Statistica Sinica Preprint No: SS-2019-0171	
Title	Rejoinder for "Entropy Learning for Dynamic Treatment Regimes"
Manuscript ID	SS-2019-0171
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202019.0171
Complete List of Authors	Binyan Jiang Rui Song Jialiang Li and Donglin Zeng
Corresponding Author	Jialiang Li
E-mail	stalj@nus.edu.sg

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0171

Statistica Sinica

REJOINDER

Binyan Jiang[1], Rui Song[2], Jialiang Li[3] and Donglin Zeng[4]

The Hong Kong Polytechnic University[1], North Carolina State University[2]

National University of Singapore[3] and University of North Carolina[4]

We thank Statistica Sinica for providing the venue for this paper and its discussion, and all discussants for their many contributions, insights, and thought-provoking questions. The area of dynamic treatment regimes is developing rapidly, and we hope that our paper and the subsequent discussion will add further momentum to this exciting field. In this rejoinder, we focus on the following four topics: (1) the nonregularity issue, when neither treatment is more beneficial for a nontrivial subgroup (comments by Lu; Qian and Cheng; Qiu et al.); (2) the linear decision boundary (comments by He, Xu, and Wang; Lu; Qiu et al.); (3) extensions that incorporate smooth weights, multiple classes, or a nonconvex loss (comments by Wager; Kallus; Lu; Qian and Cheng; He et al.; Qiu et al.; Zhang and Laber); (4) interpreting the p-value in a real application (comment by Wager).

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0171

1. Nonregularity

The nonregularity issue $P(X_t^{*T}\beta_t^0 = 0) > 0$ is a long-standing and challenging inference problem in estimations of dynamic treatment regimes. Our assumption A3 rules out this situation; in particular, we allow a relatively weak condition on the distribution decay near this boundary. Recent attempts to address this issue include finding a probability upper bound, regardless of this nonregularity (Laber et al., 2014), the m-out-of-n bootstrap method (Chakraborty et al., 2013), data-adaptive hard-thresholding (Zhu, Zeng, and Song, 2018), penalized Q-learning (Song et al. , 2014), and adaptive Q-learning (Goldberg et al. , 2012). However, inferences may be either conservative or unreliable in the case of small sample sizes. Thus, there remains much scope for research on improving inferences with nonregularity.

Although such inferences are theoretically interesting, the impact of nonregularity on practical evaluations of optimal treatment regimes may not be that significant. Essentially, the treatments work very similarly near the boundary. Even if some patients near the decision boundary are allocated to less beneficial treatments, owing to an incorrect inference, the changes to the estimated value function and its inference are practically negligible. This is observed in our numerical studies that demonstrate the

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0171

robustness of our methods. On the other hand, as suggested by Qiu et al., a more realistic consideration is to test whether the treatment effect exceeds a certain level (i.e., $X_t^{*T}\beta_t^0 \leq \gamma$, for some $\gamma > 0$). Theoretically, we can always choose some γ close to a clinically meaningful threshold such that $P(X_t^{*T}\gamma = 0) = 0$ to void the nonregularity issue.

2. Linear decision boundary

Some discussants suggested there may be restrictions on the applicability of the linear form of the treatment decision. Specifically, He et al. suggested nonparametric treatment rules for entropy learning under the RKHS framework, and Qiu et al. obtained nonparametric decision rules using the highly adaptive LASSO approach. Many extensions to our rule are possible, following these suggestions. For example, a simple extension to our linear rule is to incorporate quadratic terms in our estimation to capture possible interactions between the feature covariates. Such ideas emerged recently in the discrimination and regression analysis literature (Jiang et al., 2018; Wang et al., 2019), and have enjoyed consistency for interaction detection. Furthermore, we may consider smoothing splines to obtain fully nonparametric rules, although the current inference results need to be adapted to reflect the nature of a sieve estimation.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0171

We argue that linear decision rules themselves are still of considerable value in practice, owing to their simplicity and better interpretability. Several discussants noted that the computational demand could become prohibitively heavy when big data such as electronic transaction records or medical images are present. In this case, the simple form of linear rules coupled with a convex objective function, such as the entropy learning loss in our work, becomes most appealing (Shi et al., 2018). Finally, partly because of the dichotomous nature of the treatment rule, applying linear rules to derive the value function may not be disadvantageous compared with using rules that are more complex. However, further empirical and theoretical investigation is necessary.

3. Extensions to incorporate smooth weights, multiclass, or nonconvex loss

While many discussants provided helpful suggestions, in this section, we provide brief replies to selected issues; certainly, many deserve a much longer explanation.

Kallus suggested replacing the indicator functions in the estimation equations (e.g., equation (2.8)) with optimal balancing weights to avoid omitting too many samples when T is large. The balanced approach is in-

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0171

teresting, and can produce better estimation results than those of outcome-weighted approaches. Here, recent research has led to a greater understanding of the theoretical properties of covariate balancing in causal inferences (Zhao, 2019). However, because the weights are data-driven, it is often difficult to conduct inferences, and the computational complexity might be high for particularly big data. Nevertheless, we agree that it would be meaningful to replace the indicator functions in some early stages with optimal balancing weights. This will enable proper inferences in the later stages, and alleviate the issue of omitting too many samples during the backward estimation procedure. On the other hand, with appropriate smoothness assumptions, it is also possible to obtain valid inferences, with extra effort required to take care of the kernel approximation bias.

Dr. Lu inquired whether E-learning is adaptable to treatments with multiple categories at each stage. Our answer is yes. Note that for the two-class case, the minimizer of (2.4) is $\log \frac{E[R|A=1,\mathbf{X}=\mathbf{x}]}{E[R|A=-1,\mathbf{X}=\mathbf{x}]}$, which attains a form similar to that of an odds ratio. Mimicking this form, we may adopt a simple approach to, for example, set the first treatment option as the baseline, and then estimate the pairwise contrast for the other option versus the first option. This operation is similar to the extension of the classical binary logistic regression model to the multiclass logistic regression model.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0171

In addition to E-learning, proposed in this work, many learning approaches for individual treatment selection have been established under various objective (see the introduction for further examples). Subsequent to this work being accepted for publication, we were informed that C-learning (Zhang and Zhang , 2018; Hager et al., 2018), augmented O-learning (Liu et al., 2018), concordance assisted learning (Fan et al., 2017; Liang et al., 2018), maximin projection learning (Shi et al., 2018), and quantile optimal treatment regimes (Wang et al. , 2018) had since been proposed, among many others. In this discussion, discussants continued to suggest further modifications. Qian and Cheng provided theoretical results for the excess risk and excess value of entropy learning, based on the construction in Bartlett et al. (2006). Qiu et al. studied the behavior of entropy learning under model misspecification, proposing a framework for nonparametric decision rules. Zhang and Laber developed a direct search approach, in which they replace the 0-1 loss with a nonconvex surrogate, to estimate an authentic linear rule that ensures value optimization.

4. Interpretation of p-values

Dr. Wager raised a concern on how to interpret the p-values from the regression tables. We agree that when more than one linear rule leads to

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0171

the same optimal value, as demonstrated in his numerical example, using a p-value to conclude an important feature for a treatment decision could be misleading.

However, information contained in p-values usually cannot be recovered by other measures. As such, we may not want to completely retire them, for the following detailed reasons:

(a) For an estimated linear rule, such as that in our application, p-values can be used to assess statistical evidence on whether a feature contributes to a rule. However, identifying an important feature does not necessarily imply its utility in the treatment decision for value improvement. This significance is useful in practice when examining the uncertainty of a rule in a finite sample.

(b) The p-values given in the tables provide a computationally simple way to assess the importance of features in the estimated optimal treatment rule. Thus, it is potentially useful for screening out noisy features in the high-dimensional data settings (for example, Zhu, Zeng, and Song (2018)). In contrast, using value-based methods to select important features may be computationally intensive or unstable, especially when more than one rule yields the same optimal value.

(c) The p-values given in the tables are associated with the particular sur-

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0171

rogate loss (entropy loss) we used. In this sense, each inference used to test a feature's contribution is unique and reliable, in practice. However, value-based inferences are infeasible owing to a lack of uniqueness.

Finally, we believe that the best way to assess the importance of features is a combination of our approach and a value-based method. The former yields an unambiguous treatment rule and associated inference, which is useful in practice. The latter ensures that the selected features truly lead to clinically meaningful benefits.

References

Bartlett, P., Jordan, M. & McAuliffe, J. (2006). Convexity, classification, and risk bounds. *J. Am. Statist. Assoc.*, **101**, 138–156.

Chakraborty, B., Laber, E. & Zhao, Y. (2013). Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme. *Biometrics* **69**, 714–723.

Hager, R., Tsiatis, A. & Davidian, M. (2018). Optimal Two-Stage Dynamic Treatment Regimes from a Classification Perspective with Censored Survival Data. *Biometrics* **74**, 11801192.

Jiang, B., Wang, X., & Leng, C. (2018). A direct approach for sparse quadratic discriminant analysis. *J. Mach. Learn. Res.*, **19**, 1098–1134.

Song, R., Wang, W., Zeng, D. & Kosorok, MR. (2014). Penalized Q-learning for Dynamic Treatment Regimes. *Statistica Sinica*, **25**(3), 901–920.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0171

REFERENCES

Goldberg, Y., Song, R. & Kosorok, MR (2012). Adaptive Q-learning. *IMS Collections: From Probability to Statistics and Back: High-Dimensional Models and Processes*, **9**, 150–162,

Laber, E., Lizotte, D., Qian, M., Pelham, W., & Murphy, S. (2014). Dynamic treatment regimes: Technical challenges and applications. *Electron. J. Stat.*, **8**, 1225–1272.

Liu, Y., Wang, Y., Kosorok, M., Zhao, Y., Zeng, D. (2018). Augmented outcome-weighted learning for estimating optimal dynamic treatment regimens. *Statistics in Medicine*, **37**, 3776-3788.

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. *Proceedings of The Second Seattle Symposium in Biostatistics*, pp. 189-326. Springer, New York, NY.

Shi, C., Fan, A., Song, R. & Lu, W. (2018). High-dimensional A-learning for optimal dynamic treatment regimes. *Ann. Stat.*, **46**, 925–957.

Wang, C., Jiang, B., & Zhu, L. (2019). Penalized interaction estimation for ultrahigh dimensional quadratic regression. *arXiv preprint arXiv:1901.07147*.

Wang, L., Zhou, Y., Song, R. & Sherwood, B. (2018). Quantile-Optimal Treatment Regimes. *J. Am. Statist. Assoc.*, **113**, 1243–1254.

Zhang, B and Zhang, M (2018). C-Learning: A New Classification Framework to Estimate Optimal Dynamic Treatment Regimes. *Biometrics*, **74**, 891–899.

Fan, C., Lu, W., Song, R. & Zhou, Y. (2017). Concordance-Assisted Learning for Estimating

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0171

REFERENCES

Optimal Individualized Treatment Regimes. *J R Stat Soc Series B*, **79**(5), 1565–1582.

Liang, S., Lu, W., Song, R. & Wang, L. (2018). Sparse Concordance-assisted Learning for Optimal Treatment Decision. *J. Mach. Learn. Res.*, **18** (202): 1-26.

Shi, C., Song, R., Lu, W., & Fu, B. (2018). Maximin Projection Learning for Optimal Treatment Decision with Heterogeneous Individualized Treatment Effects. *J R Stat Soc Series B*, **80** (4), 681–702.

Zhao, Q (2019). Covariate balancing propensity score by tailored loss functions. *Ann. Stat.*, **47**, 965-993.

Zhu, W., Zeng, D., & Song, R. (2018). Proper inference for value function in high-dimensional Q-Learning for dynamic treatment regimes. *J. Am. Statist. Assoc.*, 1–14.

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China.

E-mail: by.jiang@polyu.edu.hk

Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA.

E-mail: rsong@ncsu.edu

Department of Statistics and Applied Probability, National University of Singapore, 117546, Singapore.

E-mail: stalj@nus.edu.sg

Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, USA.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0171

REFERENCES

E-mail: dzeng@email.unc.edu