Why a Bayesian researcher might prefer observational data

Macartan Humphreys

May 2, 2016

Abstract

I give an illustration of a simple problem in which a Bayesian researcher can choose between random assignment of a treatment or delegating assignment to an informed—but motivated—agent. In effect she compares between learning from an RCT or from an observational study. She evaluates designs according to an expected squared error criterion (ESE). I show that for a small problem ($n = 2$) if she starts with a prior expectation of no treatment effect but believes that the agent is an advocate of treatment with probability q (and otherwise an opponent) then for all values of q she does at least as well delegating assignment as she does from an RCT and she does strictly better as long as $q \neq 0.5$. For other priors on treatment effects, randomization can dominate delegation or be dominated by it. As n grows the expected errors from an RCT design fall but errors from delegated assignment do not. Although there is always *some* prior such that a delegated procedure beats randomized assignment, the converse is not true. For a given prior there may be no delegated procedure that trumps an RCT. With uniform priors for example the RCT dominates delegated assignment for all beliefs on agent motivations when $n \geqslant 4$.

1. Introduction

Researchers generally prefer analyzing experimental data—data generated by controlled, and preferably random assignment of treatments—to observational data. A driving concern is that when treatment is not controlled it may be systematically related to potential outcomes in ways that muddy inference. This short paper questions this preference. I examine a simple problem to assess the performance of researcher-controlled assignment (including random assignment) relative to a delegated procedure that may use information not available to researchers. This lets us identify conditions under which delegated assignment trumps controlled assignment. Ironically, when observational data is optimal it is optimal for the very reason that researchers dislike it: the realized assignments carry information about potential outcomes. When controlled assignment dominates it is because it allows for the observation of data patterns that are censored by motivated assignment procedures.[1]

[1]Note that this problem is distinct to the problem of assessing how a researcher should optimally assign units to treatment conditional on available data. A longstanding argument suggests that when ancillary information is available about research units, randomization is not a strictly dominant procedure to generate data to learn about causal effects. If randomization is optimal, so too must be every assignment that receives positive probability under the randomization scheme, assuming that the inferences from these possible assignments do not depend on the assignment procedure in question (see Stone (1969) for an older treatment, see Rubin (1978) and Berry and Kadane (1997) for defenses of randomization; see Kasy et al. (2013) and Banerjee, Chassang and Snowberg (2016) for more

2. Set up

There is a binary treatment $X = \{0,1\}$ and outcome $Y = \{0,1\}$. A unit is of causal type a if $Y(0) = 1, Y(1) = 0$, b if $Y(0) = 0, Y(1) = 1$, c if $Y(0) = 0, Y(1) = 0$, and d if $Y(0) = 1, Y(1) = 1$.

There are $2m$ units and a researcher's prior, p, places probability on three events:

- ω_1: all units are b types
- ω_2: half the units are c types and half are d types
- ω_3: all units are a types

These three events imply average treatment effects of 1, 0, and -1, respectively.

This set up evidently admits only a very constrained set of events but for our purposes it has the advantage of allowing for simple analytic solutions and sharp inferences. Note that if half the units are randomly assigned to treatment, then in event ω_2 all the c units are assigned to treatment with probability $\phi(m) = \frac{m!m!}{(2m)!}$, all are assigned to control with the same probability and there are mixed assignments with probability $1 - 2\phi(m)$.

The researcher has access to no further information on units and so all assignments controlled by the researcher are equivalent to random assignment from her perspective. This helps distinguish the problem of delegation (or optimal assignment *procedure*) from the problem of optimal assignment. For this reason I will use the terms "randomized assignment" and "controlled assignment" interchangeably in what follows.

Assume that the researcher is interested in assessing the average treatment effect (ATE) and suffers squared loss in errors in her assessments. If she has a uniform prior over ω_1, ω_2, and ω_3, then her prior expectation of the ATE is 0. This means with probability 1/3 her squared error is 1^2, with a 1/3 probability it is 0, and with a one third probability it is $(-1)^2$, for a total expected error of $\frac{2}{3}$. More generally, with prior p her expected value for the ATE is $p_1 - p_3$ and her expected loss is $p_1 + p_3 - (p_1 - p_3)^2$. This prior loss is the same for all m.

To improve on this loss she compares experimentation strategies in which (a) assignment is randomized (b) assignment is determined by an agent who is informed about treatment effects but is also motivated in the sense that she has preferences over findings. In both cases exactly half the units are to be assigned to treatment and half to control and inference is undertaken taking the assignment procedure into account (unlike in Berry and Kadane (1997)).

3. Results

I provide results first for the controlled assignment procedure ("RCT") and then for the delegated procedure.

recent treatments). This note focuses instead on assignment *procedures* and contrasts the optimal controlled assignment to procedures in which researchers give up control. In this case inferences from realized data depends on how assignments were made since assignment can carry information about potential outcomes.

3.1 Posteriors from an RCT

For intuition, consider first the case with just two units ($m = 1$). In this case only two data patterns will be observed, one with $Y = X$ for both units and one with $Y = 1 - X$ for both units. The first pattern is consistent with ω_1 and ω_2; the second with ω_2 and ω_3. If a $Y = X$ pattern is observed the researcher places a $\frac{p_1}{p_1+\frac{1}{2}p_2}$ probability on ω_1 and zero probability on ω_3 for an expected ATE of $\frac{2p_1}{2p_1+p_2}$. Similarly if a $Y = 1 - X$ pattern is seen the researcher's expected ATE is $\frac{-2p_3}{p_2+2p_3}$. From this we can calculate the expected squared error as:

$$p_1\left(1-\frac{p_1}{p_1+.5p_2}\right)^2 + p_2\frac{1}{2}\left(\frac{p_1}{p_1+.5p_2}\right)^2 + p_2\frac{1}{2}\left(-\frac{p_3}{.5p_2+p_3}\right)^2 + p_3\left(-1+\frac{p_3}{.5p_2+p_3}\right)^2$$

This simplifies to: $p_2\left(\frac{p_1}{2p_1+p_2}+\frac{p_3}{2p_3+p_2}\right)$.

This generalizes in a simple way for arbitrary m. If data pattern $Y = X$ is observed, the probability of ω_1 is $\frac{p_1}{p_1+p_2\phi(m)}$. If a mixed pattern is observed ω_2 is inferred with certainty. From this we calculate the ESE from the RCT as:

$$\text{ESE}_{\text{RCT}} = p_2\left(\frac{\phi(m)p_1}{p_1+\phi(m)p_2}+\frac{\phi(m)p_3}{p_3+\phi(m)p_2}\right)$$

This expected error falls rapidly with m because as m increases it becomes increasingly unlikely that ω_2 will give rise to a data pattern that could be confused with ω_1 or ω_3.

3.2 Posteriors from assignment by delegation

Say now that the researcher could ask an agent to assign half the units to treatment and half to control. Say that with probability $q \in (0,1)$ the agent is an *informed advocate* ("pro") of the treatment and wants the researcher to believe that there is a large treatment effect. With probability $1-q$ the agents wants the researcher to believe there is a small (or negative) treatment effect ("anti"). We then have a Bayesian game in which the state of the world is known to the agent but not the researcher, the agent makes an optimal assignment given the inferences that will be made by the researcher, and the researcher makes optimal inferences given the strategies of the agent. If data arises that should arise with 0 probability in equilibrium we require only that the researcher makes an inference on ω consistent with the data.

Equilibrium. In the perfect Bayesian equilibrium of this game the *pro* agent assigns the *c* case to control and the *d* case to treatment in ω_2, producing a $Y = X$ data pattern. She has no meaningful discretion in Events 1 and 3. Similarly the *anti* agent assigns the *c* case to treatment and the *d* case to control in ω_2. The researcher's estimated ATE is $\frac{1}{1+qp_2/p_1}$ upon observing pattern $Y = X$ and $\frac{-1}{1+(1-q)p_2/p_3}$ upon observing pattern $Y = 1 - X$.

To check the inferences of the researcher in this equilibrium note that with these strategies a $Y = X$ pattern arises in ω_1 and in ω_2 when the agent is pro; a $Y = 1 - X$ pattern arises in ω_3 and in ω_2 when the agent is anti. Thus, with prior p, pattern $Y = X$ is observed with probability $p_1 + qp_2$ leading to estimated ATE of $\frac{1}{1+qp_2/p_1}$. Pattern $Y = 1 - X$ is observed with probability $p_3 + (1-q)p_2$ leading to estimated ATE of: $\frac{-1}{1+(1-q)p_2/p_3}$. Note that with q bounded away from 0 and 1 there is no need to worry about out-of-equilibrium beliefs in the $m = 1$ case since both

$Y = X$ data $Y = 1 - X$ data arise with positive probability. For the $m > 1$ case any data pattern in which Y and X are not perfectly correlated can result only from ω_2 and results in an inference of ATE = 0. To check the optimality of the agent's actions it is enough to note the different sign of the inferences for each data pattern.

In this equilibrium the ESE is:

$$\text{ESE}_{\text{DEL}} = p_2 \left(\frac{qp_1}{p_1 + qp_2} + \frac{(1-q)p_3}{p_3 + (1-q)p_2} \right)$$

This expected error does not fall with m because for any m the agent can produce a data pattern that makes ω_2 indistinguishable from ω_1 or ω_3.

3.3 Implications

In the $m = 1$ case, delegation yields the same ESE as an RCT at the point of maximal uncertainty $q = 0.5$. The RCT trumps the delegated design if $p_1 \left(\frac{1}{2p_1+p_2} - \frac{q}{p_1+qp_2} \right) + p_3 \left(\frac{1}{2p_3+p_2} - \frac{(1-q)}{p_3+(1-q)p_2} \right) < 0$. More simply, control trumps delegation for q in the interval $\left(\frac{1}{2}, \frac{p_1(3p_3p_1+p_2p_1-p_3^2)}{p_1^2(2p_3+p_2)+p_3^2(p_2+2p_1)} \right)$. This interval is empty when the researcher has uniform priors on treatment effects ($p_1 = p_2 = p_3$). Equivalently, with uniform priors random assignment never trumps delegated assignment.

With uniform priors and $m > 1$, the ESE from delegation exceeds the RCT for *all* values of q.

For non uniform priors, the delegated design can still trump the RCT however. This is the case for example for $m = 2$ with $p_1 = p_3$, $q = 1$ and $p_2 > 1 - 2\phi(m)$. Here the large prior on ω_2 means that the agent has considerable discretion. If an advocate produces data with a $Y = 1 - X$ pattern this gives great confidence in what was thought to be an unlikely event, ω_3.

Figure 1 illustrates how with uniform priors the RCT is trumped by delegation for $m = 1$ $(n = 2)$ but trumps delegation for $m = 2$ $(n = 4)$ for almost all values of q. For non-uniform priors, an RCT can be trumped or trump a delegation design depending on q.

4. Conclusion

I examine a problem in which a researcher can decide whether to assign a treatment using randomization (controlled assignment) or delegate assignment to a motivated agent.[2] A key findings is that delegation can improve inference because it provides information about potential outcomes that is lost when treatment assignment is randomized. In the case with two units and uniform priors, delegation generically does better than randomization. However the relative gains from delegation can decline rapidly with scale. In the case examined here, inference from delegated assignment does not enjoy *any* gains from an increase in scale while inferences from randomized designs gain rapidly in precision. Randomized assignment enjoys gains from scale while delegated assignment does not because it allows for a range of assignments that get censored by non random processes. The fact that these assignments are possible alters inferences, whether or not they occur.

[2]These results hold also with a non-strategic assignment mechanism or with self selection when these produce similar assignment procedures. For example in this case allowing self selection is optimal if c types are expected to select into treatment and d types select into control.

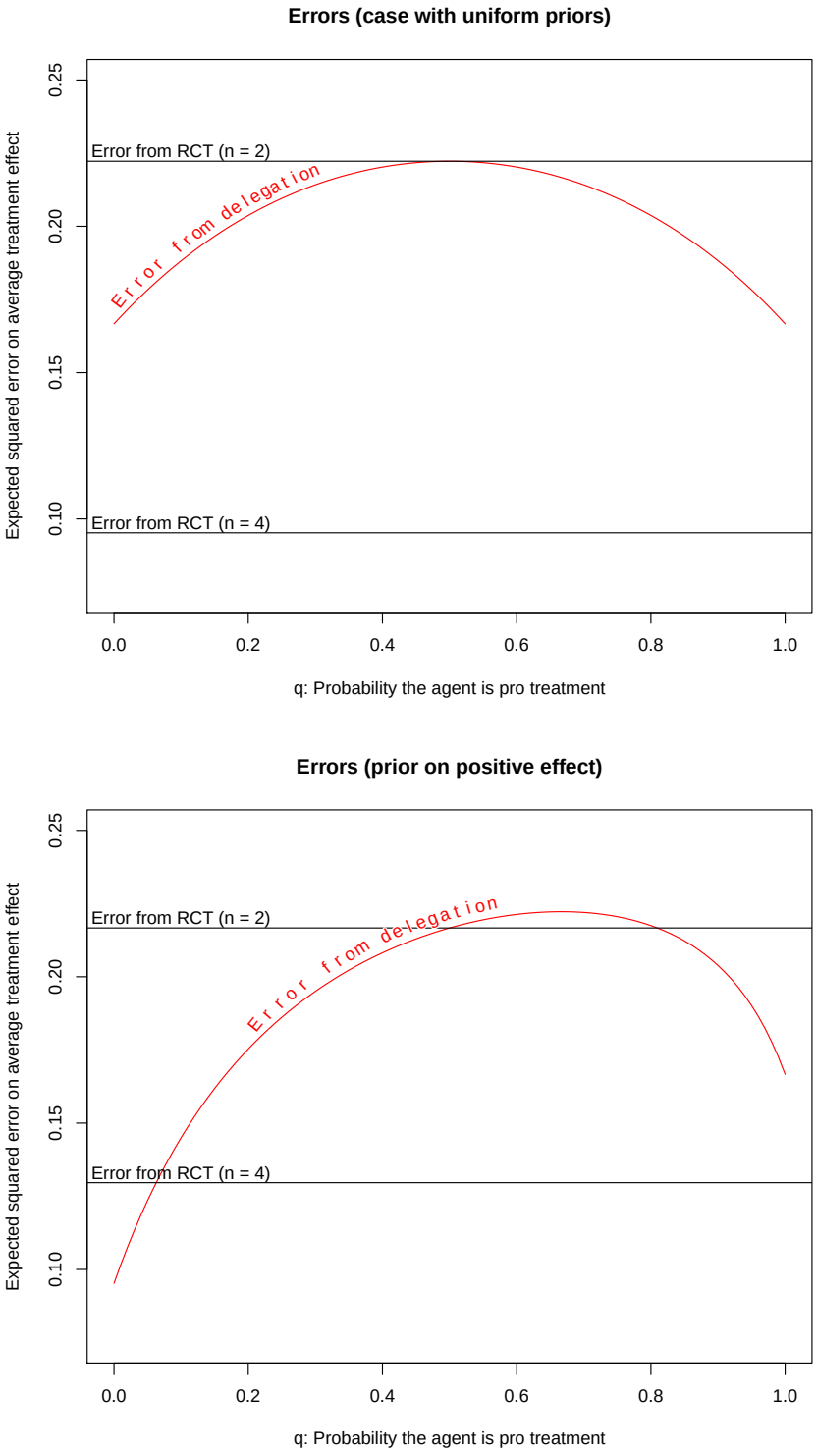

Figure 1: Expected error as a function of q given uniform priors $p = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ (upper figure) and non uniform priors $p = \left(\frac{2}{9}, \frac{6}{9}, \frac{1}{9}\right)$ (lower figure). Under uniform priors and with two units ($m = 1$), delegation trumps experimentation except at the point of maximal uncertainty about agent motivations. With four units ($m = 2$) the RCT always trumps delegated assignment. With non uniform priors, the RCT can trump or be trumped by delegation.

References

Banerjee, Abhijit, Sylvain Chassang and Erik Snowberg. 2016. "Decision Theoretic Approaches to Experiment Design and External Validity." *NBER WP 22167* .

Berry, Scott M and Joseph B Kadane. 1997. "Optimal bayesian randomization." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 59(4):813–819.

Kasy, Maximilian et al. 2013. "Why experimenters should not randomize, and what they should do instead." *European Economic Association & Econometric Society* .

Rubin, Donald B. 1978. "Bayesian inference for causal effects: The role of randomization." *The Annals of statistics* pp. 34–58.

Stone, M. 1969. "The role of experimental randomization in Bayesian statistics: Finite sampling and two Bayesians." *Biometrika* pp. 681–683.