Exploiting partial knowledge to evaluate the average causal effect via an ABC perspective

L'introduzione di conoscenze parziali per valutare l'effetto medio causale tramite una prospettiva ABC

Giulia Cereda, Fabio Corradi, Cecilia Viscardi

Abstract To evaluate the average causal effect, we propose a method to be used when the variables required to remove confounding are not observed. We consider the case in which the causal structure is known, and there is also some additional partial knowledge. The evaluation is probabilistic, and the partial knowledge may consist of an ordering of some relevant conditional probabilities and/or of the marginal distributions of some variables. A simple way to integrate such additional information is to adopt an approximate Bayesian computation perspective to derive an approximate posterior distribution for the average causal effect. We experimentally evaluate our methodology through an example from the literature, and we compare the results with the exact evaluation of the average causal effect.
Abstract *Al fine di valutare l'effetto causale medio, proponiamo un metodo che possa essere usato quando le variabili richieste per eliminare il confondimento non sono osservabili. Consideriamo il caso in cui la struttura causale del modello è nota e sono disponibili alcune informazioni aggiuntive. La valutazione fornita è probabilistica e la conoscenza aggiuntiva può consistere in un ordinamento di alcune rilevanti probabilità condizionate e/o nella conoscenza di distribuzioni marginali. Un modo semplice per incorporare tali informazioni è quello di adottare una prospettiva di calcolo Bayesiano approssimato per derivare una distribuzione a posteriori approssimata dell'effetto causale medio. La metodologia proposta è stata valutata su un esempio noto in letteratura e i risultati sono stati confrontati con una valutazione esatta.*

Key words: ABC, ACE, Adjustment formula, Causal graph, Simpson's Paradox.

1 Introduction and preliminaries

In causal inference, the evaluation of the Average Causal Effect (ACE) [5], measuring the difference between the expected values of the outcome under different scenarios, has significant relevance. A starting point to determine ACE is to assume the knowledge of a causal mechanism represented by a graph, more specifically a causal direct acyclic graph

Cecilia Viscardi
Dipartimento di Statistica, Informatica, Applicazioni e-mail: cecilia.viscardi@unifi.it

Fabio Corradi
Dipartimento di Statistica, Informatica, Applicazioni e-mail: fabio.corradi@unifi.it

Giulia Cereda
Dipartimento di Statistica, Informatica, Applicazioni e-mail: giulia.cereda@unifi.it

(DAG), defined by two sets $\{V,E\}$. V is the set of nodes representing the model variables, and E is the set of direct edges connecting nodes. Two connected nodes are linked by a direct causal relation, according to the orientation of the edge.

In a simple setting, two nodes of the causal graph are mainly of interest: X, the cause (often a treatment), and Y, the outcome. ACE is evaluated by measuring the difference in the expectation of Y for different interventions on X. An intervention on X, expressed by the so-called $do(\cdot)$ operator, consists of assigning the same x values to quantify changes in the system. This action is purely fictitious (observed x values correspond to past events that cannot be modified), but it has been proved that $do(X = x)$ corresponds to conditioning to $X = x$ in the model obtained as a surgery modified graph removing all the edges incident on X [4].

In the case of boolean Y and dichotomous X, the $do(\cdot)$ operator can be used to calculate the (adjusted) ACE according to:

$$ACE(X \to Y) = p(Y|do(X = x)) - p(Y|do(X = x')) \tag{1}$$

The discrepancy between the estimation of the causal effect obtained by (1) and

$$p(Y|X = x) - p(Y|X = x') \tag{2}$$

is due to confounding. Its detection may be obtained by using some topological properties of the causal graph, such as the well-known Back-door or Front-door criteria [4].

In a simple causal model with only one identified but unobserved confounder, we propose to incorporate some additional knowledge for the probabilistic evaluation of ACE via an algorithm based on the ABC type of reasoning. We experimentally evaluate our methodology through a running example, and we compare our results with a situation where the confounder is observed and the intervention can be evaluated correctly.

2 The running example

As a simple running example, we refer to a well-known causal mechanism with treatment and an outcome both influenced by a third variable, as represented in the causal graph in Fig. 2 (left). This graph represents the causal mechanism behind Simpson's paradox which occurs in many applications.

Fig. 1 (Left) Direct acyclic graph representing the causal mechanism assumed in the running example. (Right) Surgery modified causal graph.

To be concrete, we refer to an example provided by a real-life medical study concerning kidney stone treatments [3]. Let $Y \in \{0,1\}$ represent the no-recovery/recovery of a patient; $X \in \{A,B\}$ stand for two possible treatments; $Z \in \{S,L\}$ represent the size (Small or Large) of the stones of the treated individuals. The study gathers data from 700 patients with kidney stones undergoing the two treatments. Data are summarised in Table 1.

The causal story behind the graph tells us that both treatment and recovery are influenced by the size of the kidney stones of the exposed individual. Actually, doctors prefer

			Treatment (X)	
		Outcome (Y)	$X = A$	$X = B$
Size (Z)	Small	$Y = 1$	81	234
		$Y = 0$	6	36
	Large	$Y = 1$	192	55
		$Y = 0$	71	25
	Total	$Y = 1$	273	289
		$Y = 0$	77	61

Table 1 Complete data for the kidney stones example.

to assign the most effective and expensive treatment (A) to patients with the most severe kidney large stones pathology which experienced a lower rate of recovery. The opposite happens for patients affected by small kidney stones, which are more likely to receive treatment B and are more prone to recover.

If the leftmost graph in Fig. 2 correctly represents the causal mechanism, Z acts as a confounder, and the average causal effect can be evaluated as the difference in the probability of recovery under the two treatment regimes evaluated by using propagation of evidence in the surgery modified causal graph in Fig. 2 (right) or, directly and equivalently, by the so-called Adjustment formula [2]:

$$p(Y = 1|do(X = i)) = \sum_{j\in\{S,L\}} p(Y = 1|X = i, Z = j)p(Z = j) \qquad i \in \{A,B\}. \tag{3}$$

This can be used to calculate ACE according to (1):

$$ACE = \sum_{j\in\{S,L\}} \big(p(Y = 1|X = A, Z = j) - p(Y = 1|X = B, Z = j)\big)p(Z = j). \tag{4}$$

To estimate the components of (3) we need to observe the triplet (x, y, z) over a set of n observations; for the case at hand, the availability of the triplets, including observations on Z, is also the condition established by the Back-door criterion to remove confounding.

3 Dealing with unobserved confounders

Consider the causal graph in Fig. 2 again and assume that we are convinced of the validity of the depicted causal relations, but observations on Z are not accessible to us so that only the totals of Table 1 are available. The empirical joint distribution of X and Y leads to the conclusion that recovery is more likely undergoing treatment B (0.82) than undergoing treatment A (0.78), and the difference in predictions (2) amounts to -0.04.

Actually, the correct evaluation of ACE according to (4) amounts at 0.05, showing that treatment A is better than treatment B. This means that evaluating the causal effect relying on the probability of Y conditionally to X would be misleading.

A way to mitigate the effect of our lack of knowledge on Z is to exploit some additional pieces of information we are willing to assume. The main issue is finding a way to incorporate the available information in estimating the adjustment formula. In our running example we considered a set of three available pieces of information, $\mathscr{I} = \{\mathscr{I}_1, \mathscr{I}_2, \mathscr{I}_3\}$, defined as follows:

- $\mathscr{I}_1$ is the marginal distribution of Z, corresponding to assume that the distribution of small and large kidney stones among people affected by the disease is known in the population of interest;
- $\mathscr{I}_2$ is an ordering of $p(X|Z)$, in particular $p(X=A|Z=L) > p(X=B|Z=L)$ and $p(X=B|Z=S) > p(X=A|Z=S)$, corresponding to assume that Treatment A is more likely assigned to people with large kidney stones, while treatment B is assigned preferentially to people with small kidney stones;
- $\mathscr{I}_3$ is an ordering of $p(Y|X,Z)$, in particular

$$p(Y=1|Z=S,X=A) > p(Y=1|Z=S,X=B) > p(Y=1|Z=L,X=A) > p(Y=1|Z=L,X=B), \tag{5}$$

corresponding to assume that Treatment A is better than treatment B in both small and large kidney stones groups, and that the recovery with small stones is more probable than the recovery with large stones. Notice that this assumption implies a positive ACE.

In principle, a possible solution to estimate (4) is to build a complete database which require to impute Z values by sampling from the predictive

$$p(Z|X,Y) \propto p(Z)p(X,Y|Z) \tag{6}$$

which is based on the unavailable $p(X,Y|Z)$ in (3).

To deal with this issue and to incorporate $\mathscr{I}$ in the estimation, we adopt the same type of reasoning as ABC methods [6]. In particular, we consider the following distribution defined on an augmented space where X' and Y' are two auxiliary random variables:

$$p(Z,X',Y'|X,Y) \propto p(Z)p(X',Y'|Z)p(X,Y|X',Y',Z). \tag{7}$$

Generally speaking, samples from the above distribution can be obtained by implementing an ABC sampling scheme: 1) sample Z from its marginal $p(Z)$; 2) sample pairs (X',Y') from $p(X',Y'|Z)$; 3) approximate $p(X,Y|X',Y',Z) \approx \mathbb{1}\{(X,Y)=(X',Y')\}$. In practice, a two-fold approximation is usually introduced relying on $p(X,Y|X',Y',Z) \approx \mathbb{1}\{d\big(s(X,Y),s(X',Y')\big) \leq \varepsilon\}$, where $d(\cdot,\cdot)$ is a proper distance function, $s(\cdot)$ are summary statistics reducing the size of the data, and ε is a positive tolerance threshold.

Note that in the case at hand, the tuple (X',Y') represents pseudo-data to be generated from a *simulator*. The simulator would be a computer program producing samples from $p(X',Y'|Z)$ in the ABC framework. Here, due to the unavailability of such a generative mechanism, we consider an approximation $\hat{p}_{\mathscr{I}}(X',Y'|Z)$ that complies with $\mathscr{I}_2$ and $\mathscr{I}_3$. Such a mechanism may be based on the decomposition $\hat{p}_{\mathscr{I}}(X',Y'|Z) = \hat{p}_{\mathscr{I}_2}(X'|Z)\hat{p}_{\mathscr{I}_3}(Y'|X',Z)$ and the fact that $X'|Z$ and $Y'|X',Z$ are Bernoulli random variables whose parameters, denoted as $\theta_{X=A|Z}$ and $\theta_{Y=1|X,Z}$ for $X \in \{A,B\}$ and $Z \in \{S,L\}$, satisfy constraints implied by $\mathscr{I}_2$ and $\mathscr{I}_3$. More precisely, we get samples from $\hat{p}_{\mathscr{I}}(X',Y'|Z)$ as detailed in Algorithm 1.

Algorithm 1 SIMULATOR

Draw $\theta_{X=A|L} \sim Uniform(0.5,1)$; $\theta_{X=A|S} \sim Uniform(0,0.5)$ and $\theta_{Y=1|B,L} \sim Uniform(0,1)$
Draw $\theta_{Y=1|A,L} \sim Uniform(\theta_{Y=1|B,L},1)$
Draw $\theta_{Y=1|B,S} \sim Uniform(\theta_{Y=1|A,L},1)$
Draw $\theta_{Y=1|A,S} \sim Uniform(\theta_{Y=1|B,S},1)$

Once defined a simulator, we can exploit the knowledge deriving from observed data and get samples from an approximate distribution $\hat{p}_{\mathscr{I}}(Z,X',Y'|X,Y)$ following the ABC scheme in Algorithm 2.

Algorithm 2 ABC SCHEME

Draw $Z^{(t)} \sim p(Z)$ according to $\mathscr{I}_1 \quad \forall t \in \{1,...,T\}$
Generate $(X',Y')^{(t)} \sim \hat{p}^{(t)}_{\mathscr{I}}(X',Y'|Z)$ simulated according to Alg.1 $\quad \forall t \in \{1,...,T\}$
Accept $((X',Y')^{(t)},Z^{(t)})$ if $d(\widehat{p}^{(t)}_{X',Y'},\widehat{p}_{X,Y}) \leq \varepsilon \quad \forall t \in \{1,...,T\}$

Note that here observed and simulated data are compared by means of sufficient summary statistics, the empirical distributions $\hat{p}_{X,Y}$ and $\hat{p}_{X',Y'}$.

Looking at Alg. 2, it is apparent that the outlined sampling scheme can incorporate whatever state of knowledge simply by replacing Alg. 1 with a proper simulator based on different information $\mathscr{I}$. In any case, the output of the algorithm is a sample from an approximate posterior distribution $\hat{p}_{\mathscr{I}}(X',Y',Z|X,Y)$ which can be used to evaluate $p(Y=1|do(X=i))$ as

$$\sum_{j\in\{S,L\}} \hat{p}_{\mathscr{I}}(Y'^{(t)}=1|X'^{(t)}=i,Z^{(t)}=j)p(Z^{(t)}=j) = \sum_{j\in\{S,L\}} \theta^{(t)}_{Y=1|i,j}\cdot p(Z^{(t)}=j) \qquad i\in\{A,B\}$$

for each retained sample t. It follows that we came up with a posterior distribution of ACE based on the accepted triplets $((X',Y')^{(t)},Z^{(t)})$.

4 Experiments

We consider the example described in the previous section. Using the frequencies of Tab. 1, and formula (4) we obtain an ACE equal to 0.05.

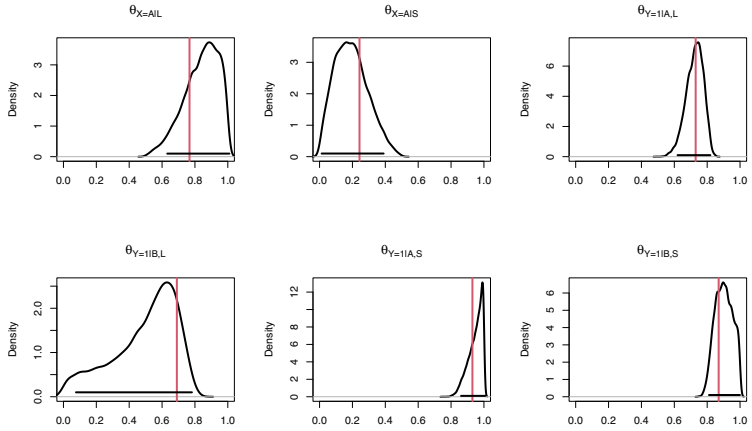


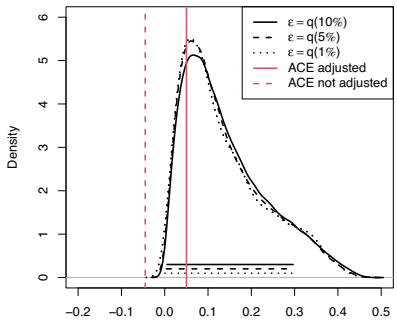
Fig. 2 Posterior distribution of the six θ parameters corresponding to the conditional probabilities of Algorithm 1.

To evaluate the performance of our method, we pretend that the values of Z are not observable and compute a posterior distribution for ACE by using $800,000$ samples from $\hat{p}_{\mathscr{I}}(Z,X',Y'|X,Y)$, obtained with the procedure described in Alg. 1 and Alg. 2.

Figure 2 shows the posterior distributions for $\theta_{X=A|S}$, $\theta_{X=A|L}$, $\theta_{Y=1|A,S}$, $\theta_{Y=1|B,S}$, $\theta_{Y=1|A,L}$ and $\theta_{Y=1|B,L}$, $\varepsilon = 0.076$, corresponding to the 1% quantile of the distance empirical dis-

tribution. Considering that before the analysis we had a uniform prior belief about their values, we can conclude that the procedure allows to learn appreciably about the six parameters. Figure 3 shows the ACE posterior distributions obtained using three different values of the ABC threshold, ε, selected through the quantile method [1].

Note that the three posterior modes are almost coincident with the true ACE, and they get closer and closer as the threshold ε decreases.



quantile	epsilon	HPD	MAP
1%	0.076	(0.000, 0.299)	0.057
5%	0.128	(0.004, 0.293)	0.059
10%	0.164	(0.005, 0.297)	0.062

Fig. 3 Posterior distributions for the adjusted ACE obtained using three different values of ε, along with the corresponding 90% HPD intervals. Table contains the specification of the ε values along with corresponding 90% HPD intervals and MAP.

5 Conclusions

In causal inference, this work provides a flexible method to evaluate ACE when observations of the variable(s) required to remove confounding are not available. The research stage is at its infancy, and we plan to consider more complex causal stories to verify the ability of the method to exploit qualitative information about the conditional probabilities beyond the causal DAG to provide useful probabilistic information about the required ACE. As a final remark, we want to stress that the method measures the confounding only under the assumption of its existence.

References

1. M. Beaumont, "Approximate Bayesian computation" *Annual Review of Statistics and Its Application*, vol. 6, no. 1, pp. 379–403, 2019.
2. S. Greenlan, "Relation of probability of causation, relative risk,and doubling dose: a methodlogical error that has become a social problem" *American Journal of Pubblic Healt*, vol. 89, no. 8, pp. 1166–1169, 1999.
3. S. A. Julious and M. A. Mullee, "Confounding and Simpson's paradox," *British Medical Journal*, vol. 309, pp. 1480–1481, 1994.
4. J. Pearl, "Causality" *Cambridge University Press*, 2009.
5. D. Rubin, "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies" *J. Educ. Psychol.*, vol. 66, no. 5, pp. 6688–701, 1995.
6. S. A. Sisson, Y. Fan, and M. Beaumont, *Handbook of approximate Bayesian computation.* Chapman and Hall\CRC., 2018.