Research Proposal #3

Noah Dixon

10-2-25

Research Proposal

Topic ($\neq$ Question): Can heterogeneous applications of policy learning be used to predict the effects of an optimally designed experiment, in a different environment than that which was initially run?

Main Idea: The idea is to synthesize the methodology of optimal policies (Athey & Wager 2021) [2] and optimally-designed experiments (Higbee 2024; Tabord-Meehan 2022; Watson et al. 2023) [5] [7] [8] to predict the "likelihood of generalization," conditional on a wide stratum of covariates and a restricted budget set. The former involves algorithmic welfare maximization; the latter, an optimal experimental design.

To test this, I will use two experimental counterparts. The first experiment is the "bread and butter." Its purpose is to utilize an optimal experimental design to determine the efficacy of "AI integration" – that is, literally giving AI to people as treatment – in a developing country. Covariates should be selected for the purpose of generalizing to another environment. That is to say, they should match census data; or, if census data is unavailable – or unreliable – easily ascertained via survey methods. Building off the results of the first experiment, the idea is to use this data to predict the optimal policy in another environment, using two methods: (i) the ATE, absent of covariates, and (ii) the welfare-maximizing policy, as devised in Athey and Wager (2021) and actualized from the transportability literature. In particular, by

using the welfare-maximizing policy as an additional piece of information, it may be possible to reconstruct a posterior probability map, conditional on the covariates extracted from census data (given that certain transportability conditions are satisfied). From here, a second experiment should be ran, using the current, standard methodology in the field – a "benchmark" RCT. This allows us to test the prediction from the first experiment. Clustering is not required – or allowed, by the methodologies utilized.

In particular, by leveraging two experimental counterparts, it allows me to simultaneously test numerous research questions:

I. The prediction accuracy of the first experiment when extrapolated to the second experiment's environment vs. the prediction accuracy of the second experiment when extrapolated to the first experiment's environment.

II. The added value of an optimally-designed experiment.

III. The relative efficacy of AI integration on different subpopulations.

Framework

Motivation: There is a conflation, in the literature right now, between two distinct types of experiments: (i) an experiment conducted for the purposes of *uncovering* a causal relationship, typically in a *contained environment* and (ii) an experiment conducted for the purposes of *projecting* a causal relationship. The former can be satisfied by quasi-experimental methods – and to the extent to which that is not possible – experimental methods. However, the latter asks a distinctly different question: how can we simultaneously minimize variance and cost, conditional on wide set of covariates, and *project* this relationship onto a larger population? While these methodologies overlap, I believe this distinction is important.

The standard – in lab experiments as well as field experiments – is to use ex-ante frequentist power analysis to justify an effect size – and consequently, a sample size that satisfies effect size. Ex-ante power analysis is conducted for two reasons: (i) to annihilate the "endogenous stopping rule" – the process of perpetually generating data until we statistically significant results – and (ii) as a means of uncovering a larger relationship in the population. However, this process, while satisfying (i), has unnecessarily restricted the scope and application of (ii). To illustrate the perils, statistically, of this process, consider the following example, which stems from Cohen (1973) [3]: suppose $\mu_A = 5.5$, $\mu_B = 5$, $\sigma_A = 1$, $\sigma_B = 1$, $\alpha = .05$, and $\beta = .8$. Under unequal sample sizes, let $d = \frac{\mu_A - \mu_B}{\sigma}$, where $\sigma = \sqrt{\frac{s_A^2 + s_B^2}{2}}$, and $n \approx \frac{2(Z_{1-\alpha/2} + Z_{1-\beta})^2}{d^2}$. Then,

$d = \frac{5.5-5}{\sqrt{(1^2+1^2)/2}} \approx .5$. However, suppose this canonical study is underpowered, such that $\mu_A^{true} = 5.25$ and $\sigma_A^{true} = .9$ reflect the true population parameters. Consequently, $d^{true} \approx 0.26$. Transmuting this to the required sample size, we have $n \approx \frac{2(1.96+.84)^2}{.5^2} = 62.72$. However, under d^{true}, $n^{true} = 253.76$. *Fundamentally, slight deviations, from the true population values, can lead to stark differences, in power analysis derivation.* The effects are even more severe when: (i) studies are compounded on top of each other, (ii) the designs of the initial study differ, (iii) the results are only prevalent in different subsets of the entire population, and (iv) when the "winner's curse" is prevalent (it always is). Additionally, it is fundamentally unclear what the effect sizes are, conditional on covariates (if implementing the CATE, for instance).

There are four main ways the literature has attempted to fix this issue: (I) simulation-based Bayesian methods (commonly used in medicine), (II) a sufficiently-large pilot, which can thereafter be used to identify the exact sample size, (III) Bayesian updating, which adheres to an algorithmic process, and which updates the sample size, *conditional* on new data coming in, and (IV) MDE, an ex-ante analysis of the minimum effect you are likely to see, conditional on a sample size and previously calculated sample sizes. (IV) is, like ex-ante frequentist power analysis, highly dependent on the ex-ante standard deviation, which is subject to the same scrutiny described above. A combination of (II) and (III) is likely the most applicable in the context of constructing an experiment for the purpose of *projecting* a causal relationship.

All of this ties into the larger question of external validity. Likely, it is not a formal, closed-form rule that we can follow, but a series of complicated algorithms, conditional on a well-defined canonical study (Lee, Correa & Bareinboim 2019) [6]. The econometrics literature has expanded into this area, but the applied literature is lagging behind (likely due to the complexity of the underlying methodology). ***Note that my goal, more generally, is to synthesize and operationalize existing econometric methods in the policy learning literature, optimal experimental design literature, and transportability literature – not to develop an entirely new econometric methodology.***

Athey and Wager (2021) Methodology: Define the data as (X_i, T_i, Y_i) where $T_i \in \{0, 1\}$, a policy class Π as the set of all possible rules, and define $W(\pi) = \mathbb{E}[Y(\pi(X))]$ as the expected outcome if treatments are assigned to rule $\pi : \mathcal{X} \to \{0, 1\}$.[1] The rule π can be anything (within reason). For instance, π could be that you only treat individuals over the age of 60 –

[1] In the literature, the individual-level estimate is given by, $\hat{\theta} = \frac{1}{n}\sum_i \psi_i$, where ψ_i is the doubly robust score for each individual i. A doubly robust score combines the observed outcome, treatment, and estimate nuisance functions. Formally,

$$\psi_i = \hat{m}_1(X_i) - \hat{m}_0(X_i) + \frac{T_i}{\hat{e}_i(X_i)}\big(Y_i - \hat{m}_1(X_i)\big) - \frac{1 - T_i}{1 - \hat{e}_i(X_i)}\big(Y_i - \hat{m}_0(X_i)\big),$$

where $\hat{m}_t(x)$ is the estimated regression function and $\hat{e(x)}$ is the estimated propensity score. In Athey and Wager (2021), $\hat{m}_t(x)$ is obtained using any supervised learning regression method (linear regression or logistic regression [the classical choice], lasso estimation or ridge estimation, random forests, boosted trees, neural nets, or causal forests) and $\hat{e(x)}$ is obtained using similar (albeit simpler) methods. Note that it is just the predicted probability of treatment assignment given $X = x$. Next, you can express the welfare of a

which might be the best policy, in terms of *maximizing* welfare in society.

The first step of this process is to *split* the data into two – or more – "folds". The first fold will be used to estimate the treatment effect, $\tau(X)$, and the second fold will be used to choose a policy rule that maximizes welfare.

To estimate the individual-level CATE, we must first compute individual doubly robust scores, $\psi(x)$ (see footnote 1). To estimate the CATE for *each* covariate profile, we apply ML methods (random forests, causal trees [Athey & Imbens, 2016], boosting, lasso, etc.) to smooth and aggregate the individual doubly robust scores $\hat{\psi}_i$ across values of X to derive $\hat{\tau}(x)$, where the true CATE is defined by:

$$\tau(x) = \mathbb{E}[Y(1) - Y(0)|X = x].$$

Next, we need to construct candidate policies, $\pi \in \Pi$, which essentially

---

policy π as,

$$\hat{\pi} = \underset{\pi}{\text{argmax}} \left\{ \frac{1}{n} \sum_{i=1}^{n} (2\pi(X_i) - 1)\hat{\psi} : \pi \in \Pi \right\},$$

where $\pi(X_i) = 1$ implies treatment, $\pi(X_i) = 0$ implies no treatment, and $\hat{W}(\hat{\pi})$ is the welfare from the optimal policy itself. The "-1" is a coding trick that gives a positive weight of "+1" to people treated and a negative weight of "-1" to people who are not treated. In this way, somebody who would negatively benefit from treatment will ***increase*** the welfare of society. The main result of this paper is that $\hat{\pi}$ has a regret bound,

$$R(\hat{\pi}) = \max_{\pi' \in \Pi} \mathbb{E}[Y_i(\pi'(X_i))] - \mathbb{E}[Y_i(\pi(X_i))] = \underset{\pi' \in \Pi}{max}\{W(\pi') - W(\hat{\pi})\},$$

which gives us the economic surplus from a policy, and which is "bounded on the order of $\sqrt{VC(\pi)/n}$". That is to say, $R(\hat{\pi}) = O_p(\sqrt{\frac{VC(\Pi)}{n}})$, which means that, as the sample size n grows, the regret shrinks towards 0 (although this depends on the complexity of the policy class, $VC(\Pi)$, where "VC" = VC dimension.

functions as a treatment assignment strategy. A typical policy rule is $\pi(x) = \mathbf{1}\{\beta' x > 0\}$, although there are certainly other alternatives. In general, the amount of covariates we use at this stage – versus the amount of covariates we use in the CATE stage – is significantly lower, to avoid overfitting and to make the rules interpretable. Lasso, ridge, and abiding by theoretical concerns (i.e., there are specific variables which are weighted more heavily in the analysis) or structural models are all viable strategies.

Now, on the *evaluation* fold (the second fold), we use inverse probability weighting (IPW) to assign people based on *each* policy rule,

$$\hat{W}(\pi) = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i \cdot \mathbf{1}\{T_i = \pi(X_i)\}}{p(T_i)},$$

where $p(T_i)$ is the randomization probability (we need to re-weight to get the true welfare effect since only a proportion of people are actually treated). In practice – and to fix bias and variance issues – they use doubly robust scores (see footnote). We iterate this process for *each* policy rule.

Finally, the algorithm selects $\hat{\pi} = \underset{\pi \in \Pi}{\text{argmax}}\{\hat{W}(\pi)\}$. This is the rule with the highest estimated welfare on the evaluation sample.[2]

There are a multitude of other components, in the paper, that are worth mentioning. The *regret bound* here is given by, $W(\hat{\pi}) \geqslant \sup_{\pi \in \Pi}\{W(\pi)\} - o_p(1)$.

[2]There are numerous assumptions used in this paper, which are quite complex. To condense these assumptions: if unconfoundedness holds, if the nuisance components are estimated consistently and remain well-behaved, if the CATE is consistently estimated, and if the policy class Π is properly restricted to avoid overfitting, then the learned policy $\hat{\pi}$ is asymptotically equivalent to the optimal policy $\pi^\star$. Note that heteroskedastic units are allowed, in this design.

Essentially, this says that, as your sample size grows, the welfare of the optimal policy is *almost* as good as the *true* policy, $\pi^{\star}$. To state it more simply, $W(\hat{\pi}) \xrightarrow{p} W(\pi^{\star})$ (and the gap shrinks at the rate $\sqrt{1/n}$). *In fact, this is precisely the main point of the paper.* As long as Π is properly controlled – that is, if is restricted and regularized – then the complexity is manageable. Cross fitting – which involves running the process on *each* fold, at separate times – should also be used to help control Π (this is the process we have iterated above). By leveraging an "optimal experimental design," it becomes more likely that the assumptions will hold.

Optimal Experimental Methodology: (Higbee 2024; Tabord-Meehan 2022; Watson et al. 2023) [5] [7] [8] are placeholders for a plethora of papers that must be further examined. One issue is ensuring that robust standard errors do not "corrupt" the process; that is, that the *i.i.d.* assumption can be relaxed. As a general rule, medicine seems to be significantly ahead of economics, in terms of devising an optimal study (perhaps because it is more willing to adopt Bayesian methods), while computer science is significantly ahead of all other fields, in the *projection* of causal effects (perhaps because it is more willing to adopt messy, algorithmic solutions).

External Validity: One way to generalize is to use a sufficiently large amount of clusters to fully encapsulate the larger population. However, in many contexts, this is infeasible. For instance, in 1997, Mexico launched Pro-

gresa, the first large-scale conditional cash transfer (CCT) program, which included 506 rural villages and 24,000 households. In practice, this is infeasible for most researchers. Further, while this study inspired similar programs in Latin America, there is little statistical evidence that this is a welfare-maximizing policy in those specific countries. In particular, for treatments of equal monetary status – e.g., farming subsidies, vaccines, etc. – there may be a policy that achieves even more welfare. Clustering can only generalize to the larger population in which it clusters (assuming that G is sufficiently large and the sample estimate $\hat{\theta}$ is fully representative of the larger population).

As described above, the process given by Athey and Wager (2021) [2] defines an optimal policy, based on a pre-defined data set (not an optimally designed experiment, like I propose). For instance, in Athey et al. (2024), they define an optimal policy to "nudge" students prior to financial-aid applications, showing that nudging all students is sub-optimal. However, in many contexts – and specifically in developmental interventions – these sorts of policy rules are unethical. "Treat everyone over the age of 50," for instance, is not something that can be advocated for, as a policy, especially if the treatment could *potentially* – and significantly – benefit any arbitrary member of the subpopulation. That being said, we can use this as an extra piece of information, when deciding how to extrapolate these results to new contexts. Beyond this, I believe that we can invert this policy function, to determine the *worst* policy. In this way, we can reconstruct a posterior

probability map,, conditional on the covariates we extract from census data. **More generally, the goal should be to significantly outperform the ATE – not recover a completely unbiased estimate.** If we have multiple treatments – and multiple policies – we can use this information to determine the optimal policy for each region, although this may be beyond the scope of this study.

In theory, there are a number of assumptions underlying this process, which I will need to examine further.

AI Integration as a Treatment Strategy: There are numerous aid strategies to study, in practice. However, the most applicable – and perhaps the most novel – is treating AI integration as the treatment itself. The idea is to give some people computers and to give some people computers *with* explicit – and conditional – AI training. In the literature, thus far, there is sufficient evidence that giving people computers is ineffective (Cristia et al. 2017; Angrist and Lavy 2002) [4] [1]. Likely, this is the result of: (i) targeting the incorrect people, and (ii) targeting people the wrong way. To my knowledge, nobody has attempted to test AI integration as a treatment strategy yet. Obviously, significantly more work needs to go into fleshing this idea out.

References

[1] Angrist, J. and V. Lavy (2002). New evidence on classroom computers and pupil learning. *The Economic Journal 112*(482), 735–765.

[2] Athey, S. and S. Wager (2021). Policy learning with observational data. *Econometrica 89*(1), 133–161.

[3] Cohen, J. (1973). Statistical power analysis and research results. *American Educational Research Journal 10*(3), 225–229.

[4] Cristia, J., P. Ibarrarn, S. Cueto, A. Santiago, and E. Severn (2017, July). Technology and child development: Evidence from the one laptop per child program. *American Economic Journal: Applied Economics 9*(3), 295–320.

[5] Higbee, S. D. (2024). Experimental design for policy choice.

[6] Lee, S., J. D. Correa, and E. Bareinboim (2019). General identifiability with arbitrary surrogate experiments. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 389–398.

[7] Tabord-Meehan, M. (2023, October). Stratification trees for adaptive randomisation in randomised controlled trials. *The Review of Economic Studies 90*(5), 2646–2673.

[8] Watson, S. I., A. Girling, and K. Hemming (2023, July). Optimal study designs for cluster randomised trials: An overview of methods and results.