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
Notice: Accepted version subject to English editing.	

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 rapidly developing and hopefully 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 non-regularity issue when neither treatment is more beneficial for a non-trivial 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) the extensions to incorporate smooth weights, multiclass, or non-convex loss (comments by Wager; Kallus; Lu; Qian and Cheng; He et al; Qiu et al.; Zhang and Laber); (4) the interpretation of p-value in the real application (comment by Wager).

1. Non-regularity

The non-regularity issue $P(X_t^{*T}\beta_t^0 = 0) > 0$ has been known as a long-standing and challenging inference problem in estimating dynamic treatment regimes. Our assumption A3 rules out this situation; in particular, we allow a relatively weak condition regarding the distribution decay near this boundary. Recent attempts to address this issue include finding a probability upper bound regardless of this non-regularity (Laber et al., 2014), 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, inference can be either conservative or unreliable with small example sizes. There remains a large room to improve the inference with non-regularity.

Although the inference with non-regularity is theoretically interesting, the impact on practical evaluation of the optimal treatment regimens may not be that significant. Essentially the treatments work very similarly near the boundary. Even if some patients near the decision boundary were allocated to less beneficial treatments due to incorrect inference, the change to the estimated value function and its inference is practically negligible. This has been observed in our numerical studies which demonstrated the robustness of our methods. On the other hand, as suggested by Qiu et al., a more

realistic consideration is to test whether 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 clinical meaningful threshold such that $P(X_t^{*T}\gamma = 0) = 0$ to void the non-regularity issue.

2. Linear decision boundary

Some discussants suggested that there were restrictions on the applicability of the linear form of treatment decision. Specifically, He et al. suggested a nonparametric treatment rules for entropy learning under the RKHS framework; Qiu et al. also obtained nonparametric decision rules via the Highly Adaptive LASSO approach. There are so many alternative extensions along these suggestions. For example, a simple extension to our linear rule is to incorporate quadratic terms in our estimation so as to capture possible interactions among the feature covariates. Such ideas emerged recently in the discrimination analysis literature (Jiang et al., 2018; Wang et al., 2019) and enjoyed consistency for misclassification rate. Furthermore, we may consider smoothing splines to obtain fully nonparametric rules, although the current results for inference needs to be adapted to reflect the nature of sieve estimation.

We argue that linear decision rules themselves are still of considerable

value in practice, owing to their simplicity and better interpretability. Acknowledged by several discussants that computational demand could be prohibitively heavy when big data such as electronic transaction records or medical images are present, 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 in a huge disadvantage as compared to more complex rules. Further empirical and theoretical investigation is necessary.

3. Extensions to incorporate smooth weights, multiclass, or non-convex loss

All discussants have provided other miscellaneous suggestions that help us see this work from heretofore unappreciated angles. In this section we provide brief replies to some of the issues; certainly many deserve much longer explanation.

Kallus suggested to replace the indicator functions in the estimation equations (e.g., equation (2.8)) by the optimal balancing weights, so as to avoid omitting too many samples when T is large. The balanced approach is interesting and can produce better estimation results than outcome-

weighted approaches. In fact more theoretical properties were understood about covariate balancing in causal inference lately (Zhao , 2019). However, since the weights are data-driven, it is in general hard to conduct inferences, and the computational complexity might be high when facing really big data. Nevertheless, we agree that it would be meaningful to replace the indicator functions in some early stages by the optimal balancing weights, so as to enable proper inferences in the later stages and alleviate the issue of omitting too many samples during the backward estimation procedure. With appropriate smoothness assumptions, we shall be able to obtain valid inference with slightly extra effort 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. We 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 an odds ratio. Mimicking this form, we may adopt a simple approach to set for example, the 1st treatment option, as the baseline and estimate the pairwise contrast for the other option vs. the 1st option. This operation is similar to the extension of classical binary logistic regression model to the multi-level logistic regression model.

Besides E-learning proposed in this work, many versions of learning

approaches for individual treatment selection were established under different objectives. See introduction for earlier examples. By the time this work was accepted for publication we were further 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 appeared, among many others. In this discussion, discussants continued proposing further modification. Qian and Cheng provided theoretical results for the excess risk and the excess value for entropy learning based on the construction in Bartlett et al. (2006); Qiu et al. studied the behavior of entropy learning under model misspecification and further proposed a framework for nonparametric decision rules; Zhang and Laber developed a direct search approach, which replaces the 0-1 loss by a non-convex 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 p-values outputted in regression tables. We agree that when more than one linear rule leads to the same optimal value as demonstrated in his numerical example, using

p-value to conclude an important feature for treatment decision could be misleading.

However, information contained in p-values usually cannot be recovered by other measures and consequently we may not want to completely retire them. The following are some detailed arguments:

(a) For an estimated linear rule such as the one in our application, p-values can still be used to assess the statistical evidence regarding whether the contribution of a feature to this particular rule is important, although such an importance may not necessarily imply its utility in the treatment decision for value improvement. Such significance is practically useful if one wonders about the uncertainty of the rule itself in a finite sample.

(b) 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 high-dimensional data setting (for example, Zhu, Zeng and Song (2018)). By contrast, using value to select important features may be computationally intensive or unstable, especially when there are more than one rules yielding the same optimal value.

(c) P-values given in the tables are associated with the particular surrogate loss (entropy loss) we used. In this sense, the inference for testing each

feature's contribution is unique and reliable for practice. However, value-based inference is infeasible due to lack of uniqueness.

Finally, we believe that the best way for assessing feature importance is a combination of our approach and value-based method: the former yields an unambiguous treatment rule and its associated inference, which is useful for practice; while the latter ensures the importance of selected features can 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. *Statistical Sinica*, **25**(3), 901–920.

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

REFERENCES

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.

E-mail: dzeng@email.unc.edu