What is a relevant control?: An algorithmic proposal

Fernando Delbianco[1,2][0000−0002−1560−2587] and Fernando Tohmé[1,2][0000−0003−2988−4519]

[1] Departamento de Economía, Universidad Nacional del Sur
[2] Instituto de Matemática de Bahía Blanca (INMABB), CONCET.
fernando.delbianco@uns.edu.ar
ftohme@criba.edu.ar

Abstract. Individualized inference (or prediction) is an approach to data analysis that is increasingly relevant thanks to the availability of large datasets. In this paper, we present an algorithm that starts by detecting the relevant observations for a given query. Further refinement of that subsample is obtained by selecting the ones with the largest Shapley values. The probability distribution over this selection allows to generate synthetic controls, which in turn can be used to generate a robust inference (or prediction). Data collected from repeating this procedure for different queries provides a deeper understanding of the general process that generates the data.

Keywords: Individualized inference · Relevance selection · Relevance classification · Synthetic controls.

1 Introduction

Sometimes, a statistical analysis requires only focusing on an observation or individual from which an unknown variable of interest can be inferred based just on its characteristics. In such cases, a critical issue faced by the analyst is whether all the available data (the actual sample) is equally relevant to the question at hand. Or to put it another way, there may exist a more efficient way to exploit the dataset to obtain a result that is both robust and relevant to that question. This idea led Liu and Meng (2016) to propose what they call *transitional inference*, striking a balance between finding a specific answer (ensuring *relevance*) satisfying desirable statistical properties (*robustness*).

This type of problem, in which it is required to infer or predict answers to individual queries, is an instance of what Vovk et al. (2022) deem as (one-off) *online learning*. Delbianco et al. (2021) and Delbianco and Tohmé (2023) presented some initial approaches to address the question of obtaining robust and relevant individualized inferences using an algorithmic approach. The latter contribution incorporated the idea of using *conformal intervals* (Xie and Singh (2013)) to

determine the reliability of individualized inferences.

The literature presents different ways of exploiting the idea of individualization by focusing on local data. Cai et al. (2021) and Alaa and Van Der Schaar (2017) exhibit different approaches to that problem. The former, closer to the traditional methods of Statistics, mixes the information drawn from instances similar to that of a given query; the latter contribution, in turn, presents a Machine Learning solution.

Different aspects of the problem of individualization are analyzed in the literature. So, for instance, Msaouel et al. (2022) proposes a framework for individualized medicine; Gelman and Hennig (2017) discusses, in a Bayesian setting, the question of subjective and objective aspects in the analysis of the trade-off between relevance and robustness in statistical inference; Duan et al. (2022) presents a radical approach of running completely individual inferences (N-of-1 trials). Lu et al. (2018) suggest using random forests to obtain answers from individualized treatments while Gong and Meng (2021) contextualizes the problem as a way of solving Simpson's paradox. While these and other contributions address the same question they attack it from a diversity of angles.

Our idea here is that a query splits the space of observations in two, namely those relevant to it and those that can be disregarded since they carry little information to answer it. This classification can be understood as being associated with a relevance metric for each individual, indicating how relevant is an observation in the online learning framework. This means that an observation has several weights, one referring to its general relevance and the others to its relevance for particular queries.

We present this approach by exploiting the concept of *Shapley value*, adapted from Cooperative Game Theory (Buckmann et al. (2021)) as well as the properties of *distance functions* (Deza and Deza (2013)). We generate then a notion of relevance for a sample which further allows us to obtain a classification and standardization of the data.

The rest of the work is structured as follows. Section 2 describes the motivation and the main ideas behind our proposal. Section 3 presents the methodology while in Section 4 we describe the actual algorithm. Section 5 assesses the classification task derived from the solution to the individualization problem. Section 6 analyzes possible extensions and upgrades to the algorithm. Section 7 concludes.

2 Motivation

While our long-term project is motivated by the original proposal of Liu and Meng, we focus here on the detection of relevant observations as a first step to

the generation of synthetic data. The latter would incorporate the main properties of the relevant observations as to allow the specification of scenarios in which robust predictions can be made while preserving the relevance of the data used to generate them.

We assume, as in Delbianco et al. (2021), a statistical model of a data generation process, which can be described as $\{\mathcal{O}, \mathcal{P}\}$, where $\mathcal{O}$ is the set of observations while $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is a family of probability distributions over $\mathcal{O}$ and Θ is the space of parameters of the model. The goal is again to estimate the parameters in response to any *query* q, where q is a specific request for information under a given inference method $\mathcal{I}$ applied on the database.

The query defines several dimensions. First, $\mathcal{O}$ consists of entries $\{o_i = \langle x_i, y_i \rangle\}_{i=1,\ldots,n}$, where x_i is a vector of p variables, the *tail* of o_i, while y_i is the *head*, a vector whose components may become potential answers to different queries. The query q consists of a *tail*, x_0, with no *head*, and with $x_0 \neq x_i$, for $i = 1, \ldots, n$. We assume that there exists a class of *latent variables* $\mathcal{S}$ such that given a query q there exists a corresponding $s_q \in \mathcal{S}$ yielding a class of *relevant* observations $\mathcal{O}_q \subseteq \mathcal{O}$ The use of controls from a subset of observations can be justified on the grounds of the Law of Large Populations. As pointed out in Meng (2018) it is relevant to have a large *quantity* of high *quality* data.

The relevant set of observations is given by $\mathcal{O}_q = f_{\mathcal{S}}(s_q)$, where $f_{\mathcal{S}} : \mathcal{S} \to 2^{\mathcal{O}}$ characterizes a *selection* procedure. A further step of refinement may yield a set $\mathbf{O}_q \subseteq \mathcal{O}_q$, representing *strongly* representative observations. We then can generate a class of controls $\bar{\mathcal{O}}_q$ verifying that $\mathbf{O}_q \subseteq \bar{\mathcal{O}}_q$. Based on $\bar{\mathcal{O}}_q$ the application of an inference procedure $\mathcal{I}$ yields θ^q such that $\bar{\mathcal{O}}_q$ can be understood as a set of draws from a distribution $\bar{P}_{\theta^q} \in \mathcal{P}$ (Also known as *Transitional Inference*, as in Li and Meng (2021)).

As an example consider the case of n different economies, each one described by a vector of a few macroeconomic variables $o_i = \langle x_i, y_i \rangle$, where y_i is the GDP of country i. We can ask, for any given economy $o' \notin \mathcal{O}$ what is its expected GDP, knowing only x_0. If we assume that a latent variable is the productivity of the leading sectors of an economy, we can group all the countries in $\mathcal{O}$ of productivity similar to that of o' to generate a class of controls to answer the question.

For another example, can consider a dataset of programs of study, students, and their grades. Given a new student, a relevant question is how long it will take to complete the chosen academic program. Another question could be to assess the probability that she will drop out of her studies. Of course, a different query may require different controls. So the s_q latent variable will yield the portion of observations that is relevant given q and will let us obtain $\mathcal{O}_q$ to simulate new

controls and make a robust and relevant inference to answer q.

For each q we need to specify latent variables relevant for the query. Then, we have to distinguish its scope using some proxy or measurement on the available $\mathcal{O}$. This will yield the class of relevant observations $\mathcal{O}_q$. In the particular example of the student, let us assume that there are two types of students, associated with a latent variable (which can capture, for instance, socioeconomic or cognitive advantages). Then, a query q about a particular student will be associated with her corresponding type and according to $f_{\mathcal{S}}$, mapped to the class of observations of students of that type.

But once obtained the relevant observations, a *robust* inference requires generating new controls, not present in $\mathcal{O}_q$. This is achieved by creating pairs $\langle x, y\rangle$ similar to those in $\mathcal{O}_q$ but without assuming that they share with the latter a common value of the latent variable. This means that only the *observable* features of the entries $o \in \mathcal{O}_q$ must be used to create fictitious controls in $\bar{\mathcal{O}}_q$, for instance as in the examples of Section 5 of Delbianco et al. (2021).

This problem can be depicted as in Figures 1 y 2 (drawn from Delbianco and Tohmé (2023)). Figure 1 represents the existence of a set of observations relevant to a query q without specifying a particular answer y. In low-dimensional cases, the determination of $\mathcal{O}_q$ seems rather trivial, since it can be obtained by clustering the observations that are at a rather short distance according to, say, classical metrics like the cosine distance. With a larger number of variables, non-classical notions of distance may be more appropriate. Then, as shown in Figure 2, the set of relevant observations can be enlarged by simulating new ones according to the empirical distribution of $\mathcal{O}_q$. This is known in Computer Science as the creation of *synthetic controls* (Emam et al., 2020) and as the generation of *repro samples* in the field of conformal prediction (Xie and Wang, 2022).

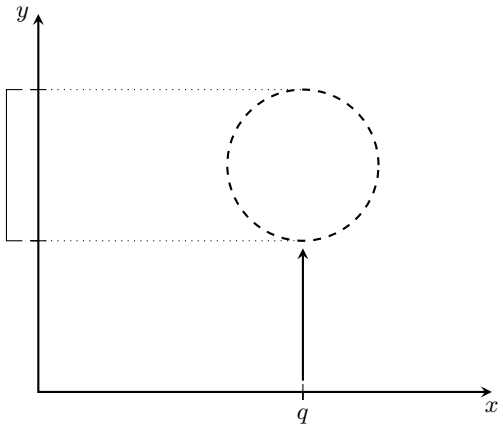
3 Methodology

Our methodology ensues from the following result Delbianco et al. (2021):

Theorem 1 *There exists $\mathbf{\Phi} : \Theta \to \mathcal{O}$ such that for any query q with parameter θ_q it returns $\mathcal{O}_q \subseteq \mathcal{O}$ and a function $\nu_q : \mathcal{O} \to \{0, 1\}\}$ yielding the minimal expected value $\Delta_{\theta^q} = E(\bar{P}_{\theta^q}) - E(P_{\theta^q})$ according to ν_q.*

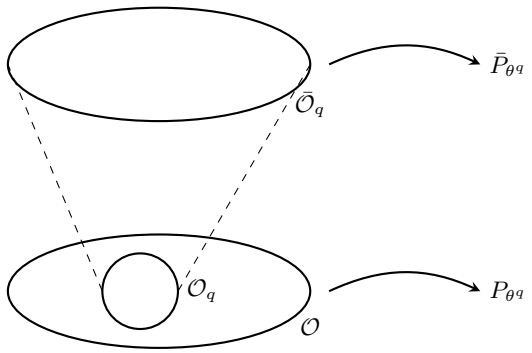
To implement $\mathbf{\Phi}$ we go through different stages. The first one is the definition of a *metric* that will allow us to define a distance between x_0 and the tails of the observations in $\mathcal{O}$. The selected observations yield a particular model $y = f_q(x)$. The contribution of each observation to this model is captured by its *Shapley value*. We can then regress a restricted loss function against the Shapley values of the observations in $\mathcal{O}_q$, to keep only the significant observations. This ensures that only those that have a weight that matters for accurate predictions. The

Fig. 1: Latent relevant area S_q



A particular value of a tail q, is associated with a latent variable s_q that yields the relevant space of tails $x \in X_q$ and as a consequence allows to infer the corresponding class of heads $y \in Y_q$. Then, this bi-dimensional figure can be extended to a three-dimensional one, with the sequence of different queries as a third axis. This represents how the relevant set varies according to q, learning the associated θ^q.

Fig. 2: Simulated controls $\in \bar{\mathcal{O}}_q$



The relevant set $\mathcal{O}_q$, detected by means of $f_{\mathcal{S}}$, is extended to an enlarged set, $\bar{\mathcal{O}}_q$. The difference in the inferences made with and without the new controls is $\Delta_{\theta^q} = E(\bar{P}_{\theta^q}) - E(P_{\theta^q})$.

following subsections present in detail these stages.

3.1 Relevant distance

We want to find, given a query q (identified with a tail x_0) a subset $\mathcal{O}_q \subseteq \mathcal{O}$. Since the definition of $\mathcal{O}_q$ might not be exact, we need to define a distance μ_q such that $\mu_q(x_0, x_i) > 0$ for each x_i corresponding to an observation $o_i = \langle x_i, y_i \rangle \in \mathcal{O}_q$. A possibility is to identify μ_q with a *fuzzy T-metric* μ, such that $\mu_q(x_0, x_i) = \mu(x_0, x_i, \alpha)$ for a given positive $\alpha \in \mathbb{R}$.

A *fuzzy T-metric* in our case would be a function

$$\mu : X \times X \times \mathbb{R}_+ \to [0, 1] \tag{1}$$

where X is a space of p dimensions in which the tails of the observations (and the queries) live. μ satisfies the following axioms for all $x, y, z \in X$, $\alpha, \beta > 0$:

- $\mu(x, y, \alpha) \in (0, 1]$,
- $\mu(x, y, \alpha) = 1$ iff $x = y$,
- $\mu(x, y, \alpha) = \mu(y, x, \alpha)$,
- $T(\mu(x, y, \alpha), \mu(y, z, \beta)) \leq \mu(x, z, \alpha + \beta)$,
- $\mu : \mathbb{R}_{++} \to [0, 1]$ is a continuous function.

Here $T : [0, 1]^2 \to [0, 1]$ is a *continuous triangular norm* satisfying that for all $a, b, c, d \in [0, 1]$:

- $T(a, 1) = a$ (1 is the identity),
- If $a \leq b$ and $c \leq d$, then $T(a, c) \leq T(b, d)$ (monotonicity),
- $T(a, b) = T(b, a)$ (commutativity),
- $T(a, T(b, c)) = T(T(a, b), c)$ (associativity).

An example of norm T is the *Gödel norm*: $T(a, b) = \min\{a, b\}$. Under this norm a rather natural characterization of a possible distance μ_q obtains by, fixing α, as follows Romaguera (2022):

$$\mu_q(x_0, x_i) = \mu(x_0, x_i, \alpha) = \begin{cases} 1, & \text{if } d(x_0, x_i) \leq \alpha \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where $d(\cdot, \cdot)$ is a standard metric on X. Then, $\mathcal{O}_q = \{o_i = \langle x_i, y_i \rangle \in \mathcal{O} : \mu_q(x_0, x_i) = 1\}$.

3.2 Shapley Value

Consider a predictive model $y = f(x)$ computed on $\mathcal{O}_q$. A way of assessing the contribution of an individual $o_k = \langle x_k, y_k \rangle$ to the prediction $f(x_0)$ for the query q. This can be obtained by using this version of the Shapley value for Machine Learning Buckmann et al. (2021):

$$\phi_k(f, x_0) = \sum_{S \subseteq \mathcal{O}_q \setminus \{o_k\}} \frac{|S|!(|\mathcal{O}_q| - |S| - 1)!}{|\mathcal{O}_q|!} (f(x_0 | S \cup \{x_k\}) - f(x_0 | S)) \tag{3}$$

where S is a subset of $\mathcal{O}_q$ *without* o_k. Then, two new models are computed, $f(\cdot|S \cup \{o_k\})$ and $f(\cdot|S)$, using only the observations in S plus o_k and those in S, respectively. The computed Shapley values $\{\phi_k(f, x_0)\}_{o_k \in \mathcal{O}_q}$ must satisfy a normalization condition:

$$\phi_0(f, \emptyset) + \sum_{k=1}^{n} \phi_k(f, x_0) = f(x_0) \tag{4}$$

If we assume that the Shapley values of observations in $\mathcal{O}_q$ are inversely proportional to the distances between their tails and x_0, an algorithm Fatima et al. (2008) of complexity $\mathbf{O}(n)$ approximates to the contribution of each o_k, where $n = |\mathcal{O}_q|$.

3.3 Selecting observations in $\mathcal{O}_q$

Now let us consider a restricted loss function over $\mathcal{O}_q$, namely $L(o_k) = 1$ if $|(y_k - f(x_k))| \leq \gamma$, for a given $\gamma \geq 0$ and $L(o_k) = 0$, otherwise.

Then, we consider the model

$$L(o_k) = \beta_0 + \sum_{o_j \in \mathcal{O}_q} \beta_j \phi_j(f, x_0) + \beta_x x_k + \epsilon_k \tag{5}$$

Now we consider the null hypothesis of the non-positive weights of the Shapley values:

$$H_0 : \{\beta_j \leq 0 \text{ for } o_j \in \mathcal{O}_q\} \tag{6}$$

The o_js that are significant are the only ones we keep in a restricted $\mathbf{O}_q \subseteq \mathcal{O}_q$.

3.4 Estimation of the distribution over $\mathbf{O}_q$

To generate the distribution that might approximate P_{θ^q} we start by applying the **EDAs** (*Estimation of Distribution Algorithms*) approach. This involves applying stochastic optimization methods to build and sample explicit probabilistic models of promising candidates to be the actual model for a given sample. A series of incremental updates of a probabilistic model obtain by optimization, starting with the model encoding an uninformative prior over admissible solutions and ending with a model that generates only the global optima.

The main difference between EDAs and most conventional evolutionary algorithms is that EDAs use an *explicit* probability distribution. The quality of candidate solutions is evaluated using one or more objective functions.

Our EDA proceeds as follows:

```
t := 0
initialize model M(0) to represent an uniform distribution over O_q
while (termination criteria not met) do
P := generate N > 0 candidate solutions by sampling M(t)
F := evaluate all candidate solutions in P
M(t+1) := adjust_model(P, F, M(t))
t := t + 1
```

The evaluation step F is critical. We accept a new observation o' sampled from a distribution $M(t)$ if $L(o') = 1$. In turn, the termination criterion would be satisfied if the N candidates in P are accepted under F. The resulting set of observations is $\bar{\mathcal{O}}_q$ and the corresponding distribution is $\bar{P}_{\theta^q}$, which is used to infer the actual head y_0 of the query q with tail x_0.

4 Algorithm

We will now describe how the previous procedures are integrated into an algorithm to compute the answer to a query q. The sequence is as follows:

1. *Query*: The first step involves defining the exact specification of the query q. It is constituted by a *tail* x_0, which can be conceived as a vector with as many dimensions as explanatory variables of an (unknown) head y_0 that might also be multi-dimensional. The entire set of observations is used to find a model $y = \mathbf{f}(x)$ and compute $\mathbf{f}(x_0)$.
2. *Neighbourhood*: The computational cost of determining a sample of *relevant* observations requires the application of a simple procedure to detect appropriate candidates. The fuzzy metric defined above allows to select the observations that are closer than a given α, selected according to the precision demanded by the researcher. The more information is applied in this selection, the more precisely defined is $\mathcal{O}_q$.
3. *Shapley value*: We compute the class of Shaple values $\{\phi_k(f, x_0)\}_{o_k \in \mathcal{O}_q}$.
4. *Regression*: The loss function $L(\cdot)$ is regressed on the Shapley values to find their corresponding weights. The observations with weights that do not reject the null hypothesis, are discarded. The remaining observations constitute a set $\mathbf{O}_q$. Then we can identify $\{\phi_j(f, x_0)\}_{o_j \in \mathbf{O}_k}$ with the *latent variable* s_q.
5. *Estimation*: Applying the **EDA** described above, we generate a distribution over $\mathbf{O}_q$.
6. *Simulation*: New observations are drawn from the distribution over $\mathbf{O}_q$, generating a class of observations $\bar{\mathcal{O}}_q$.
7. *Inference*: A model $y = \mathbf{f}_q(x)$ is inferred on the basis of the data in $\bar{\mathcal{O}}_q$. Then, the answer to q is $\hat{\mathbf{f}}(x_0)$.
8. *Comparison*: We define $G_q = |\mathbf{f}(x_0) - \mathbf{f}_q(x_0)|$, the gap between the inference with *all* the observations and the inference with the relevant ones. Different choices in the previous steps may lead to different values of G_q.

9. *Iteration*: Different queries $q_1, q_2, \ldots, q_t$ ($t = \dim(Y)$) based on x_0 lead to values G_{q_j}, $j = 1, \ldots, t$. Then $\bar{G} = \frac{\sum_{j=1}^{t}}{t}$ can be used to infer the average difference between the answers over specific queries and all the set of observations.

Algorithm 1 Individualized Prediction with Relevant Synthetic Controls

Input: $\mathcal{O} = \{\langle x, y\rangle, x \in X, y \in Y\}$, $X_0 \subseteq X$, Regression method $\mathcal{A}$, Similarity intensity α
Output: $\hat{y}_0$ for each element $x_0 \in X_0$ of q, $\mathcal{O}_q$, $\mathbf{O}_q$, $\bar{\mathcal{O}}_q$
$\alpha \in [0, 1]$
$\mathcal{A}$ ▷ Regression method (e.g. OLS)
$\mathcal{L}$ ▷ Loss function (e.g. $L(o_k)$)
$\mathcal{T}$ ▷ Fuzzy T-metric (e.g. *Gödel norm*)
$\mathcal{E}$ ▷ Search algorithm (e.g. EDA)
for each $x \in X_0$ **do**
1 Predict y_0 with $\mathcal{A}$ and X
2 Use α and $\mathcal{T} \rightarrow \mathcal{O}_q$
3 Use $\mathcal{O}_q \rightarrow$ *Shapley values* in $\mathcal{O}_q$
4 Regress $\mathcal{L}$ on the *Shapley values* $\rightarrow X_{rel}$ that reject the null hypothesis H_0
5 Use $\mathcal{E}$ on $X_{rel} \rightarrow \bar{P}_{\theta^q}$
6 Generate synthetic data based on $\bar{P}_{\theta^q} \rightarrow \bar{\mathcal{O}}_q$
7 Predict y_0 with $\bar{\mathcal{O}}_q \rightarrow \hat{y}_0$
8 Compare steps **1** and **7** $\rightarrow$ relevant gain G_q
end for each $x_0 \in X_0$
Return: Y_0, $\bar{G}$

5 Classification

Repeating the procedure for different queries with a common x_0, makes our procedure an instance of online learning, instead of just a one-off one. The information gained by the different rounds provides a classification of the corresponding sets of relevant observations.

An interesting question is to consider as many queries as pairs $\langle x_i, y_i^l\rangle$, where $i = 1, \ldots, n$ and $l = 1, \ldots, \dim(Y)$ to see the proportion of cases in which a given observation $\langle x_i, y_i\rangle \in \mathcal{O}$ is chosen as relevant. This proportion is $\frac{\bar{s}_i n_i}{\dim(Y)-1}$, where n_i is the number of times the observation is deemed relevant for different queries, weighted by a representative Shapley value, $\bar{s}_i$. The representative $\bar{s}_i$ can be either a simple or weighted average, the median, or any other way of deriving a value from the Shapley values corresponding to the different queries.

6 Room for improvement

Since our algorithm is modular, each of its components may be replaced by an alternative procedure. Here we discuss some of them:

- **Bayesian updating**: The Dirichlet distribution $Dir(\alpha)$ is a family of continuous multivariate probability distributions parameterized by a vector α. $Dir(\alpha)$ is a multivariate generalization of the *Beta* distribution. It is commonly used as a *prior distribution* in Bayesian statistics. This is because it is the *conjugate prior* (i.e. $p(\theta|x)$ belongs to the same family of distributions as $p(\theta)$) to two important probability distributions: the *categorical* distribution and the *multinomial* distribution.

  Then, if $\theta \sim Dir(\alpha)$, where $\alpha = (\alpha_1, \ldots, \alpha_K)$ and $\theta = (\theta_1, \ldots, \theta_K)$, we have that

  $$Dir(\theta|\alpha) = \frac{1}{Beta(\alpha)} \prod_{i=1}^{K} \theta_i^{\alpha_i - 1} \tag{7}$$

  Starting with an *a priori* distribution over relevant observations, a Bayesian sequence (for the same x_0 but different queries) of *a posteriori* distributions may simplify the generation of answers to the queries. Using Dirichlet distributions reduces the burden of determining the *a posteriori* distributions, each one becoming then the *a priori* one for the next query. While the answers will no longer be independent of each other, this may be helpful in the cases in which the variables in the heads y are correlated.
- **Discriminating variables**: Another alternative is to define which variables in the tails must be taken into account to determine the relevant observations. Since some of the variables may not contribute to distinguishing whether an observation is close (for a given query) to x_0. A previous step to the selection of the observations in $\mathbf{O}_q$ could be to run a LASSO regression using the observations in $\mathcal{O}_q$ to detect the variables contributing to the answer to reduce the set X and consequently the distance μ_q.

7 Conclusions

An individualized inference algorithm provides a way of learning accurately the differences among different observations in a dataset. Each step records those that are relevant according to a metric of similarity. Furthermore, it registers the Shapley values of the observations that pass a significance test among those that are deemed relevant.

The usefulness of this procedure is straightforward: classifying relevant observations contributes to detecting the specific elements that may contribute to yielding the right answers to queries. Further questions like, which query or category of problems is relevant to a given observation, can be answered by this

one-off learning procedure. The different parameters that are to be chosen can be found optimally. The procedure is transparent, avoiding the risks associated with black box ones. Then, besides yielding answers to queries, the procedure learns about specific cases, recording the inferences for future use on new cases.

Acknowledgments. This research was funded by a research project funded by the Universidad Nacional del Sur (PGI 24-E172).

Disclosure of Interests. We have no conflicts of interest to disclose.

Bibliography

Alaa, A. M. and Van Der Schaar, M. (2017). Bayesian inference of individualized treatment effects using multi-task gaussian processes. *Advances in neural information processing systems*, 30.

Buckmann, M., Joseph, A., and Robertson, H. (2021). Opening the black box: Machine learning interpretability and inference tools with an application to economic forecasting. In *Data Science for Economics and Finance: Methodologies and Applications*, pages 43–63. Springer International Publishing Cham.

Cai, C., Chen, R., and Xie, M.-g. (2021). Individualized group learning. *Journal of the American Statistical Association*, pages 1–17.

Delbianco, F., Fioriti, A., and Tohmé, F. (2021). A methodology to answer to individual queries: finding relevant and robust controls. *Behaviormetrika*, 48(2):259–282.

Delbianco, F. and Tohmé, F. (2023). Individualized conformal prediction: Using synthetic data as relevant controls. *Available at SSRN 4506759*.

Deza, M. and Deza, E. (2013). *Encyclopedia of Distances*. Berlin: Springer-Verlag.

Duan, N., Norman, D., Schmid, C., Sim, I., and Kravitz, R. L. (2022). Personalized data science and personalized (n-of-1) trials: Promising paradigms for individualized health care.

Emam, K. e., Mosquera, L., and Hoptroff, R. (2020). *Practical synthetic data generation: balancing privacy and the broad availability of data*. O'Reilly, Beijing Boston Farnham Sebastopol Tokyo, first edition edition.

Fatima, S., Wooldridge, M., and Jennings, N. R. (2008). An anytime approximation method for the inverse shapley value problem.

Gelman, A. and Hennig, C. (2017). Beyond subjective and objective in statistics. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 180(4):967–1033.

Gong, R. and Meng, X.-L. (2021). Judicious judgment meets unsettling updating: dilation, sure loss and simpson's paradox.

Li, X. and Meng, X.-L. (2021). A multi-resolution theory for approximating infinite-p-zero-n: Transitional inference, individualized predictions, and a world without bias-variance tradeoff. *Journal of the American Statistical Association*, 116(533):353–367.

Liu, K. and Meng, X.-L. (2016). There Is Individualized Treatment. Why Not Individualized Inference? *Annual Review of Statistics and Its Application*, 3:79–111.

Lu, M., Sadiq, S., Feaster, D. J., and Ishwaran, H. (2018). Estimating individual treatment effect in observational data using random forest methods. *Journal of Computational and Graphical Statistics*, 27(1):209–219.

Meng, X.-L. (2018). Statistical paradises and paradoxes in big data (i) law of large populations, big data paradox, and the 2016 us presidential election. *The Annals of Applied Statistics*, 12(2):685–726.

Msaouel, P., Lee, J., Karam, J. A., and Thall, P. F. (2022). A causal framework for making individualized treatment decisions in oncology. *Cancers*, 14(16):3923.

Romaguera, S. (2022). Basic contractions of suzuki-type on quasi-metric spaces and fixed point results. *Mathematics*, 10(21):3931.

Vovk, V., Gammerman, A., and Shafer, G. (2022). *Algorithmic learning in a random world*. Springer.

Xie, M.-g. and Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review. *International Statistical Review*, 81(1):3–39.

Xie, M.-g. and Wang, P. (2022). Repro samples method for finite-and large-sample inferences. *arXiv preprint arXiv:2206.06421*.