DOI: 10.1002/sta4.571

DISCUSSION

WILEY

# A Discussion of "A Note on Universal Inference" by Tse and Davison

Minge Xie

Department of Statistics, Rutgers University, Piscataway, New Jersey, 08854, USA

Correspondence
Minge Xie, Department of Statistics, Rutgers University, Piscataway, NJ 08854, USA.
Email: mxie@stat.rutgers.edu

Congratulations to Timmy and Anthony for the in-depth note that provides a great insight for understanding of the performance of the recently developed universal inference method by Wasserman et al. (2020). Based on my own reading and research experiences on irregular inference approaches including universal inference, I echo the messages that Timmy and Anthony have presented so clearly, some of which I tried to con-vert to my students but could not do as nicely and convincingly. Beside the messages by Timmy and Anthony, I would like to comment on that the universal inference method has a unique advantage for sequential testing problems. I will also provide some additional remarks on tackling irregular inference problems.

The universal inference framework is a novel development that can provide a performance-guaranteed finite-sample inference without (amazingly) any regularity conditions (Wasserman et al., 2020). The method typically requires a tractable likelihood function and data splitting, and the validity (guaranteed test size) of the method is justified by an Markov inequality. However, as demonstrated by Timmy and Anthony in their eqs. (5)–(7) of the standard mean test using the simple normal example, the method systemically suffers loss of power due to the use of the Markov inequality and sample splitting. In addition, Timmy and Anthony used a multivariate normal mean example (with a known variance matrix) to show the impact of the number of nuisance parameters on power and provided an a simple correction to mitigate the problem. The discussion is extended to some other cases but relies on asymptotics. Finally, the impact on the testing power for different numbers of folders of data split-ting is also studied, again using the Gaussian example. These discussions on testing power provided a full picture of the universal inference method, in complementary to its validity result under the regular inference settings.

The universal inference is particularly effective for sequential testing problems. For instance, consider a setting described in Ramdas et al. (2020): we have a sequence of observations $y_t, t = 1,2,\ldots,$ and we would like to do a sequential test and make a simultaneous inference for $\{\mu_t, t = 1,2,\ldots\}$, where $\mu_t = E(\overline{Y}_t)$ and $\overline{y}_t = t^{-1}\sum_{s=1}^{t} y_s$. Ramdas et al. (2020) outlined the issues encountered in the standard approaches and showed that the universal inference method can derive a stronger result of anytime-valid simultaneous inference. The universal inference method for sequential tests also has close connections to the probabilistic game-theoretic approach of sequential optimal betting (Ramdas et al., 2022). These developments have stimulated an emerging development of the so-called e-value concept and its sequential analogue e-processes, which provide a useful alternative to the traditional p value and other notions for anytime-valid inference (Turner et al., 2021; Vovk & Wang, 2021; Wang & Ramdas, 2020).

Another attraction of the universal inference method is that it is readily applicable for many irregular inference problems where inference solutions are either unavailable or difficult to obtain by the traditional methods. Here, following Wasserman et al. (2020), the irregular inference problems refer to those problems for which the regular statistical inference conditions and large-sample theories do not apply. The universal infer-ence development is invaluable for irregular inference problems (even though it may loss some power), since these problems are prevalent in data science. For instance, inherently discrete parameters are routinely seen, for example, model indices in model selection problems, number of clus-ters and membership in tree partition and other classifications/clustering problems, and architecture parameters (numbers of layers and vertices) in a deep neural network. For these irregular problems, point estimators (if they exist) do not provide uncertainty quantification; bootstrap approaches lack theoretical support and often underperform; Bayesian credible sets tend to be highly sensitive to the (additional) prior choices and their frequency coverage is often poor even with a large-sample size. These difficulties arise because the central limit theorem and

Bernstein–von Mises theorem do not apply on an intrinsically discrete parameter space. The universal inference is perhaps the first general framework that can possibly provide validity guarantees for such irregular inference problems.

Speaking of irregular inference problems, our group is also developing a new artificial-sample-based approach, called a repro samples method, to tackle these problems (cf. Xie & Wang, 2022). Our method, developed under the frequentist framework, borrows the ideas of "inversion" and "matching of artificial and observed samples" that are deeply rooted in fiducial inference and approximate Bayesian computing (ABC) method. It provides guaranteed inference without the need to rely on a likelihood function or the large-sample central limit theorem. It is especially effective for difficult inference problems such as those involving discrete/non-numerical parameters and/or non-numerical data. The repro samples method has been used to provide a finite-sample two-sided confidence set for the (smallest possible) unknown number of components in a Gaussian mixture (Xie & Wang, 2022), noting that the universal inference method of Wasserman et al. (2020) does not give an upper bound for the unknown number of component. The repro samples method has also been used to provide performance-guaranteed (finite- and large-sample) confidence sets for the unknown true model (and other model parameters) in a high-dimensional regression setting (Wang et al., 2022). Compared with the universal inference method, the repro samples method does not systemically suffer a loss of power. However, it has the same issues encountered in the standard approaches and does not typically provide an anytime-valid simultaneous inference for sequential testing problems.

## REFERENCES

Ramdas, A., Grünwald, P., Vovk, V., & Shafer, G. (2022). Game-theoretic statistics and safe anytime-valid inference. arXiv preprint arXiv:2210.01948.

Ramdas, A., Ruf, J., Larsson, M., & Koolen, W. (2020). Admissible anytime-valid sequential inference must rely on nonnegative martingales. arXiv preprint arXiv:2009.03167.

Turner, R., Ly, A., & Grünwald, P. (2021). Safe tests and always-valid confidence intervals for contingency tables and beyond. arXiv preprint arXiv:2106.02693.

Vovk, V., & Wang, R. (2021). E-values: Calibration, combination and applications. The Annals of Statistics, 49(3), 1736–1754.

Wang, P., Xie, M., & Zhang, L. (2022). Finite and large sample inference for model and coefficients in high dimensional linear regression with repro samples. arXiv e-prints, arXiv.2209.4500515.

Wang, R., & Ramdas, A. (2020). False discovery rate control with e-values. arXiv preprint arXiv:2009.02824.

Wasserman, L., Ramdas, A., & Balakrishnan, S. (2020). Universal inference. Proceedings of the National Academy of Sciences, 117(29), 16880–16890.

Xie, M., & Wang, P. (2022). Repro samples method for finite- and large-sample inferences. arXiv e-prints, arXiv.2206.06421.

How to cite this article: Xie, M. (2023). A Discussion of "A Note on Universal Inference" by Tse and Davison. Stat, 12(1), e4571. https://doi.org/10.1002/sta4.571