Consistent treatment selection based on Manski midpoints

Juraj Medzihorsky
Durham University

Tamás Rudas
Eötvös Loránd University

Abstract

The fundamental difficulty in comparing efficacies of two treatments based on observational data is that those who actually selected or received a particular treatment may give different responses to it than those who did not select or receive it. Thus, based on those who received one treatment, one cannot estimate the effect in case the treatment is applied to the entire population. Manski (1990) showed, that also in this situation, one may derive bounds for the effect size on the entire population. When only observational data are available, Rudas (2015a) suggested, in order to find the better effect, comparing the differences in the numbers the positive versus negative outcomes, in contrast to the usual method, when their ratios are used. He showed that this procedure is consistent in the sense that it never leads to Simpson's paradox. In this research, we develop Manski bounds for these differences, and show that the observed differences are the midpoints of the ranges obtained.

Treatment selection and effect measures

The task of selecting the best treatment from set $\mathcal{D}$ of available treatments consists of finding

$$d^* = \arg\max_{d \in \mathcal{D}} u(d),$$

where $u(d)$ is the net benefit of treatment d when selected for the whole population. For two treatments ($\mathcal{D} = \{0, 1\}$), this maps to decision function

$$\operatorname{sgn}\left[\delta(T)\right] := \begin{cases} -1 & \text{if } u(0) > u(1) \\ 0 & \text{if } u(0) = u(1) \\ +1 & \text{if } u(0) < u(1) \end{cases},$$

where $\delta(.)$ is an effect measure and T a table that holds the data.

Often, $u(d)$ is specified as $\mathrm{E}\left[Y(d)\right]$, where $Y(d)$ is the population distribution of outcome Y under treatment d. For two treatments, this yields effect measure

$$ATE(1) = \mathrm{E}\left[Y(1)\right] - \mathrm{E}\left[Y(0)\right], \quad \text{which under binary } Y \text{ equals} \quad \mathrm{P}\left[Y(1) = 1\right] - \mathrm{P}\left[Y(0) = 1\right].$$

In a setup with two treatments and two outcome values, the population is held by the 2×2 table

$$T = \begin{array}{|c|c|c|} \hline & Y_i = 0 & Y_i = 1 \\ \hline D_i = 0 & p_{10} & p_{11} \\ \hline D_i = 1 & p_{00} & p_{01} \\ \hline \end{array},$$

such that $\forall p_{dy} \in [0, 1]$ and $\sum_d \sum_y p_{dy} = 1$.

Keele and Quinn (2017) define four latent classes $Z \in \{0, 1, 2, 3\}$ in terms of their potential outcomes,

$Z_i = 0$: Never $Y_i = 1$

$Y_i(0) = 0$	
$Y_i(1) = 0$	

$Z_i = 1$: Helped

$Y_i(0) = 0$	
	$Y_i(1) = 1$

$Z_i = 2$: Hurt

	$Y_i(0) = 1$
$Y_i(1) = 0$	

$Z_i = 3$: Always $Y_i = 1$

	$Y_i(0) = 1$
	$Y_i(1) = 1$

giving a decomposition of T

Never + Helped	Hurt + Always
$Z_i \in \{0, 1\}$	$Z_i \in \{3, 4\}$
Never + Hurt	Helped + Always
$Z_i \in \{0, 3\}$	$Z_i \in \{2, 4\}$

and

$$ATE(1) = \mathrm{P}\left[Y(1) = 1\right] - \mathrm{P}\left[Y(0) = 1\right] = \mathrm{P}\left[Z \in \{1, 3\}\right] - \mathrm{P}\left[Z \in \{0, 2\}\right] = \mathrm{P}[Z = 1] - \mathrm{P}[Z = 2].$$

Consistent treatment selection with observational data

For any unit i, at most one of $\left\{Y_i(0), Y_i(1)\right\}$ is observable. Thus, only estimates $\left\{\hat{Y}(0), \hat{Y}(1)\right\}$ are available. In experiments, randomized treatment assignment makes the assumption of $\left(Y(1), Y(0)\right) \perp\!\!\!\perp D$ and thus also $D \perp\!\!\!\perp Z$ viable, which yields the conventional estimator

$$\widehat{ATT}_r(1) = \mathrm{P}\left[Y(1) = 1\right] - \mathrm{P}\left[Y(0) = 1\right] = \mathrm{P}[Y = 1, D = 1] - \mathrm{P}[Y = 1, D = 0] = \frac{p_{11}}{p_{10} + p_{11}} - \frac{p_{01}}{p_{00} + p_{01}}.$$

In observational data, confounders X may affect both treatment assignment and outcomes. Thus, only conditional independence $(Y(1), Y(0)) \perp\!\!\!\perp D|X$ holds, and if X is not observed, one cannot estimate $\{Y(0), Y(1)\}$.

Suppose two strata $S \in \{0, 1\}$ held by tables

$$T_{S=0} = \begin{array}{|c|c|} \hline p_{10}^{S=0} & p_{11}^{S=0} \\ \hline p_{00}^{S=0} & p_{01}^{S=0} \\ \hline \end{array} \qquad \text{and} \qquad T_{S=1} = \begin{array}{|c|c|} \hline p_{10}^{S=1} & p_{11}^{S=1} \\ \hline p_{00}^{S=1} & p_{01}^{S=1} \\ \hline \end{array}$$

where the frequencies sum up to 1 within each table. Under conventional effect measures such as $\widehat{ATT}_r$, the stratified tables may lead to one decision and the combined table to another,

$$\operatorname{sgn}\left[\delta(T_{S=0})\right] = \operatorname{sgn}\left[\delta(T_{S=1})\right] \neq \operatorname{sgn}\left[\delta(T)\right],$$

a consequence of association reversal known as Simpson's paradox. This poses a problem for treatment selection if S is not observed or scientific information lacks whether to condition on it.

Manski bounds and midpoints

Manski (1990) showed, that also in settings with observational data and unobserved confounding, one may derive bounds for the effect size on the entire population. For treatment d, the bounds on the response in the entire population are

$$mb\left(\mathrm{P}\left[Y(d) = 1\right]\right) = [2p_{d1} - 1, 1 - 2p_{d0}], \quad \text{with midpoint} \quad mm\left(\mathrm{P}\left[Y(d) = 1\right]\right) = p_{d1} - p_{d0},$$

where $mb()$ and $mm()$ are Manski bounds and midpoints, respectively. Under two treatments,

$$mb\left(ATE(1)\right) = \left[-(p_{01} + p_{10}), (p_{00} + p_{11})\right], \quad \text{and} \quad mm\left(ATE(1)\right) = \frac{1}{2}\left(\left(p_{00} + p_{11}\right) - \left(p_{01} + p_{10}\right)\right) = \frac{1}{2}DI,$$

where DI is a special case of DI_k (Rudas, 2015a). Decision function $\operatorname{sgn}\left[DI(T)\right]$ selects

$$d^* = \arg\max_{d \in 0,1} mm\left(ATE(d)\right), \quad \text{always yielding} \quad \operatorname{sgn}\left[DI(T_{s=0})\right] = \operatorname{sgn}\left[DI(T_{s=1})\right] = \operatorname{sgn}\left[DI(T)\right].$$

Statistical properties

Rudas (2015a) shows that under multinomial sampling with $p_{00} + p_{11} = r$,

$$\mathrm{P}\left[DI > 0\right] = \sum_{k=[N/2]+1}^{N} \binom{N}{k} r^k (1 - r)^{N-k}, \quad \text{approximated for large } N \text{ by} \quad \phi\left(\sqrt{N}\frac{r - 0.5}{\sqrt{r(1 - r)}}\right).$$

where ϕ is the standard normal CDF, and the probability of the correct decision depends only on N and the population value of DI.

Relationship to sensitivity analysis

For $T = \{n_{00}, n_{01}, n_{10}, n_{11}\}$, $\sum_d \sum_y n_{dy} = N$, there are $\prod_d \prod_y (n_{dy} + 1)$ possible allocations of the N observations into the four Z classes. Inspecting all such possible allocations relates to conditional volume tests (Diaconis and Efron, 1985), as $2 \times 2 \times 4$ tables (D, Y, Z) with the same 2×2 marginals (D, Y) and 8 of the cells fixed to zero. Then, DI equals the expected effect if all allocations in Z are equally likely. Keele and Quinn (2017) propose a generalization that weights the allocations according to extra-data information, with simulated cell-splitting parameters $\{\psi\}_{dy}$ on the (D, Z) margin

	$Z_i = 0$	$Z_i = 1$	$Z_i = 2$	$Z_i = 3$
$D_i = 0$	$(1 - \psi_{00})\theta_{00}$	$\psi_{00}\theta_{00}$	$(1 - \psi_{01})\theta_{10}$	$\psi_{01}\theta_{01}$
$D_i = 1$	$(1 - \psi_{10})\theta_{10}$	$(1 - \psi_{11})\theta_{11}$	$\psi_{10}\theta_{10}$	$\psi_{11}\theta_{11}$

,

where $\{\theta\}_{dy}$ are posterior cell probabilities.

Discussion

- just a cites holder Rudas (2010) (Rudas, 2015b) Rudas, 2015a

References

Diaconis, Persi and Bradley Efron (1985). "Testing for independence in a two-way table: new interpretations of the chi-square statistic". In: *The Annals of Statistics* 13.3, pp. 845–874.
Keele, Luke and Kevin M Quinn (2017). "Bayesian sensitivity analysis for causal effects from 2 × 2 tables in the presence of unmeasured confounding with application to presidential campaign visits". In: *The Annals of Applied Statistics* 11.4, pp. 1974–1997.
Manski, Charles F (1990). "Nonparametric bounds on treatment effects". In: *The American Economic Review* 80.2, pp. 319–323.
Rudas, Tamás (2010). "Informative allocation and consistent treatment selection". In: *Statistical Methodology* 7.3, pp. 323–337.
– (2015a). "Directionally collapsible parameterizations of multivariate binary distributions". In: *Statistical Methodology* 27, pp. 132–145.
– (2015b). "Effects and interactions." In: *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 11.4, p. 142.

Acknowledgments

This research project was supported by European Research Council, Grant 724191, PI: Staffan I. Lindberg, V-Dem Institute, University of Gothenburg, Sweden.