Treatment and Welfare Learning

Edward Jee EDJEE@UCHICAGO.EDU
Kenneth C. Griffin Department of Economics
University of Chicago
Chicago, IL 60637, USA

Abstract

I describe an adaptive trial algorithm that jointly learns experimental participant preferences and treatment effects across multiple outcomes. By aggregating treatment effects using estimated marginal rates of substitution I can adaptively assign participants to treatment arms to maximise *welfare* and not just some feature of the outcome vector. Finally, by randomising participants into their preferred choice or alternative choices I directly observe both the average treatment effect on the treated (ATT) and average treatment effect on the untreated (ATU) vectors which cannot usually be uncovered by conventional randomised control trials.

1. Introduction

Field experiments in development economics often observe multiple, common outcomes across treatment arms and must identify an optimal arm for a policymaker. However, without knowledge of individual's preferences they cannot map changes in outcomes to changes in welfare. Therefore, economists using adaptive trials typically maximise a single outcome or some standardised index. Taking such a decision seriously implies the economist's imposed utility function reflects participants' preferences and these preferences only depend on one feature of the outcome vector. Alternatively, using a standardised index implies participants value the marginal value of a variance weighted good equally across outcomes[1]. Instead, I propose an algorithm that jointly learns participant preferences and treatment arm effectiveness. Aggregating across the two leads to estimates of posterior *welfare* which can be used to adaptively assign participants to the welfare-optimal arm.

Moving from experimenter's preferences to trial participants' introduces several complications. A rich literature in economics, pyschology, and experimental fields more generally demonstrate the importance of designing mechanisms such that participants are incentivised to reveal their true preferences (Savage, 1971; Delavande, 2014). When the decision maker and experimenter are one and the same, incentives are clearly aligned and stated preferences can be treated as the ground truth. However, when the experimenter must learn participant preferences he/she must account for the possibility of strategic behaviour and "cheap talk". Therefore, my work differs from Lin et al. (2022) and methods outlined by (Furnkranz and Hullermeier, 2010) by proposing an incentive compatible preference elicitation step. I incentivise participants to tell the truth by allowing individuals to rank treatment arms and assigning individuals to arms with a probability concordant with their stated preferences.

Finally, this paper seeks to further embed the principles of the Belmont report in randomised control trials run by development economists and other researchers. By enshrining participant preferences at the center of randomised trials, my proposed algorithm speaks directly to respect for persons; beneficence; and justice as outlined by the report. Respect for participants and their preferences, who in development economics often reside in low-income countries, is particularly important in a field dominated by rich, US-based academics

1. For a static example see Ashraf et al. (2010); Blattman et al. (2017); Bandiera et al. (2017)

(Stansbury and Schultz, 2022) running trials on those with little agency, income, or human capital.

2. Setup and Method

The experimenter faces a vector of outcomes $Y_i = \begin{bmatrix} y_i^1 & y_i^2 & ... & y_i^j \end{bmatrix}'$ for individual i and must determine the optimal treatment arm k with associated J-length reward vector $\boldsymbol{\mu}_k = \begin{bmatrix} \mu_k^1 & \mu_k^2 & ... & \mu_k^j \end{bmatrix}'$. Individuals arrive in waves of size N_t. Participant's utility functions are parametrised using McFadden (1973)'s discrete choice random utility model[2]: $U_k = V_k + \varepsilon_k$ where U_k corresponds to the utility an individual receives from treatment arm k with $V_k = \gamma_1 \mu_k^1 + \gamma_2 \mu_k^2 + ... + \gamma_j \mu_k^j = \boldsymbol{\mu}_k \boldsymbol{\gamma}'$.

Algorithm 1 Treatment and structural participant preference estimation

Generate a prior (Q_0, F_0, Π_0) over $\left(\boldsymbol{\mu}_k, \boldsymbol{\gamma}, (\boldsymbol{\mu}_0, \boldsymbol{\tau}_0^{-1})\right)$
for $t = 1, ..., T$ **do**
  if $e_i < \alpha_t, e_i \sim U(0,1)$ **then**
    Elicit participant priors, $\boldsymbol{\mu}_0, \boldsymbol{\tau}_0^{-1}$, using BRS and update Π_t
  end if
  Sample $\boldsymbol{\nu}_t \sim Q_{t-1}(\cdot | p_k^{t-1}, Y^{t-1})$ and inform each participant of a single $\mu_k(\nu_t)$ draw
  Observe participant rankings, K_t, and update $F_t(\cdot | K^t, \boldsymbol{\mu}_k(\boldsymbol{\nu}^t))$ given Π_t
  Sample $\boldsymbol{\omega}_t \sim F_t(\cdot | K^t, \boldsymbol{\mu}_k(\boldsymbol{\nu}^t)), \boldsymbol{\nu}_t \sim Q_{t-1}(\cdot | p_k^{t-1}, Y^{t-1}), \boldsymbol{u} \sim U(0,1)$
  Choose $p_k = \frac{1}{N_t} \sum_{n=1}^{N_t} \mathbb{I}\{\boldsymbol{\mu}_k(\nu_t)\boldsymbol{\gamma}(\omega_t)' > \boldsymbol{\mu}_l(\nu_t)\boldsymbol{\gamma}(\omega_t)'\}, k \neq l$
  Assign participants to treatment arm k with probability p_k using a strategy proof mechanism
  Observe Y_t and update the posterior $Q_t(\cdot | p_k^{t-1}, Y^t)$ over $\boldsymbol{\mu}_k$.
end for

Each wave, the experimenter chooses a subsample to elicit participants' priors over treatment arm effectiveness and samples N_t draws, or *signals,* from the joint posterior of treatment arm effects. Next, the experimenter individually informs participants of a set of menus over expected outcomes, comprised of the private signals drawn previously, and asks individuals to rank menus. Estimating a rank-ordered discrete choice model of rankings on signals and normalising estimated coefficients by the first signal coefficient gives the marginal rate of substitution *across signals about outcomes.* Unfortunately, this complicates identification somewhat as we must disentangle how much an individual values an additional unit of an outcome from their private information, i.e. how sceptical they are about outcome signals. Estimating an auxillary model using the prior subsample allows the experimenter to separate prior scepticism from preferences over outcomes.[3]

Incentive compatibility is ensured by using a strategy proof mechanism with probability of arm assignment increasing in participant rankings. One example would be the *random serial dictatorship* mechanism whereby participants are randomly ordered from 1 to N_t, assign the first participant their first choice, the next participant their top choice amongst the remaining choices, and so on. Each treatment arm accepts remaining participants until

2. Any non/semi/fully-parametric choice model could be used here.
3. In the interest of brevity, I skip details of this procedure for the extended abstract. Alternatively, the experimenter may elicit posterior beliefs directly using a binarised scoring rule, although this cost may be prohibitive in many field settings.

   Finally, it may be possible to discretize the unobserved private information using subsequent observed outcomes and identify posteriors over outcomes using moment conditions derived from participants' rankings.

their assignment proportion, p_k, is reached. With rankings and participant posteriors in hand the experimenter estimates a rank-ordered logit, updating F_t.

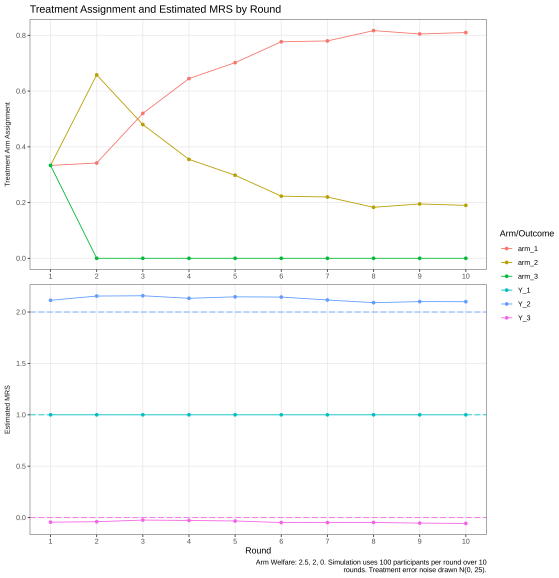
To calculate assignment probabilities the experimenter draws from the treatment effect, Q_{t-1}, and discrete choice model posterior, F_t, to generate $\boldsymbol{\mu}_k(\nu_t), \boldsymbol{\gamma}(\omega_t)$ draws. Taking the linear combination of these draws, $\boldsymbol{\mu}_k(\nu_t)\boldsymbol{\gamma}(\omega_t)'$, gives posterior arm welfare and p_k is chosen using probability matching in proportion to the probability an arm's welfare is highest. Finally, the experimenter assigns participants to treatment arms, observes Y_t and updates their treatment effect posterior, Q_t. By jointly estimating the treatment and preference posterior, the modified Thompson sampling algorithm balances the exploitation-exploration tradeoff by assigning participants to treatment arms reflecting uncertainty across both arm effectiveness and preference uncertainty.

3. A Simulated Example

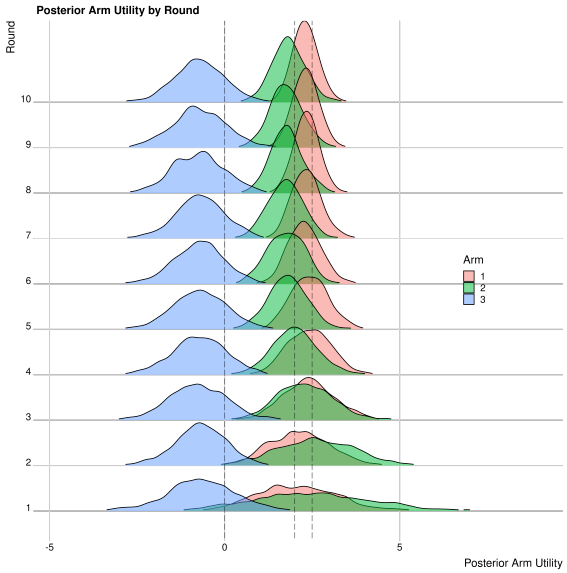
Suppose we face three treatment arms and measure three outcomes for each participant. Participants enter the trial in batches of 100 over 10 rounds. Let $\mu_1 = \begin{bmatrix} 1 & 2 & -3 \end{bmatrix}, \mu_2 = \begin{bmatrix} 2 & 1 & -3 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 & 0 & 3 \end{bmatrix}$. That is, the first treatment arm has treatment effect one for Y_1, two for Y_2, and negative three for Y_3. The second treatment arm has treatment effect two for Y_1, one for Y_2, and negative three for Y_3. The final treatment arm has no effect apart from increasing Y_3 by three units. Let the population MRS across outcomes be given by γ below. Whilst μ_1 and μ_2 are just permutations of each other, participants would prefer to receive treatment two since they value Y_2 twice as much as Y_1. In fact, as the extreme case of arm three demonstrates, even with such a large increase in outcome three, this arm is dominated by the other arms since consumers don't value Y_3 at all. Arm utility is given by $\mathcal{V}$:

$$\boldsymbol{\mu} = \begin{pmatrix} 1 & 2 & -3 \\ 2 & 1 & -3 \\ 0 & 0 & 3 \end{pmatrix}, \gamma = \begin{pmatrix} 1/2 & 1 & 0 \end{pmatrix}$$
$$\mathcal{V} = \boldsymbol{\mu}\gamma' = \begin{pmatrix} 2.5 & 2 & 0 \end{pmatrix}'$$

Figure 1a shows assignment proportions and estimated marginal rates of substitution after each round of a simulated bandit problem. Individual errors are drawn $N(0, 5^2)$. The algorithm quickly uncovers preferences across outcomes, shown in the lower left panel. In the upper left panel, the third treatment arm, which corresponds to 0 "welfare", is swiftly ignored. After a few more rounds arm three, which only returns four-fifths the utility of the first arm, is no longer played as frequently. In Figure 1b I plot posterior arm utility over rounds. Whilst uncertain, it's clear even by round one that arm three is sub-optimal. As more participants are observed in arms one and two posterior utility becomes less uncertain and arm one starts to pull ahead of arm two.



(a) Treatment Assignment and Posterior Median Marginal Rates of Substitution



(b) Posterior Utility Over Time

Figure 1: A Simulated Example Using Algorithm 1

4. Monte-Carlo Simulation Results

Table 1 shows results from 100 Monte Carlo draws using 15 rounds of 100 participants per wave with four treatment arms and three outcomes to aggregate across. Simulation parameters are drawn from:

$$\boldsymbol{\gamma} \sim N(\mathbf{0}, I_3), \boldsymbol{\mu}_k \sim N(\mathbf{0}, I_3), \ k = 1, ..., 4$$
$$\eta_i \sim N(0, 1), \ \varepsilon_{ki} \sim T1EV$$

where η_i, ε_{ki} represent participant-level outcome and ranking errors respectively. Models are estimated in Stan (Carpenter et al., 2017).

Table 1: Monte-Carlo Results

Assignment Type	Pr(Optimal Arm)	Mean Welfare Rank
Estimated	0.95	1.28
Random Assignment	0.87	3.06
Equal	0.39	2.74
First	0.30	2.92

Assignment type "Estimated" corresponds to Algorithm 1 outlined above and identifies the optimal arm by the end of the trial 95% of the time. In contrast, static random assignment only identifes the optimal arm in 87% of draws. "Equal" corresponds to Thompson sampling maximising a standardised index of the three outcomes whilst "First" only targets the first element of the outcome vector to maximise. Since I use a closed form solution for participant utility, using the multinomial logit and generated $\boldsymbol{\gamma}$ parameters, I calculate average welfare across participants within a simulated draw and rank each algorithm, denoted by "Mean Welfare Rank". As expected, Algorithm 1, which estimates participant preferences directly produces the greatest mean welfare for participants whilst static random assignment the lowest.

5. Conclusion

By carefulling incentivising research trial participants, I've shown how to estimate participant preferences and aggregate treatment effects across disparate outcomes to estimate of treatment arm welfare. By adaptively assigning participants using posterior welfare, rather than posterior outcomes, researchers can conduct adaptive trials that identify the optimal arm whilst maximising the welfare of participants using their own loss function and not one imposed by the economist. Simulation results show the experimental design and adaptive algorithm are able to detect the optimal arm more often, and in-sample welfare is higher, compared with conventional alternatives. A small MTurk trial is currently in pilot.

References

Nava Ashraf, Dean Karlan, and Wesley Yin. Female empowerment: Impact of a commitment savings product in the philippines. *World Development*, 38(3):333344, 2010. doi: ISSN0305-750X.doi. URL https://doi.org/10.1016/j.worlddev.2009.05.010.

Oriana Bandiera, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. Labor markets and poverty in village economies*. *The Quarterly Journal of Economics*, 132(2):811870, 2017. doi: doi:10.1093/qje/qjx003. URL https://doi.org/10.1093/qje/qjx003.

Christopher Blattman, Julian C. Jamison, and Margaret Sheridan. Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in liberia. *American Economic Review*, 107(4):11651206, 2017. doi: doi:10.1257/aer.20150503. URL https://www.aeaweb.org/articles?id=10.1257/aer.20150503.

Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76(1), 2017.

Adeline Delavande. Probabilistic expectations in developing countries. *Annual Review of Economics*, 6(1):1–20, 2014. doi: 10.1146/annurev-economics-072413-105148. URL https://doi.org/10.1146/annurev-economics-072413-105148.

Johannes Furnkranz and Eyke Hullermeier. Preference learning and ranking by pairwise comparison. In *Preference learning*, page 6582. Springer, 2010.

Zhiyuan Jerry Lin, Raul Astudillo, Peter I. Frazier, and Eytan Bakshy. Preference exploration for efficient bayesian optimization with multiple outcomes, 2022.

Daniel McFadden. Conditional logit analysis of qualitative choice behavior. 1973.

Leonard Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971. doi: 10.1080/01621459.1971.10482346.

Anna Stansbury and Robert Schultz. Socioeconomic diversity of economics phds, 2022. URL https://ssrn.com/abstract=4068831.