Statistics and Probability Letters 107 (2015) 178–182

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Balancing scores for simultaneous comparisons of multiple treatments

Yannis G. Yatracos

Faculty of Communication and Media Studies, Cyprus University of Technology, Cyprus

ARTICLE INFO

Article history:
Received 12 April 2015
Received in revised form 22 August 2015
Accepted 22 August 2015
Available online 1 September 2015

Keywords:
Balancing factorization criterion
Causal inference
Matching
Propensity function
Sufficient statistic

ABSTRACT

Criteria based on the T-treatment's posterior density, $q_T(t|x)$, provide *conclusively* a balancing-sufficient score, s, and the coarsest, c, for the treatments indexing the X-covariates' *unknown* density. Strong ignorability of treatment assignment given s, allows using s and c for simultaneous comparisons of multiple treatments.

1. Introduction

In this work, Bayesian and frequentist theories together provide balancing scores conclusively, shedding additional light to the propensity score and its extensions. The main tool is the Balancing Factorization Criterion (BFC) for posterior densities, $\{q(t|x), x \in \mathcal{X}\}$, which provides the sufficient statistic, s, for the parameter, t, in the unknown family of the X-covariates' densities, $\{p(x|t), t \in \mathcal{T}\}$.

In observational studies for treatments' causal effects, units in different treatment groups may have substantial pre-treatment differences due to lack of random assignment. The remedy is to include in each treatment group units with pre-treatment covariates, x, "similar" to those of units in the other groups. When each reservoir population is available for testing a different treatment, t, the sufficient statistic s of the covariates' X-densities, one density for each population, can be used to create similar treatment groups. However, the X-densities are usually unknown and it is not easy either to model or to estimate $p(x|t)$ because x's dimension is usually large. In Causal Inference s is called balancing score, the minimal sufficient statistic c is the coarsest score and the posterior density, $q(t|x)$, of the treatment variable, T, has known form. Thus, $q(t|x)$ *is used herein instead of* $p(x|t)$ to obtain either s or c, respectively, with BFC or the Coarsest Balancing Score Criterion (CBSC).

The consequences of this work are: (i) tools are provided to obtain s and c from the treatment variable model $q(t|x)$, (ii) it is seen that often s and c are not scalar, (iii) it is shown that strong ignorability of treatment assignment given any s-value holds, and (iv) it is shown that for matched units from the reservoir populations with the same s-value, s_0, *any* differences between the sample treatment means are *simultaneously* unbiased for the corresponding average causal effects of treatments, conditionally on $s = s_0$ and also unconditionally.

For two treatments, $t = 1, 2$, Rosenbaum and Rubin (1983) propose the groundbreaking scalar propensity score $e(x)$, i.e., $q(1|x)$, to balance the x-covariates of the n units in the treatment groups. They show that if treatment assignment and the potential units' responses to treatments, $r_i(1)$ and $r_i(2)$, $i = 1, \ldots, n$, are conditionally independent given x, then the

E-mail address: yannis.yatracos@cut.ac.cy.

http://dx.doi.org/10.1016/j.spl.2015.08.018
0167-7152/

difference between the sample treatments' means given $e(x)$ is unbiased for the average causal effect $E\{r(2) - r(1)\}$; E denotes expectation over the whole population.

For more than two treatments, Joffe and Rosenbaum (1999) study causal effects using a small number of balancing linear functions of x. For a finite number of multi-valued categorical treatments, Imbens (2000) introduced the generalized propensity score $P(T = t|x)$, i.e., $q(t|x)$, and used it to estimate average causal effects for treatments' pairs but not for simultaneous causal comparisons of all treatments; see Imbens (2000, page 709, lines 15–17 and lines -4 to -1).

For general treatment regimes with any type of t-values, Imai and van Dyk (2004) introduce the propensity function $e_\psi(\cdot|x) = q_\psi(.|x)$ (p. 856) that has the form of $q(t|x)$; ψ is a known parameter. *Assumption* 3 is made, that for all x-values $e_\psi(\cdot|x)$ depends on x through a unique, finite dimensional parameter $\theta_\psi(x)$, i.e., $e_\psi(\cdot|x) = e_\psi(\cdot|\theta_\psi(x))$. Strong ignorability of treatment assignment given $e_\psi(\cdot|x)$ is also established (*Result* 2). In the examples, $\theta_\psi(x)$ is used for matching and sub-classification instead of $e_\psi(\cdot|\theta_\psi(x))$. This raises questions on the real need for e_ψ; see also Example 3.3 herein where it is shown that $\theta_\psi(x)$ is balancing score. Our results are obtained without any restriction on $q(t|x)$.

Bahadur recognized that for two treatments $e(x)$ is equivalent to the likelihood-ratio of the X-densities which is minimal sufficient (Rubin and Thomas, 1996, p. 250). Noorbaloochi et al. (2010) considered likelihood-ratios as balancing scores when the number of treatments is either finite or infinite but assumed that strong ignorability of treatment assignment given s holds (p. 12, lines -2, -1, p. 13, lines 1–4). The approach and some missing details on the likelihood-ratios matching theory can be found in Yatracos (2014). The likelihood-ratios approach is criticized by followers of the propensity score (and its extensions) because $p(x|t)$ is usually unknown and the curse of the x-dimensionality affects the likelihood ratios' estimates. This criticism does not hold when *BFC* and CBSC are used to obtain, respectively, s and c.

The goal of this work is to obtain balancing scores in a general framework. BFC and CBSC, both motivated and derived from sufficiency, are the criteria-tools used to obtain s and c. As criteria, BFC and CBSC are neither related, nor extensions of the generalized propensity score (Imbens, 2000) and the propensity function (Imai and van Dyk, 2004) which are balancing scores, similar in form to the propensity score. In addition, the summaries in Imbens (2000) and Imai and van Dyk (2004) clarify from the outset that their goals are, respectively, the extension and the generalization of the propensity score.

The approach herein explains clearly what "matching of similar units" means. When reservoir populations are available each for a different treatment, units from these populations form a matching group when their covariates provide the same information, i.e. have the same s-value, s_0. Such groups can be used in causal comparisons, e.g., to determine the "right" dose for a new drug, by examining *simultaneously* the sample means estimating either the expected response differences $E\{r(t_2) - r(t_1)|s = s_0\}, \ldots, E\{r(t_k) - r(t_{k-1})|s = s_0\}$ or their expected values over all s_0, for doses' levels $t_1 < \cdots < t_k$.

Estimates of c and s may not be sufficient statistics and the same also holds for estimates of the propensity score, the generalized propensity score and the propensity function. This has been neglected so far in the Causal Inference literature that has not its own tools to confirm the balancing property, unlike the already existing frequentist theory for estimates of s and c; see, e.g., Le Cam (1964) and Joyce and Marjoram (2008).

The framework is presented in Section 2. The main results are in Sections 3 and 4. The proofs are in the Appendix.

2. Causal inference framework and assumptions

For a random vector U use $p_U(u)$ to denote its density (but also its probability). When random vector V is also available use $p(u|v)$ to denote the conditional density of U given V. Let $\mathcal{T}$ denote the treatments and let T be the treatment variable with values t in $\mathcal{T}$ and prior density π_T. Treatment t is used in selected units of population $\mathcal{P}_t$ having balanced x-covariates with respect to $\mathcal{T}$. The units in $\mathcal{P}_t$ have covariates $x \in \mathcal{C}(\mathcal{P}_t) \subset R^d$ and unless otherwise stated it is assumed that $\mathcal{C}(\mathcal{P}_t) = \mathcal{C}, t \in \mathcal{T}$. Let $p(x|t)$ denote the covariates' X-density of units in $\mathcal{P}_t$ and let $\mathcal{D}_\mathcal{T} = \{p(x|t),\ t \in \mathcal{T}\}$; $p_X(x)$ is the marginal density of the x-covariates. The notation $p(x|t)$ does not mean necessarily that p is the same density with the parameter t changing, $t \in \mathcal{T}$, but simply denotes the covariates' distribution in $\mathcal{P}_t$. Use $q(t|x)$ to denote T's posterior density (or probability) given the x-covariates. For unit i, $r_i(t)$ is the response for treatment t and the potential outcomes $\mathcal{R}$ is the set $\{r_i(t), t \in \mathcal{T},\ \text{for } i = 1, \ldots, n\}$. The average outcome for treatment t is denoted by $Er(t),\ t \in \mathcal{T}$, and we are interested simultaneously in the conditional and unconditional average causal effects, respectively, $E\{r(t_k) - r(t_m)|s = s_0\}$ and $E\{r(t_k) - r(t_m)\}$, for various choices of t_k and t_m from $\mathcal{T}$. Conditional independence of x and y given z is denoted by $x \perp\!\!\!\perp y|z$ (Dawid, 1979). The expression "covariates u, v match" means that the units with these covariates match.

Assumption 1 (*Stable Unit Treatment Value Assumption (SUTVA), Rubin, 1980, 1990*)**.** The distribution of potential outcomes for one unit is assumed to be independent of potential treatment status of another unit given the observed covariates.

Assumption 2 (*Strong Ignorability of Treatment Assignment Given x, Rosenbaum and Rubin, 1983*)**.**
(i) $\mathcal{R}$ and T are conditionally independent given x: $\mathcal{R} \perp\!\!\!\perp T\ |x$, and
(ii) for every $t \in \mathcal{T}$, $0 < p(t|x)$ (or equivalently $0 < p(x|t)$).

Recall that $b(x)$ is a balancing score if the conditional distribution of x given $b(x)$ is the same for all treatment values, i.e.,

$$p(x|t, b(x)) = p(x|b(x)) \quad \forall\, t \in \mathcal{T}. \tag{1}$$

From (1), thinking of t as parameter value for the distribution of x it follows that $b(x)$ is a sufficient statistic for the family $\mathcal{D}_\mathcal{T} = \{p(x|t); t \in \mathcal{T}\}$. The terminology balancing and coarsest balancing scores for $\mathcal{D}_\mathcal{T}$ are used herein more often, respectively, than sufficient and minimal sufficient statistics.