What really matters (and to whom): Investigating heterogeneous effects in high-dimensional conjoint analyses

Thomas Reiter Philipp Sterner

Introduction

In the social sciences, conjoint analyses are a powerful research design to assess preferences in humans. Under the popular special case of *forced choice designs*, participants are presented with two options of choice form which they have to choose one, for example certain products in marketing research, or two potential candidates in political research on voting behavior. The two possible options are characterized by certain *attributes* (also: treatments), which can have different *attribute levels* (treatment levels), for example the attribute color of a car might have levels white, black, and blue. By randomly mixing the attribute levels and "forcing" participants to choose one of two options across multiple rounds, researchers are able to estimate the relative importance that participants give to certain attribute levels.

Arguably the most often estimated causal quantity in conjoint analyses is the *average marginal component effect* (AMCE). The AMCE is the effect of a specific attribute level of interest compared to another level of the same attribute, holding equal the joint distribution of all other attributes and averaged over this joint distribution as well as the sampling distribution from the population CITE. This quantity is estimable by regression models and provides us an effect of an attribute on the probability that a certain option (or profile) will be chosen. As the name AMCE suggestes, this estimate is an average. That is, potential differences in effects among participants cannot be considered directly. However, this *effect heterogeneity* is often of interest, when trying to determine whether an attribute is equally important to all participants regardless of background covariates (e.g., age, gender, education, etc.) or whether there are subgroup differences. Ultimately, it is a question of generalisation of results to the entire population and, even more advanced, of causal mechanisms in choice behavior.

The simplest approach to take into account potential effect heterogeneity in conjoint analyses is to estimate the AMCE-models in different subgroups, for example a model for all male and female participants. Altough appealing due to its simplicity, this approach quickly reaches its limits and has additional downsides. First, it is not clear how to a priori define the subgroups let alone how and where to include certain interactions. If a continuous covariate age is included in the data, arbitrary splits into distinct groups are necessary. Potential interaction effects would have to be included manually with the likely chance to miss important ones. Needless to say, sample sizes in the subgroups are always smaller than in the whole sample as well, increasing the uncertainty in estimations. This leads to a second issue: conjoint data can quickly become high-dimensional due to many different pairings of different attribute levels (across attributes). Consequently, there are more combinations of attribute levels (and thus, possible options) that should be presented to participants than there are actual rounds of choosing per participant. This leads to some kind of "double high-dimensionality issue": there are many different attribute (level) combinations *and* many potential covariates characterizing the participants. This results in a number of possible interactions that are untractable for classical methods (these interactions correspond to the effect heterogeneity mentioned above).

New Approaches to Handle Effect Heterogeneity under High-Dimensionality

To remedy the downsides of the "naive" subgroup approach outline above, researchers developed more advanced methods based on machine learning. Most notable in this regard are (a) an approach based on *Bayesian additive regression trees* (in the following called *cjbart* CITE), (b) an approach based on Bayesian mixture of regularized logistic regressions (*FactorHet* CITE), and (c) a testing approach based on conditional randomization tests (*CRT* CITE). In this article, we focus on the explanation and demonstration of cjbart. At the end of this section, we briefly delineate similarities and differences in these three approaches.

cjbart

The premise of cjbart as introduced by CITE is that there are several *nested causal quantities* underlying the AMCE. Specifically, the AMCE can be decomposed into an individual-level (i.e., participant-level), round-level, and observation-level marginal component effect (IMCE, RMCE, and OMCE, respectively). Let us consider N individuals choosing between J profiles across K rounds (where in the simplest case of binary choices $J = 2$). In each round, an individual i is presented with J profiles in which the attribute levels of L attributes are assigned randomly. In the final data set, there are $N \times J \times K$ rows and $L + X$ columns (with X being the covariates characterizing the individuals). From these data we now want to estimate the causal parameters of interest, that is, AMCE, IMCE, RMCE, and OMCE. Nested in the AMCE described above, the IMCE is the change in probability that subject i choses a profile given a specific attribute level (compared to a reference category), again averaged over the effects of all other attributes. This corresponds to subgroup analyses of the AMCE with the convenient addition that the IMCE considers conditional effects based on all possible individual-level covariates. By inspecting the IMCE, effect heterogeneity due to non-randomized characterisitics (i.e., the covariates) can be identified. The IMCE further contains the two lower-level quantities RMCE and OMCE. Because for each participant, there are usually multiple rounds of observations (i.e., choices), the RMCE can be obtained as the effect of an attribute within a specific round k of the experiment for a given individual i. Lastly, the OMCE is estimated by additionally conditioning on a specific profile-level (i.e., if $J = 2$, by conditioning on one of the two profiles). As CITE note, the OMCE does not contain too much substantial information. It is, however, of statistical importance: assuming the OMCE is an independent random draw from an individual-level distribution, we can aggregate the OMCEs to estimate the IMCEs.

In the following, we explain how the IMCEs are estimated. IMCEs can be considered the most important quantity for our purpose of investigating effect heterogeneity because it allows us to analyze how attribute importances differ depending on participant covariates. CITE propose a three-step estimation procedure.

Step 1: Modeling Potential Heterogeneity

In a first step, potential effect heterogeneity is modeled. Specifically, some function is estimated that relates the attribute levels of the L attributes that were shown to subject i in the k^{th} round in profile j and the covariate vector X_i to the observed binary outcome Y_{ijk} (which is equal to 1 if the profile was chosen and equal to 0 if it was not chosen). CITE detail this estimation procedure using Bayesian additive regression trees (BART) but other appropriate models that can estimate this potentially complex functional relationship could be used as well. BART is a supervised learning model based on a boosting procedure: many small decision trees are trained subsequently, each one aiming to explain the residual variance of the outcome variable which was not explained by all previous trees. The "Bayesian part" in BART is that parameters are seen as random variables (instead of constants, as would be the case in frequentist approaches). An advantage of BART is its robustness to the choice of hyperparameters (e.g., the number of individual trees) but they could be tuned if necessary, for example by cross-validation. The data that is used to train the BART are the data resulting from the conjoint experiment; that is, the different choices of profiles as well as the covariates, which are invariant at the individual-level.

Step 2: Predicting Counterfactual Outcomes

In a second step, the estimated function from step 1 (i.e., the trained BART model) is used to predict counterfactual outcomes by changing the values of attribute levels. These predictions are counterfactual because they did not happen — but we assess what would have been chosen had a certain attribute level been set to a different attribute level of the same attribute. This is done repeatedly by drawing multiple times from a predicted posterior distribution, once with the altered attribute level and once with the reference level of the same attribute. By subtracting these results and averaging them over the multiple draws of the posterior, we arrive at a parameter estimate of the OMCE (i.e., observation-level effects).

Step 3: Calculating IMCEs

Following the nested causal quantities structure outlines above, IMCEs are calculated by averaging the OMCEs for each individual. Specifically, the OMCE estimates from step 2 are summed and divided by $J \times K$, that is, the number of profiles times the number of rounds (which is the number of total observations).

Identification of Heterogeneity

As mentioned, the main reason why we employ these complicated procedures is because we want to estimate heterogeneous treatment effect. More specifically, we want to assess whether the effects of certain attribute levels of choice behavior are different across individuals. CITE provide two methods that yield information about which covariates are associated with heterogeneity in effects of attribute levels.

EXPLAIN THREE STEP PROCEDURE OF ESTIMATING OMCES AND THEN IMCES WITH BART.

EXPLAIN HOW RFs ARE USED TO ESTIMATE VIMP

EXPLAIN HOW DTs ARE USED TO INVESTIGATE HETEROGENEITY IN IMCEs AND WHICH COVARIATES ARE DRIVERS

-Vorgeschlagene neue Modelle in der Literatur -cjBart -Idee des Modells/der Modelle -weitere Modelle: -factorHet + idee kurz -CRT + idee kurz Warum cjBart
das Modell stark machen

Data, Research Question, and Method

-Die Daten -Fragestellung -bisherige Forschung

Results

-Ergebnisse cjBart Ergebnisse

Integration of Results

-Einordnung der Ergebnisse -Vergleich mit factorHet -Vergleich mit "normaler" Subgruppen Analyse

Discussion

Discussion & Limitations -Tuning -Wie viele Covariates (inference vs. exploration) -Stabilität

Conclusion