Published in final edited form as:

J Am Stat Assoc. 2017 ; 111(516): 1524–1525. doi:10.1080/01621459.2016.1244065.

Comments on “Personalized dose finding using outcome weighted learning”*

Jun Fan and Ming Yuan[†]

University of Wisconsin-Madison

We would like to first congratulate Professors Chen, Kosorok and Zeng for making yet another important contribution to the development of theory and methodology for using outcome weighted learning in personalized medicine. More specifically, they considered in this article how to infer an optimal individualized dose rule from a randomized trial.

Following the authors, let $X \in \mathcal{X}$ be the covariates for a patient, and $R \in \mathcal{R}$ the reward when assigning dose $A \in \mathcal{A}$ to the patient. The goal is then to find an optimal individualized dose rule $f: \mathcal{X} \rightarrow \mathcal{A}$ such that for a patient with covariate X , the dose assignment $A = f(X)$ maximizes the expected reward, or the so-called value function:

$$\mathcal{V}(f) = \mathbb{E}[R|A=f(X), X].$$

In the usual two-arm randomized trial setting where $\mathcal{A} = \{\pm 1\}$, we can rewrite

$$\mathcal{V}(f) = \mathbb{E}\left[\frac{1}{\mathbb{P}(A|X)} R(A=f(X)) | X\right],$$

leading to the notion of outcome weighted learning. The same idea, however, is not directly applicable for dose finding because $\mathcal{A}$ is now continuous. To address this challenge, Chen et al. proposed to approximate $\mathcal{V}(\cdot)$, first by $\tilde{\mathcal{V}}_\phi(\cdot)$, and subsequently by $\mathcal{V}_\phi(\cdot)$. An optimal dose assignment rule can then be learned by maximizing an empirical version

$$\mathcal{V}_\phi^n(f) = \frac{1}{n} \sum_{i=1}^n \frac{R_i}{2\phi p(A_i|X_i)} \min\left\{\frac{|A_i - f(X_i)|}{\phi}, 1\right\}$$

of $\mathcal{V}_\phi(f)$ based on trial data consisting of n independent copies $\{(X_i, A_i, R_i): 1 \leq i \leq n\}$ of the triplet (X, A, R) . Our goal here is to offer an alternative explanation of this seamless and seemingly tailor-made transition from an otherwise difficult value function $\mathcal{V}(\cdot)$ to the much more manageable $\mathcal{V}_\phi(\cdot)$ via $\tilde{\mathcal{V}}_\phi(\cdot)$, and argue that the two-step approximation employed by Chen et al. can actually be unified.

*Research supported in part by NSF FRG Grant DMS-1265202, and NIH Grant 1U54AI117924-01.

[†]Address for correspondence: Department of Statistics, University of Wisconsin-Madison, 1300 University Avenue, Madison, WI 53706.

More specifically, write

$$m(a, x) := (R|A=a, X=x) = \frac{1}{p_{A|X}(a, x)} \int_{\mathcal{R}} r p_{R,A|X}(r, a, x) dr,$$

where $p_{A|X}$ is the conditional density of A given X , and $p_{R,A|X}$ is similarly defined. Following the same idea of Nadaraya-Watson estimator for nonparametric regression, we can estimate $m(a, x)$ by

$$\hat{m}_{H,h}(a, x) = \sum_{i: X_i=x} \frac{H_h(a - A_i) R_i}{p_{A|X}(A_i, x)}$$

where H is a “density” kernel, and h is a bandwidth such that

$$H_h(z) = \frac{1}{h} H\left(\frac{z}{h}\right)$$

Thus,

$$\mathcal{V}(f) = \{[R|A=f(X), X]\} = \{m[f(X), X]\}$$

can be estimated by

$$\hat{\mathcal{V}}_{H,h}(f) = \frac{1}{n} \sum_{i=1}^n \hat{m}_{H,h}[f(X_i), X_i] = \frac{1}{n} \sum_{i=1}^n \frac{H_h(f(X_i) - A_i) R_i}{p_{A|X}(A_i, X_i)}.$$

Immediately, we notice that $\hat{\mathcal{V}}_{H,h}(\cdot)$ defined above becomes $\tilde{\mathcal{V}}_{\phi}^n(\cdot)$ of Chen et al. when using a uniform kernel

$$H(z) = (|z| \leq 1)/2$$

and bandwidth $h = \phi$. Similarly, $\hat{\mathcal{V}}_{H,h}(\cdot)$ becomes $\mathcal{V}_{\phi}^n(\cdot)$ when using a triangular kernel

$$H(z) = (1 - |z|)(|z| \leq 1)$$

and bandwidth $h = \phi$. In other words, we can view $\tilde{\mathcal{V}}_{\phi}$ and $\mathcal{V}_{\phi}$ as the same type of approximation to $\mathcal{V}$ by using different kernels.

This alternative interpretation obviously also opens the door for devising other approximations to the value function by using different density kernels, which may lead to further improvements. To explore such possibilities, we conducted a small numerical experiment similar to Chen et al. In particular, we consider two different Q functions:

$$Q_1(X, A) = 8 + 4X_1 - 2X_2 - 2X_3 - 15|f * (X) - A|,$$

$$Q_2(X, A) = 8 + 4X_1 - 2X_2 - 2X_3 - 25(f * (X) - A)^2,$$

where, in both Q_1 and Q_2 ,

$$f * (X) = 1 + 0.5X_1 + 0.5X_2.$$

We drew X uniformly from $[0, 1]^{30}$, A uniformly from $[0, 2]$, and R from $\mathcal{N}(Q(X, A), 1)$. For simplicity, we considered only linear kernels so that $f(X) = \alpha + X^\top \beta$. We estimated and β by

$$(\hat{\alpha}, \hat{\beta}) = \arg \min \left\{ \frac{1}{nh} \sum_{i=1}^n R_i H \left(\frac{A_i - \alpha - X_i^\top \beta}{h} \right) + \lambda \|\beta\|_{\ell_2}^2 \right\}.$$

Four different polynomial kernels were considered:

$$H_1(z) = (1 - |z|)(|z| \leq 1),$$

$$H_2(z) = \frac{3}{4}(1 - z^2)(|z| \leq 1),$$

$$H_3(z) = \frac{15}{16}(1 - z^2)^2(|z| \leq 1),$$

$$H_4(z) = \frac{35}{32}(1 - z^2)^3(|z| \leq 1),$$

The results, averaged over 200 simulated datasets with sample size $n = 800$, are reported in Table 1.

The averaged rewards reported in Table 1 show that it could be beneficial to use kernels other than H_1 . Obviously, this is only a small set of experiment. But the competitive performances illustrate the merits of thinking beyond $\mathcal{V}_\phi$, and warrants further investigation. These findings indeed have motivated us to more rigorously examine the effect and potential gains of using different kernels, which we hope to be able to report elsewhere in the near future.

Table 1

Effect of Kernels: realized reward value averaged over 200 simulation runs for four different choices of kernel. Standard errors are also reported in parenthesis.

Kernels	H_1	H_2	H_3	H_4
Q_1	7.52(0.03)	7.88(0.005)	7.83(0.007)	7.79(0.008)
Q_2	6.25(0.097)	7.20(0.024)	7.09(0.055)	6.91(0.074)