoted as r in the belief, plausibility, and "don't know" triple, (p, q, r), respectively.

The importance of this third, "don't know" category of inference has been drawing attention in recent years, driven in part by the discovery of the *false confidence theorem* introduced in Balch, Martin, and Ferson (2019). The consequence of the false confidence theorem is illustrated in Balch, Martin, and Ferson (2019) in the context of a phenomenon observed in a Bayesian analysis to assess the probability of satellite collision based on real satellite trajectory data. It is observed that a degradation of the data quality leads to a decreased posterior probability of a collision event. Obviously this is not because the analyst can be *more* sure of noncollision, but the phenomenon ensues because the analyst is necessarily *less* sure of a collision event. Under the usual Kolmogorov axioms of probability calculus, probabilities of mutually exclusive but collectively exhaustive events must sum to one. Since the logic for Bayesian inference is tied to probability calculus based on an estimated posterior probability distribution, this problem is foundational for Bayesian theory. In my own work, namely Carmichael and Williams (2018), we provided an illustration of simple examples where/how the false confidence theorem manifests.

An aspect that the article under discussion lacks is a more complete literature review and comparison relating the DS approach to other fiducial approaches, most importantly because the value of the contribution of the article resides in the relevance of the DS approach in the context of the modern literature on fiducial ideas. For instance, Martin (2021) showed that, in general, valid and efficient tests and confidence regions are fundamentally tied to the distributions of random subsets on the parameter space. Moreover, the IM approach, which parallels (at least this formulation of) inference in the DS framework, has exhibited growing success over the past decade, in terms of both methodological developments and computational strategies for implementation (see, e.g., Martin and Liu 2015a, 2015b, 2015c as well as Martin, Leaf, and Liu 2012; Martin 2018; Cella and Martin 2019, and Martin and Syring 2019). Nonetheless, other applications of IM may have nontrivial nonemptiness constraints for the distribution of random sets of the auxiliary variable. These applications, similar to the article being discussed, may require more sophisticated computational strategies. Accordingly, it can be expect that the authors' ideas and developments will have an impact far beyond multinomial applications, to valid and efficient probabilistic inference more generally.

Next, the GFI approach also begins with a data generating equation, say $Y = G(\gamma, U)$, for some deterministic function G, unknown parameter γ, and auxiliary random variable U with some known distribution that is independent of γ. While DS and IM are approaches that infer uncertainty about the unknown parameter γ by constructing random sets that capture the variability in U, the GFI approach is to employ a "switching principle" whereby a distributional estimator of the unknown parameter γ is inherited from the distributional assumption on the auxiliary random variable U. From the perspective of DS ideas, an argument can be made that the switching principle is philosophically problematic, but nonetheless, it has been demonstrated that GFI leads to statistically principled reasoning for inference (Hannig et al. 2016; Liu and Hannig 2016). Moreover, it continues to be demonstrated in the literature that GFI approaches to contemporary data problems lend themselves to viable computational strategies (see, e.g., Hannig, Lai, and Lee 2014; Lai, Hannig, and Lee 2015; Williams and Hannig 2019; Cui and Hannig 2019; Williams, Xie, and Hannig 2019; Li and Hannig 2020; Williams, Ommen, and Hannig 2020; Shi et al. 2021; Koner and Williams 2021).

It is also important to afford reference to the growing literature on confidence distributions (e.g., Xie and Singh 2013; Schweder and Hjort 2016). Successful development of computational strategies in the DS framework are exciting and encouraging, but they should not be evaluated in a vacuum.

3. Central Questions

The article under discussion represents an important contribution to the contemporary literature, but of course, much more work is needed to establish the DS framework as more than a novelty. The central questions for discussion that arise from the article are the following:

1. Is the DS approach worth revisiting?
2. Does the computational strategy developed in the article represent a viable path forward for real data analyses within a DS framework?

Certainly an affirming answer to the second question would provide a positive answer to the first. With this thought in mind, the most apparently limiting aspect of the proposed algorithm for the DS inference exclusively constructed for the categorical data model, is the feature that the auxiliary random variables $U_1, \ldots, U_N$ are elements of the same ambient space as the parameter θ. This is problematic because the authors' entire algorithmic development intrinsically exploits this fact. The exposition of the development of the ideas is a series of geometric arguments that directly relate ratios of components of θ to ratios of the values of components of the $u_1, \ldots, u_N$ via $\eta_{k \to \ell}(u_n)$. It is not obvious at all how these ideas could be extended to apply more generally for a broader class of data generating, or structural equations.

A question of secondary concern is what to do if the parameters in a given data generating equation are not supported on a simplex? While there is only so much responsibility that any one article can be charged with, this is an important question for the computational feasibility of the DS approach, more generally. Moreover, since this is a discussion article with such a central theme, this question stands out as relevant. The authors state in their introduction that over past decades, DS theory has seen "various applications in signal processing, computer vision and

machine learning ….'' It would be helpful if the authors could provide an overview of how DS computations are carried out in the referenced application areas. What makes these other settings more amenable to DS computations than the categorical model?

4. Prevailing Matters

Section 4.2 in the article under discussion, as it reads, is a bit of a hard sell for why, as the authors state in their concluding remarks: "One of the appeals of the DS framework is its flexibility to incorporate types of partial information which are difficult to express in the Bayesian framework." This ties into the broader theme of the article about whether DS is worth considering by mainstream statistical communities. Arguably, the overwhelming majority of practitioners that incorporate prior information in a typical Bayesian inference fashion would find Dempster's rule of combination an inaccessible obfuscation of how to formalize prior information in a statistical inference. Surely, this is an ill-founded reason to dismiss an important idea, but it *does* represent a real obstacle for transforming the DS approach into a class of tools that are readily accessible by a wide audience of practitioners.

I hope that in future work the authors will continue to develop on other equally impressive ideas that increase the computational feasibility of the DS approach. Furthermore, it would be very exciting to see an investigation of the comparison of the DS approach to a modern Bayesian approach, on a real data set of practical significance. A Bayesian analysis would be very problematic if the excluded "don't know" probability has a nontrivial value, as demonstrated for satellite trajectory data analyses in Balch, Martin, and Ferson (2019). In theory, the DS approach will not suffer from the inadequacy exhibited in the Bayesian approach.

ORCID

Jonathan P Williams http://orcid.org/0000-0001-5164-9430

References

Balch, M. S., Martin, R., and Ferson, S. (2019), "Satellite Conjunction Analysis and The False Confidence Theorem," *Proceedings of the Royal Society*, Series A, 475, 20180565. [1199,1200]

Carmichael, I., and Williams, J. (2018), "An Exposition of The False Confidence Theorem," *Statistics*, 7, e201. [1199]

Cella, L., and Martin, R. (2019), "Incorporating Expert Opinion in an Inferential Model While Maintaining Validity," in *International Symposium on Imprecise Probabilities: Theories and Applications*, PMLR, 68–77. [1199]

Cui, Y., and Hannig, J. (2019), "Nonparametric Generalized Fiducial Inference for Survival Functions Under Censoring," *Biometrika*, 106, 501–518. [1199]

Dempster, A. P. (1966), "New Methods for Reasoning Towards Posterior Distributions Based on Sample Data," *The Annals of Mathematical Statistics*, 37, 355–374. [1198]

Hannig, J., Iyer, H., Lai, R. C., and Lee, T. C. (2016), "Generalized Fiducial Inference: A Review and New Results," *Journal of the American Statistical Association*, 111, 1346–1361. [1199]

Hannig, J., Lai, R. C., and Lee, T. C. (2014), "Computational Issues of Generalized Fiducial Inference," *Computational Statistics & Data Analysis*, 71, 849–858. [1199]

Koner, S., and Williams J.P. (2021), "The EAS Approach to Variable Selection for Multivariate Response Data in High-Dimensional Settings," arXiv:2107.04873. [1198,1199]

Lai, R. C., Hannig, J., and Lee, T. C. (2015), "Generalized Fiducial Inference for Ultrahigh-Dimensional Regression," *Journal of the American Statistical Association*, 110, 760–772. [1199]

Li, G., and Hannig, J. (2020), "Deep Fiducial Inference," *Statistics*, 9, e308. [1199]

Liu, Y., and Hannig, J. (2016), "Generalized Fiducial Inference for Binary Logistic Item Response Models," *Psychometrika*, 81, 290–324. [1199]

Martin, R. (2018), "On an Inferential Model Construction Using Generalized Associations," *Journal of Statistical Planning and Inference*, 195, 105–115. [1199]

———— (2021), "An Imprecise-Probabilistic Characterization of Frequentist Statistical Inference," *Researchers One*. Available at: *https://researchers.one/articles/21.01.00002* [1199]

Martin, R., Leaf, D. E., and Liu, C. (2012), "Optimal Inferential Models for a Poisson Mean," arXiv:1207.0105 . [1199]

Martin, R., and Liu, C. (2015), "Conditional Inferential Models: Combining Information for Prior-Free Probabilistic Inference," *Journal of the Royal Statistical Society*, Series B, 77, 195–217. [1199]

———— (2015), *Inferential Models: Reasoning With Uncertainty* (Vol. 145), Boca Raton, FL: CRC Press. [1199]

———— (2015), "Marginal Inferential Models: Prior-Free Probabilistic Inference on Interest Parameters," *Journal of the American Statistical Association* 110, 1621–1631. [1199]

Martin, R., and Syring, N. (2019), "Validity-Preservation Properties of Rules for Combining Inferential Models," in *International Symposium on Imprecise Probabilities: Theories and Applications*, PMLR, pp. 286–294. [1199]

Schweder, T., and Hjort, N. L. (2016), *Confidence, Likelihood, Probability* (Vol. 41), Cambridge: Cambridge University Press. [1199]

Shi, W. J., Hannig, J., Lai, R. C., and Lee, T. C. (2021), "Covariance Estimation Via Fiducial Inference," *Statistical Theory and Related Fields*, 1–16. [1199]

Williams, J. P., and Hannig, J. (2019), "Nonpenalized Variable Selection in High-Dimensional Linear Model Settings Via Generalized Fiducial Inference," *The Annals of Statistics*, 47, 1723–1753. [1198,1199]

Williams, J. P., Ommen, D. M., and Hannig, J. (2020), "Generalized Fiducial Factor: An Alternative to The Bayes Factor for Forensic Identification of Source Problems," arXiv: 2012.05936 . [1199]

Williams, J. P., Xie, Y., and Hannig, J. (2019), "The Eas Approach for Graphical Selection Consistency in Vector Autoregression Models," arXiv: 1906.04812. [1198,1199]

Xie, M.-g., and Singh, K. (2013), "Confidence Distribution, The Frequentist Distribution Estimator of a Parameter: A Review," *International Statistical Review*, 81, 3–39. [1199]