Statistica Sinica Preprint No: SS-2019-0035	
Title	Discussion on `Entropy Learning for Dynamic Treatment Regimes'
Manuscript ID	SS-2019-0035
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202019.0035
Complete List of Authors	Xin He Shirong Xu and Junhui Wang
Corresponding Author	Junhui Wang
E-mail	j.h.wang@cityu.edu.hk

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0035

Discussion on "Entropy Learning for Dynamic Treatment Regimes"

Xin He†, Shirong Xu‡ and Junhui Wang‡

† *Shanghai University of Finance and Economics*

‡ *City University of Hong Kong*

We would like to congratulate Drs. Jiang, Song, Li, and Zeng (JSLZ) for their well-written and thought-provoking work, which bridges machine learning and statistical inferences when estimating optimal individualized treatment rules (ITRs), and opens numerous avenues for future research on related topics. Below we discuss the paper from two aspects: its extension to kernel-based nonparametric ITRs, and inferences for nonparametric ITRs.

1. Kernel-based nonparametric ITRs

JSLZ assume that the decision functions $f_t(\mathbf{x})$, for $t = 1, ..., T$, have a linear form, which facilitates the model fitting and statistical inferences. However, in the machine learning community, much research is being conducted on nonparametric decision functions, for example, in a reproducing

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0035

kernel Hilbert space (RKHS; see [2, 3, 4, 6, 7, 8], and the references therein). An RKHS provides a flexible framework for modeling nonparametric functions without explicitly enumerating the functional basis. It can be fully induced by any symmetric and nonnegative definite kernel function, where the choice of kernel functions relies on the available prior information about f_t. In practice, if no prior information is available, it is a common practice to use the Gaussian kernel, which is known to be universal in the sense that any continuous function can be well approximated by the induced RKHS under the infinity norm [5].

To extend JSLZ to estimate kernel-based nonparametric ITRs, we consider the case of $f_t \in \mathcal{H}_K$, an RKHS induced by some kernel function $K(\cdot,\cdot)$. The formulation of the kernel-based nonparametric ITRs then becomes

$$\min_{f_t \in \mathcal{H}_K} -\frac{1}{n}\Big(\boldsymbol{\omega}_t \odot \big(0.5(\mathbf{A}_t + \mathbf{1}_n) \odot \mathbf{f}_t + \ln \boldsymbol{\xi}_t\big)\Big)^T \mathbf{1}_n + \lambda_n \|f_t\|_K^2, \tag{1.1}$$

where $\boldsymbol{\omega}_t = (\omega_{t1}, ..., \omega_{tn})^T$, $\mathbf{A}_t = (A_{t1}, ..., A_{tn})^T$, $\mathbf{f}_t = (f_t(\mathbf{x}_1^t), ..., f_t(\mathbf{x}_n^t))^T$, $\odot$ denotes a componentwise product, $\boldsymbol{\xi}_t = (\xi_{t1}, ..., \xi_{tn})^T$ with $\xi_{ti} = \big(1 + \exp(f_t(\mathbf{x}_i^t))\big)^{-1}$, and $\|f_t\|_K^2 = \langle f_t, f_t \rangle_K$ is the associated RKHS-norm of f_t.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0035

By the representer theorem [2], the minimizer of (1.1) must have the form

$$f_t(\mathbf{x}^t) = \sum_{i=1}^{n} \alpha_{ti} K(\mathbf{x}_i^t, \mathbf{x}^t) = \boldsymbol{\alpha}_t^T \mathbf{K}_n(\mathbf{x}^t),$$

where $\boldsymbol{\alpha}_t = (\alpha_{t1}, ..., \alpha_{tn})^T$ and $\mathbf{K}_n(\mathbf{x}^t) = (K(\mathbf{x}_1^t, \mathbf{x}^t), ..., K(\mathbf{x}_n^t, \mathbf{x}^t))^T$. Moreover, let $K = \big((K(\mathbf{x}_i, \mathbf{x}_j))\big)_{i,j=1}^n$. Then $\mathbf{f}_t = \mathbf{K}\,\boldsymbol{\alpha}_t$ and $\|f_t\|_K^2 = \boldsymbol{\alpha}_t^T\,\mathbf{K}\,\boldsymbol{\alpha}_t$. After substituting these into (1.1), the optimization task with respect to the infinite-dimensional f_t simplifies to an equivalent optimization task with respect to the n-dimensional $\boldsymbol{\alpha}_t$, which can be solved by a slightly modified algorithm, as in JSLZ. It is evident that the kernel-based formulation in (1.1) is fairly similar to the original linear model of JSLZ, while admitting flexible model structures of f_t, thus allowing for general covariate effects on the ITRs.

We now examine the numerical performance of the kernel-based nonparametric ITRs using the simulated example in [3], where $R = Q(\mathbf{x}) + T(\mathbf{x}, A) + \epsilon$, with $T(\mathbf{x}, A) = 3.8(0.8 - x_1^2 - x_2^2)A, Q(\mathbf{x}) = 1 + x_1 + x_2 + 2x_3 + 0.5x_4$, and $\epsilon \sim N(0, 1)$. We consider the Gaussian kernel, set the training sample size as 400 and the validation sample size as 200,000, and set the ridge parameter $\lambda_n = 0.001$. The experiment is repeated 100 times, and the averaged value function values are summarized in Table 1.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0035

Table 1: Comparison of value function of linear and nonparametric ITRs with their standard errors in parenthesis.

Method	Linear ITR	Nonparametric ITR
Value function	1.568(0.010)	1.629(0.008)

Clearly, the nonparametric ITR outperforms its linear counterpart in the simulated example with nonlinear decision boundaries. In practice, as pointed out in [3], the selection of the kernel function can be regarded as a tuning parameter selection problem, with the optimal function being determined using some data-adaptive selection criterion.

2. Inference for nonparametric ITRs

Few studies examine inferences related to machine-learning-based methods, partly because of their "parameter-free" frameworks. A similar concern is raised in JSLZ, although their inference results are still developed for linear ITRs. In fact, recent attempts have been made to develop inference tools for kernel-based approaches. For example, [1] provides an inference for the prediction error of a kernel-based support vector machine, and [8] conducts an inference for kernel-based approaches to estimate the dynamic treatment regimes. The key idea is to utilize resampling techniques to draw inferences on a criterion about the prediction errors of the machine-learning-

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0035

based methods. Similar treatments can be extended to inferences for the predicated value function of the nonparametric ITRs.

Given the training sample $\mathcal{O}_n = \{(R_{1i}, A_{1i}, \mathbf{x}_i^1, ..., R_{Ti}, A_{Ti}, \mathbf{x}_i^T)\}_{i=1}^n$, the estimated optimal decision functions $\widehat{\mathbf{f}} = (\widehat{f}_1, ..., \widehat{f}_T)^T$ can be obtained as in Section 1. Then, for a new observation $\mathcal{O}_0 = (R_{10}, A_{10}, \mathbf{x}_0^1, ..., R_{T0}, A_{T0}, \mathbf{x}_0^T)$, we consider the predicted value function $V_t^0(\widehat{\mathbf{f}}_t)$, with $\widehat{\mathbf{f}}_t = (\widehat{f}_t, ..., \widehat{f}_T)$, and its estimate $\widehat{V}_t(\widehat{\mathbf{f}}_t, \mathcal{O}_n)$. Here, $V_t^0(\widehat{\mathbf{f}}_t)$ and $\widehat{V}_t(\widehat{\mathbf{f}}_t, \mathcal{O}_n)$ are defined as in JSLZ. We then randomly split $\mathcal{O}_n$ into K disjoint subsets $O_n^{(1)}, ..., O_n^{(K)}$ of equal size. For each k, we use all observations not in $O_n^{(k)}$ to obtain $\widehat{\mathbf{f}}^{(-k)} = (\widehat{f}_1^{(-k)}, ..., \widehat{f}_T^{(-k)})^T$, as in Section 1, and use $O_n^{(k)}$ to compute the cross-validated value function. The procedure is repeated for $k = 1, ..., K$, and the final cross-validated value function is

$$\widehat{\mathcal{V}}_{t,n}^{CV} = \frac{1}{K} \sum_{k=1}^{K} \widehat{V}_t(\widehat{\mathbf{f}}^{(-k)}, \mathcal{O}_n^{(k)}).$$

As shown in [1], the asymptotic distribution of $\sqrt{n}\big(\widehat{\mathcal{V}}_{t,n}^{CV} - V_t^0(\widehat{\mathbf{f}}_t)\big)$ is the same as that of $\sqrt{n}\big(\widehat{V}_t(\widehat{\mathbf{f}}_t, \mathcal{O}_n) - V_t^0(\widehat{\mathbf{f}}_t)\big)$.

To approximate the distribution of $\sqrt{n}\big(\widehat{\mathcal{V}}_{t,n}^{CV} - V_t^0(\widehat{\mathbf{f}}_t)\big)$, we consider a

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0035

perturbed version of (1.1), such that

$$\widetilde{f}_t = \underset{f_t \in \mathcal{H}_K}{\operatorname{argmin}} -\frac{1}{n}\mathbf{G}_t \odot \Big(\boldsymbol{\omega}_t \odot \big(0.5(\mathbf{A}_t + \mathbf{1}_n) \odot \mathbf{f}_t + \ln \boldsymbol{\xi}_t\big)\Big)^T \mathbf{1}_n + \lambda_n \|f_t\|_K^2, \tag{2.1}$$

where $\mathbf{G}_t = (G_{t1}, ..., G_{tn})^T$ is drawn from an exponential distribution with unit mean and variance. By sequentially solving (2.1), we obtain $\widetilde{\mathbf{f}} = (\widetilde{f}_1, ..., \widetilde{f}_T)^T$. Specifically, for stage t, we calculate

$$\widetilde{W}_t = n^{-1/2} \sum_{i=1}^{n} \big(\widehat{V}_t(\widetilde{\mathbf{f}}_t, o_i) - \widehat{V}_t(\widehat{\mathbf{f}}_t, \mathcal{O}_n)\big) G_{ti}, \tag{2.2}$$

where o_i denotes the ith sample of $\mathcal{O}_n$. Note that, given $\mathcal{O}_n$, the only random variable in (2.1) is G_{ti}. More importantly, the computed $\widetilde{W}_t$ in (2.2) can be regarded as a realization of a random variable whose distribution can approximate the distribution distribution of $\sqrt{n}\big(\widehat{\mathcal{V}}_{t,n}^{CV} - V_t^0(\widehat{\mathbf{f}}_t)\big)$ very well, given $\mathcal{O}_n$. Thus in practice, we generate $\{G_{ti}\}_{i=1}^n$ repeatedly M times, and obtain a large number of realizations $\widetilde{\mathbf{W}}_t = \{\widetilde{W}_{tm}\}_{m=1}^M$ to approximate the distribution of $\sqrt{n}\big(\widehat{\mathcal{V}}_{t,n}^{CV} - V_t^0(\widehat{\mathbf{f}}_t)\big)$. Therefore, the confidence interval for the prediction value function in stage t can be obtained based on the empirical distribution of $\widetilde{\mathbf{W}}_t$.

We now construct an approximate confidence inference for the pre-

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0035

dicted value function of the nonparametric ITRs in the simulated example in Section 1, with sample size 500 and five-fold cross-validation. We first calculate the true prediction value function by repeatedly generating $\mathcal{O}_n$ independently 1000 times. Then, for each $\mathcal{O}_n$, we calculate the cross-validated value function $\widehat{\mathcal{V}}_n^{CV}$. Therefore, the true value function can be computed as the average of $\widehat{\mathcal{V}}_n^{CV}$. To obtain the interval estimators, we generate $\mathcal{O}_n$ independently 100 times. For each $\mathcal{O}_n$, we compute $\widehat{\mathcal{V}}_n^{CV}$, generate $\mathbf{G}$ repeatedly to obtain 250 realizations of $\widetilde{W}$, and compute the estimated 95% confidence interval. This leads to a 94% coverage rate, which is comparable to the reported coverage rates in JSLZ for parametric ITRs, and may be improved upon with further computational efforts.

3. Concluding remarks

We appreciate the opportunity to contribute to the discussion on this excellent paper. JSLZ provide proper statistical inferences for machine-learning-based methods when estimating ITRs, and leave numerous open questions for further research. For example, it is of great interest to investigate the statistical inferences for the kernel-based nonparametric ITRs theoretically, which enjoy model flexibility and can be adjusted based on prior information. We would like to congratulate JSLZ again on their enlightening work,

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0035

and look forward to seeing similar future research.

References

[1] B. Jiang, X. Zhang, and T. Cai. Estimating the confidence interval for prediction errors of support vector machine classifiers. *Journal of Machine Learning Research*, **9**:521–540, 2008.

[2] G. Kimeldorf and G. Wahba. Some results on tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications*, **33**:82–95, 1971.

[3] Z. Qi and Y. Liu. D-learning to estimate optimal individual treatment rules. *Electronic Journal of Statistics*, **12**:3601–3638, 2018.

[4] X. Shen, G. Tseng, X. Zhang, and W. Wong. On psi-learning. *Journal of the American Statistical Association*, **98**:724–734, 2003.

[5] I. Steinwart. Consistency of support vector machines and other regularized kernel classifiers. *IEEE Transactions on Information Theory*, **51**:128–142, 2005.

[6] J. Wang and X. Shen. Large margin semi-supervised learning. *Journal of Machine Learning Research*, **8**:1867–1891, 2007.

[7] J. Wang, X. Shen, and Y. Liu. Probability estimation for large-margin classifiers. *Biometrika*, **95**:149–167, 2008.

[8] Y. Zhao, D. Zeng, E. Laber, and M Kosorok. New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association*, **110**:583–598, 2015.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0035

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China

E-mail: he.xin17@mail.shufe.edu.cn

School of Data Science, City University of Hong Kong, Hong Kong, China

E-mail: shirongxu2-c@my.cityu.edu.hk

School of Data Science, City University of Hong Kong, Hong Kong, China

E-mail: j.h.wang@cityu.edu.hk