Graphical Causal Models and Imputing Missing Data: A Preliminary Study

Rui Jorge Almeida[1](✉), Greetje Adriaans[2], and Yuliya Shapovalova[3]

[1] Department of Quantitative Economics,
Department of Data Analytics and Digitization, School of Business and Economics,
Maastricht University, Maastricht, The Netherlands
rj.almeida@maastrichtuniversity.nl

[2] Department of Hepatology and Gastroenterology,
Maastricht Universitary Medical Centrum+, Maastricht University,
Maastricht, The Netherlands
greetje.adriaans@mumc.nl

[3] Institute for Computing and Information Sciences,
Radboud University, Nijmegen, The Netherlands
Yuliya.Shapovalova@ru.nl

Abstract. Real-world datasets often contain many missing values due to several reasons. This is usually an issue since many learning algorithms require complete datasets. In certain cases, there are constraints in the real world problem that create difficulties in continuously observing all data. In this paper, we investigate if graphical causal models can be used to impute missing values and derive additional information on the uncertainty of the imputed values. Our goal is to use the information from a complete dataset in the form of graphical causal models to impute missing values in an incomplete dataset. This assumes that the datasets have the same data generating process. Furthermore, we calculate the probability of each missing data value belonging to a specified percentile. We present a preliminary study on the proposed method using synthetic data, where we can control the causal relations and missing values.

Keywords: Missing data · Graphical causal models · Uncertainty in missing values

1 Introduction

Datasets of real-world problems often contain missing values. A dataset has partial missing data if some values of a variable are not observed. Incomplete datasets pose problems in obtaining reliable results when analyzing the data. Many algorithms require a complete dataset to estimate models. On the other hand, in certain real-world problems obtaining reliable and complete data can be a tedious and costly task and can hamper the desired goal of the problem. An example is e-health. E-health tools often contain standardized forms (*i.e.* questionnaires) to capture data. Yet the questionnaires at times are lengthy and

M.-J. Lesot et al. (Eds.): IPMU 2020, CCIS 1237, pp. 485–496, 2020.
https://doi.org/10.1007/978-3-030-50146-4_36

this imposes a burden on the patients' time, which leads to reduced amount of patients completing questionnaires [1] causing incomplete datasets regarding e-health.

Since the introduction of the electronic health record (EHR) in Dutch clinical health care, large amounts of digital data are created on a daily basis. Furthermore, due to the emerging implementation of e-health applications in Dutch health care, large amounts of health-related data are created not only inside but also outside clinical institutions (*e.g.* hospitals). For instance, *MyIBDcoach*, is an e-health tool developed for home monitoring of disease activity for inflammatory bowel disease, a chronic disease with a relapsing-remitting disease course [9]. Results analyzing data captured in this e-health tool have shown the potential to predict disease activity. These results could potentially aid timely intervention and better health care resource allocation as the frequency of outpatient clinic visits could be scaled according to the risk of increased disease activity within a patient [10,24]. Exploring the further potential of combined data, data captured in the EHR and e-health tools, could lead to new insights by analyzing these data in a meaningful way.

In clinical studies, that use observational data, the data are often obtained by extracting information from the EHR. In addition, observational data documented in longitudinal prospective cohort studies often make use of standardized forms to register admission data of the cohort participants and to register data of certain variables during follow-up. Therefore datasets of prospective cohort studies can be considered complete. Since incomplete e-health datasets could lead to unreliable prediction results, incomplete data could, therefore, be problematic when e-health tools are used as an integral part in the care pathway [7].

In this paper we investigate if graphical causal models can be used to impute missing values. Causal discovery aims to learn the causal relations between variables of a system of interest from data. Thus it is possible to make predictions of the effects of interventions, which is important for decision making. Graphical models can represent a multivariate distribution in the form of a graph. Causal models can be represented as graphical models and represent not only the distribution of the observed data but also the distributions under interventions.

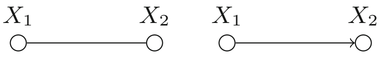
Causal inference has been applied to combine information from multiple datasets [15,16], including observational and experimental data [13,18]. Causal discovery algorithms have been adapted to deal with missing data [6]. For example, [4] presents a modification of PC algorithm [20] to be able to handle missing data, [19] and [5] present different approach to deal with mixed discrete and continuous data. We take a different perspective.

Our goal is to use the information from a complete dataset (*e.g.* cohort studies) in the form of graphical causal models to impute missing values in an incomplete dataset (*e.g.* from e-health monitoring). This assumes that these datasets represent the same population and have the same data generating process, which is implicit in setting up cohort studies. The use of causal models allows preserving causal relationships present in data, without strict assumptions of a pre-specified data generating process. Furthermore, we explore the stochastic uncertainty in imputing missing values with the proposed method. We calculate the probability of each missing data value belonging to a specified percentile. Low or high percentiles can indicate risk situations, *e.g.* existence of an active disease in e-health monitoring. In this paper we present a preliminary study using synthetic data, where we can control the causal relations and for which there is ground truth for the missing values.

2 Preliminaries

2.1 Graphical Models and Causal Discovery

A causal structure is often represented by a graphical model. A graph G is an ordered pair $<V, E>$ where V is a set of vertices, and E is a set of edges [20]. The pairs of vertices in E are unordered in an undirected graph and ordered in a directed graph. A **directed graph** G contains only directed edges as illustrated in Fig. 1(b). A **directed acyclic graph** (DAG) often represents underlying causal structures in causal discovery algorithms [17]. On the other hand, a **mixed graph** can contain more than one type of an edge between to vertices. A DAG contains only directed edges and has no directed cycles. We call the **skeleton** of a DAG an undirected graph obtained by ignoring direction of the edges in the DAG itself. See Figs. 2(a) and 2(b) for illustration. Further, if there is a directed edge from X_1 to X_2 then X_1 is called to be **parent** of X_2, and X_2 is called to be **child** of X_1. If two vertices are joined by an edge they are called to be **adjacent**. A set of parents of a vertex X_2 is denoted by $\mathrm{pa}(X_2)$, in Fig. 2(a) $\mathrm{pa}(X_2) = \{X_1\}$ while $\mathrm{pa}(X_4) = \{X_2, X_3\}$. The joint distribution implied by Fig. 2(a) implies the following conditional probability relation:



(a) Undirected (b) Directed

Fig. 1. Undirected and directed relationship between two variables

$$P(V) = \prod_{X \in V} P(X|\mathrm{pa}(X)). \tag{1}$$

Causal discovery connects the graphical theoretic approach and statistical theory. The DAG in Fig. 2(a) implies the following conditional distributions:

$$P(X|\mathrm{pa}(X)) = P\left(X \mid \bigcup_{X_j \in \mathrm{pa}(X)} \mathrm{pa}(X_j)\right), \tag{2}$$

e.g. $P(X_4|\mathrm{pa}(X_4)) = P(X_4|X_2, X_3) = P(X_4|X_1)$.

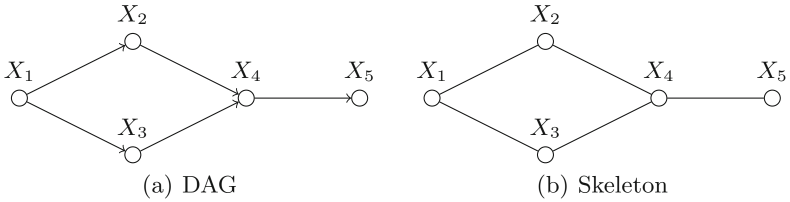


Fig. 2. Directed acyclic graph and its skeleton

A DAG encodes conditional independence relationships, which help us to reason about causality. A criteria known as **d-separation** is central in this type of inference, see for more details [20]. In particular, in any distribution P factorizing according to G, if X and Y are d-separated given Z then $X \perp\!\!\!\perp Y|Z$ in P. There are multiple algorithms that use d-separation rules to learn the graph structure; many of them are computationally intensive.

In this paper we use the PC algorithm[1] for causal discovery [20]. The idea of this algorithm is based on first forming the complete undirected graph, then removing the edges with zero-order conditional independence, then removing first-order conditional independence relations, etc. Thus, the PC algorithm heavily relies on testing conditional independence. Pearson's correlation is frequently used to test for conditional independence in the Gaussian case; other popular choices are, Spearman's rank correlation, or Kendall's tau. In addition, next to the correlation matrix, the PC algorithm requires a sample size as input. The estimate of the correlation matrix is more reliable with larger sample size, and thus we easier can reject the null hypothesis of conditional independence [5].

PC algorithm is widely applied in causal discovery algorithms and thus has been extended in various directions, including missing data cases. [3] consider causal discovery in DAGs with arbitrarily many latent and selection variables with the available R software package *pcalg* [11]. [8] use rank-based correlation and extend PC algorithm to Gaussian copula models. [4] extend this approach to mixed discrete and continuous data, while [5] further include missing data in this approach.

2.2 Graphical Models with Missing Data

In this paper we are exploiting the idea that one can infer causal structure from a cohort study and then use this information for imputing missing values in an incomplete dataset. The problem of missing data in causal inference is being studied in the literature quite extensively. [14] derive graphical conditions for recovering joint and conditional distributions and sufficient conditions for recovering causal queries. [22] consider different missingness mechanisms and present graphical representations of those. Usually, three missing mechanisms are considered in the literature [12]: missing completely at random (MCAR), missing at

[1] Named after its two inventors, Peter and Clark.

random (MAR), and not missing at random (NMAR). MCAR missingness mechanism imposes the least problems for statistical inference, while NMAR imposes most problems for statistical inference. It is important to note, that in our case there is no problem of identifying the type of missingness mechanism, however, it is useful to know and understand the distinction of missing mechanisms from the literature.

Similarly to [22] let us denote by D_{obs} observed part of the data and D_{mis} missing part of the data, and R the indicator matrix of missingness. The MCAR mechanism states that

$$P(R|D) = P(R|D_{obs}, D_{mis}) = P(R). \tag{3}$$

Equation (3) can be expressed in conditional independence statement as

$$R \perp\!\!\!\perp (D_{obs}, D_{mis}). \tag{4}$$

Thus, the missingness in this case is independent of both D_{obs} and D_{mis}. Further, MAR, a less restrictive mechanism, states that

$$P(R|D) = P(R|D_{obs}, D_{mis}) = P(R|D_{obs}), \tag{5}$$

where Eq. (5) can also be expressed in terms of a conditional independence statement

$$R \perp\!\!\!\perp D_{mis}|D_{obs}. \tag{6}$$

Thus, while the dependence between the observed data and missingness is allowed, the missingness R is independent of missing part of the data D_{mis} given information about the observed part of the data D_{obs}. Finally, for NMAR mechanism we have

$$P(R|D_{obs}, D_{mis}) \neq P(R|D_{obs}). \tag{7}$$

[22] propose a way to create m-graphs (graphs with missing data for all three mechanisms) and discuss graphical criteria for identification of means and regression coefficients. For us it is useful in a sense that while deciding on which parts of the data can be missing, we can impose requirement of identifiability.

3 Causal Models for Imputing Missing Data

In this paper, we propose using the causal information from a DAG, built from a complete sample, to impute missing values in another sample. The proposed method uses the causal discovery defined within a DAG and estimated relations between variables using the PC algorithm. The DAG and PC estimation provide the causal relations between the missing and observed variables. Once this relation is defined, the exact specification of causality between observed and missing values, together with the predictions of the missing values are obtained using nonparametric regressions. Nonparametric regressions are used to avoid assumptions on the specific functional relationship between variables.

As an illustration. Suppose that X_1, X_2, X_3, X_5 are observed in Fig. 2(a) while X_4 is missing. The DAG implies the following conditional probability relation:

$$P(X_4|\mathrm{pa}(X_4)) = P(X_4|X_2, X_3). \tag{8}$$

In case both X_4 and a parent, *e.g.* X_2, are missing, we use the following DAG-implied conditional probability relations to estimate the causal relationship between X_2 and X_1 in the training data, and obtain an estimate for the missing value of X_2. Impute the missing values of X_2 and X_4:

$$P(X_2|\mathrm{pa}(X_2)) = P(X_2|X_1) \tag{9}$$

$$P(X_4|\mathrm{pa}(X_4)) = P(X_4|X_2, X_3) = P(X_4|\mathrm{pa}(X_2), X_3) = P(X_4|X_1, X_3) \tag{10}$$

The iteration over parents of DAG implied conditionals continues until all conditioning variables are observed. When X_2, X_3, X_4 are all unobserved, we use the following DAG-implied conditional probability relation:

$$P(X_2|\mathrm{pa}(X_2)) = P(X_2|X_1) \tag{11}$$

$$P(X_3|\mathrm{pa}(X_3)) = P(X_3|X_1) \tag{12}$$

$$P(X_4|\mathrm{pa}(X_4)) = P(X_4|X_2, X_3) = P(X_4|\mathrm{pa}(X_2), \mathrm{pa}(X_3)) = P(X_4|X_1). \tag{13}$$

When the graph structure is more complicated than Fig. 3, the above procedure to obtain 'observed parents' of a missing value is more involved since backward iterations of $\mathrm{pa}(\cdot)$ are needed until none of the conditioned variables have missing values. To avoid this computational cost, we define the iterated parents of a missing observation. Let $X_{\mathrm{mis}} \subset \mathrm{pa}(X)$ denote the set of parents of X with missing values. The iterated parents of X, $\hat{\mathrm{pa}}(X)$ are defined as:

$$\hat{\mathrm{pa}}(X) = \begin{cases} \mathrm{pa}(X) & \text{if } X_{\mathrm{mis}} = \emptyset \\ (\mathrm{pa}(X) \setminus X_{\mathrm{mis}}) \cup X_1 & \text{otherwise,} \end{cases} \tag{14}$$

where the conditioning on variable X_1 is due to the graph structure in Fig. 3.

Given the conditional probability definitions in Eqs. (8)–(13), and the parent set definition in (14), we propose to obtain the predicted values of missing values using nonparametric regressions. For N observed data samples $X_{i,j}$ with $i = 1, \ldots, p$ and $j = 1, \ldots, N$, local linear regressions are estimated for each variable X_i in a training set. Each of these local linear regressions minimize the following:

$$\min_{\alpha,\beta} \sum_{n=1}^{N} \left(X_{i,j} - \alpha - \beta\left(\hat{\mathrm{pa}}(X_i) - \hat{\mathrm{pa}}(X)_{i,j}\right)\right)^2 K_h\left(\hat{\mathrm{pa}}(X_i) - \hat{\mathrm{pa}}(X)_{i,j}\right) \tag{15}$$

where $X_i = (X_{i,1}, \ldots, X_{i,N})'$ is the vector of observations from variable X_i, $\mathrm{pa}(X_i) = (\mathrm{pa}(X)_{i,1}, \ldots, \mathrm{pa}(X)_{i,N})$ and $\mathrm{pa}(X)_{i,j}$ denotes the jth observation from parents of X_i. In addition, $K_h\left(\mathrm{pa}(X_i) - \mathrm{pa}(X)_{i,j}\right)$ is defined as a Gaussian kernel with $h = 1$, but the proposed methodology is applicable to other kernel specifications or similar nonparametric regression methods.

The imputation method we propose is based on estimating (15) for a DAG based on complete data, and predicting the missing values in an incomplete dataset. This imputation, denoted by $\hat{X}_{i,j}$ for variable i in observation j is calculated using the local linear regression results:

$$\hat{X}_{i,j} = \hat{\alpha} + \hat{\beta}\,(\hat{\text{pa}}(X_i) + \hat{\text{pa}}(X)_{i,j})\,, \tag{16}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are obtained according to the minimization in (15). In addition, the Gaussian kernel defined for (15) implies local normality for all predicted values. We use this property to quantify the uncertainty of the imputed value in (16). Given a normal distribution $X_{i,j} \sim N(\hat{X}_{i,j}, \hat{\sigma}^2_{i,j})$, we calculate the probability of $X_{i,j}$ belonging to a pre-specified percentile range $[p_1, p_2]$ as:

$$pr(p_1 < X_{i,j} \leq p_2) = \int_{p_1}^{p_2} \phi\left(X_{i,j}; \hat{X}_{i,j}, \hat{\sigma}^2_{i,j}\right) dX_{i,j} \tag{17}$$

where $\phi\left(x; \mu, \sigma^2\right)$ denotes the probability density function with mean μ, variance σ^2 evaluated at point x and $\hat{\sigma}^2_{i,j}$ is estimated as the variance of the regression errors. Please note that in this preliminary study, we ignore uncertainty when estimating the model parameters α and β.

4 Simulation Results for Imputing Missing Data

We illustrate the performance of the proposed method using a DAG with eight variables. The random graph is defined for 8 variables with conditional Gaussian distributions and the probability of connecting a node to another node with higher topological ordering is set as 0.3, following [3]. The true DAG and the estimated DAG are presented in Fig. 3. The structure of this DAG implies that variables 2, 3, 5, 6, 7 and 8 can be explained by parent variables or variable 1. Variable 4, on the other hand, is completely exogenous in this graph. Hence our methodology cannot be used to impute missing values of variable 4.

We simulate 5000 training observations and estimate the DAG using these training data. The estimated DAG is presented in the right panel of Fig. 3. Given the test data with 2000 observations, we create 9 incomplete test datasets with randomly missing values (MCAR) for 6 variables that have parents in the map, *i.e.* variables 2, 3, 5, 6, 7 and 8. These 9 incomplete test datasets differ in the probability of missing observations $q = 10\%, 20\%, \ldots, 90\%$. Each observation can have none, one or more missing variables, hence the total number of missing observations in each incomplete test dataset is 2000 or less, while the expected number of missing variables is $q \times 2000 \times 6$.

For each incomplete training dataset, we use the methodology in Sect. 3 to impute the missing values. We compare our method to other baseline models, namely replacing missing values by the sample average of the variable in the test data, excluding missing values; the MissForest method, a non-parametric missing value imputation based on random forests [21]; and MICE a multivariate imputation method based on fully conditional specification [23], as implemented

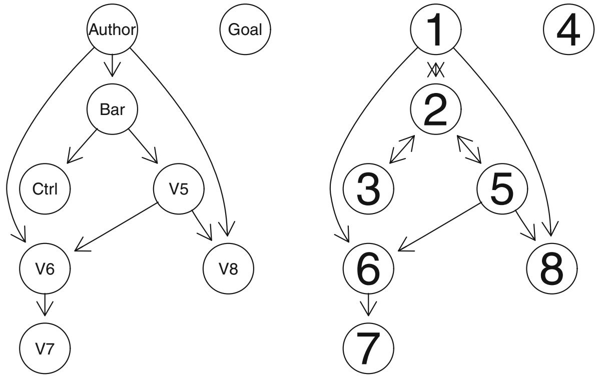


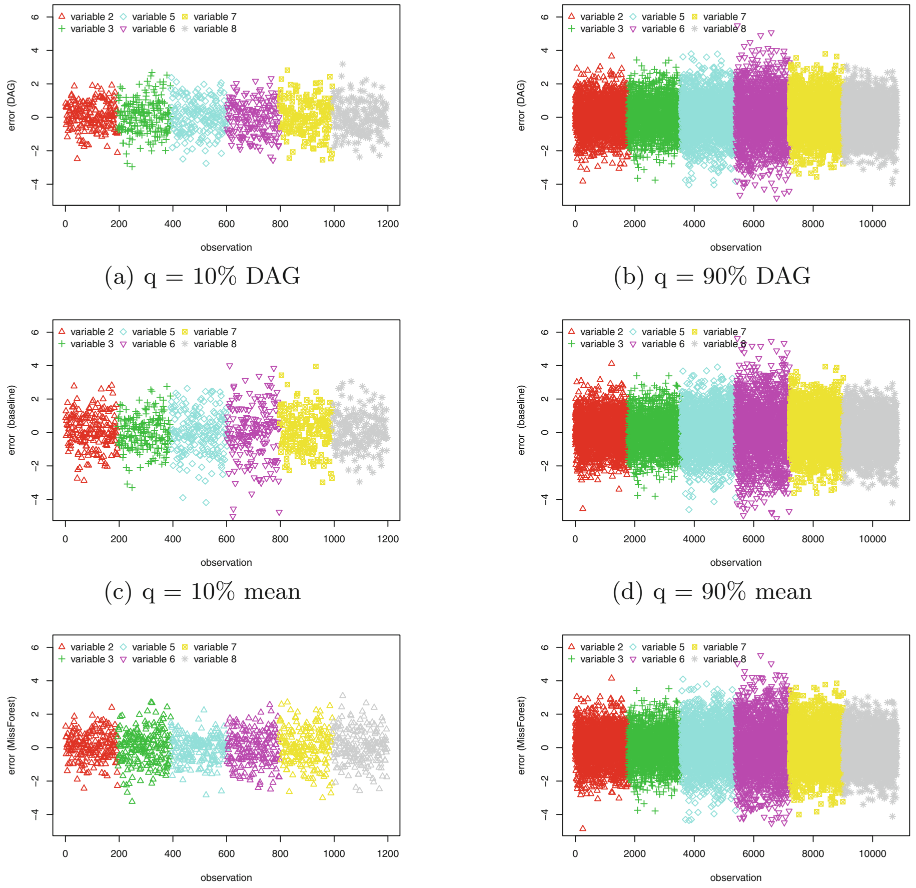
Fig. 3. True (left) and estimated (right) DAG for simulated data

in [2]. The mean squared errors of the proposed model and the baseline models are reported in Table 1. For missing values above $q = 40\%$, the proposed method performs better than all other models. For values below $q = 40\%$, the best performing model is MissForest, although the results appear to be comparable. The proposed model performance, measured by the MSE in Table 1, decreases with increasing q. This result is expected as the number of missing values for each observation increase with q. Since this increase implies that within an observation, it is more likely that the parents of a missing variable are also unobserved, hence there is an additional loss of information in the causal relations.

Table 1. MSE results from the proposed method and baseline models

	10%	20%	30%	40%	50%	60%	70%	80%	90%
Mean	1.47	1.49	1.56	1.51	1.48	1.52	1.50	1.51	1.52
DAG	0.92	0.99	1.03	**1.06**	**1.10**	**1.15**	**1.19**	**1.26**	**1.32**
MissForest	**0.91**	**0.96**	**1.02**	1.09	1.26	1.38	1.49	2.05	1.50
MICE	1.72	1.78	1.86	1.90	2.02	2.11	2.29	2.43	2.83

In addition to the overall results in Table 1, we present the errors for each variable for $q = 10\%$ and $q = 90\%$ in Fig. 4. For a small percentage of missing values, $q = 10\%$, the ranges of errors are clearly smaller in the proposed method compared to the mean baseline model. The MissForest model has some observations with a larger absolute error compared to the proposed method. Note that the variable-specific errors present the cases where the causal relations, hence the imputations are relatively less accurate. For variables 2 and 3, which have a single parent and a short link to variable 1, the obtained errors are relatively small in absolute values. Other variables, such as 6 and 8, have multiple parents, thus a higher probability of missing values in parents. When the missing parent information is replaced with the value of variable 1, some information is lost and the estimates will be less accurate. Figure 4 shows that this inaccuracy occurs especially for $q = 90\%$ where the probability of missing observations, hence the probability of missing parent information is high.



(a) q = 10% DAG

(b) q = 90% DAG

(c) q = 10% mean

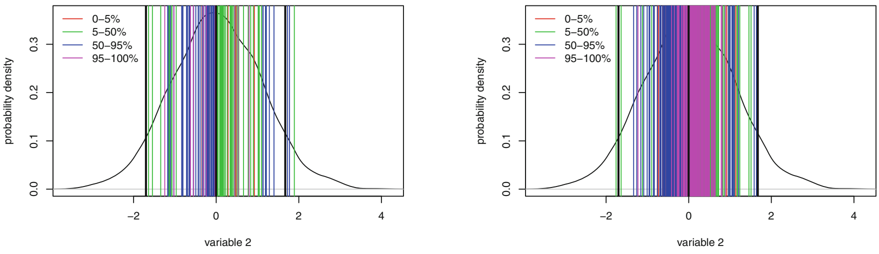
(d) q = 90% mean

(e) q = 10% MissForest

(f) q = 90% MissForest

Fig. 4. Errors per variable from imputed missing values using the proposed method (DAG), the mean and MissForest baseline models.

Finally, we illustrate the uncertainty in the missing values, quantified using the imputed values. For each variable, we set four pre-defined percentiles of 0–10%, 10–50%, 50–90% and 90–100%, corresponding to the empirical percentiles of the training data. We then calculate percentile probabilities for missing value by applying Eq. 17 for the four pre-defined percentiles. Based on these percentile probabilities, the percentile with the highest probability is selected as the estimated percentile. In Fig. 5, we present the imputed data values and percentile estimates for variable 2 for two missing value probabilities, $q = 10\%$ and $q = 90\%$. For readability, we only present observations for which the estimated and true percentiles are different. In addition, estimated percentiles are indicated with the respective colors and thick vertical lines indicate the thresholds for correct percentiles.



(a) q = 10%

(b) q = 90%

Fig. 5. Estimated percentiles for observations with different estimated and true percentiles

In this figure, overlapping colors indicate that similar imputed values can be classified in different percentiles according to the highest probability of belonging to a percentile. I.e. probabilities of belonging to a percentile can be used as an additional measure, with additional information, compared to the point estimates used as imputation. In Fig. 5, the number of overlaps are higher for a higher percentage of missing values $p = 90\%$, since there is more missing data. However, it appears that irrespective of the amount of missing values, both cases show the same pattern of overlap between estimated percentiles. This is an interesting result, since more missing values mean do not indicate more uncertainty in the estimated percentiles. This is likely due to the fact that our method derives information for imputation of missing values from causal relationships.

5 Conclusions and Future Work

In this paper we investigate if graphical causal models derived from complete datasets can be used to impute missing values in an incomplete dataset, assuming the same data generating process. We calculate the probability of each missing data value belonging to a specified percentile, to provide information on the uncertainty of the imputed values. We apply this methodology using synthetic data, where we can control the causal relations and missing values. We show that the proposed method performs better than a baseline model of imputing missing values by the mean in different simulation settings with different percentages of missing data. Furthermore, our model can still provide adequate information on missing values for very high percentages of missing values. Our results show that this methodology can be used in inputting missing values while providing information about the probability distribution of percentiles the missing value belongs to.

This is a preliminary study which opens many questions. In the future we want to investigate how to incorporate information on bidirectional causal relationships, different non-parametric models for imputing missing values and the relationship of this method with fully conditional specification.

References

1. Blankers, M., Koeter, M.W., Schippers, G.M.: Missing data approaches in eHealth research: simulation study and a tutorial for nonmathematically inclined researchers. J. Med. Internet Res. **12**(5), e54 (2010)
2. Buuren, S.V., Groothuis-Oudshoorn, K.: MICE: multivariate imputation by chained equations in R. J. Stat. Softw. **45**(3), 1–67 (2011)
3. Colombo, D., Maathuis, M.H., Kalisch, M., Richardson, T.S.: Learning high-dimensional directed acyclic graphs with latent and selection variables. Ann. Stat. **40**, 294–321 (2012)
4. Cui, R., Groot, P., Heskes, T.: Copula PC algorithm for causal discovery from mixed data. In: Frasconi, P., Landwehr, N., Manco, G., Vreeken, J. (eds.) ECML PKDD 2016. LNCS (LNAI), vol. 9852, pp. 377–392. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46227-1_24
5. Cui, R., Groot, P., Heskes, T.: Learning causal structure from mixed data with missing values using Gaussian copula models. Stat. Comput. **29**(2), 311–333 (2018). https://doi.org/10.1007/s11222-018-9810-x
6. Ding, P., Li, F., et al.: Causal inference: a missing data perspective. Stat. Sci. **33**(2), 214–237 (2018)
7. Gorelick, M.H.: Bias arising from missing data in predictive models. J. Clin. Epidemiol. **59**(10), 1115–1123 (2006)
8. Harris, N., Drton, M.: PC algorithm for nonparanormal graphical models. J. Mach. Learn. Res. **14**(1), 3365–3383 (2013)
9. de Jong, M., et al.: Development and feasibility study of a telemedicine tool for all patients with IBD: MyIBDcoach. Inflamm. Bowel Dis. **23**(4), 485–493 (2017)
10. de Jong, M.J., et al.: Telemedicine for management of inflammatory bowel disease (myIBDcoach): a pragmatic, multicentre, randomised controlled trial. Lancet **390**(10098), 959–968 (2017)
11. Kalisch, M., Mächler, M., Colombo, D., Maathuis, M.H., Bühlmann, P.: Causal inference using graphical models with the R package pcalg. J. Stat. Softw. **47**(11), 1–26 (2012). http://www.jstatsoft.org/v47/i11/
12. Little, R.J., Rubin, D.B.: Statistical Analysis with Missing Data, vol. 793. Wiley, Hoboken (2019)
13. Magliacane, S., Claassen, T., Mooij, J.M.: Joint causal inference on observational and experimental datasets. arXiv preprint arXiv:1611.10351 (2016)
14. Mohan, K., Pearl, J.: Graphical models for recovering probabilistic and causal queries from missing data. In: Advances in Neural Information Processing Systems, pp. 1520–1528 (2014)
15. Mooij, J., Heskes, T.: Cyclic causal discovery from continuous equilibrium data (2013). arXiv preprint arXiv:1309.6849

16. Mooij, J.M., Magliacane, S., Claassen, T.: Joint causal inference from multiple contexts. arXiv preprint arXiv:1611.10351 (2016)
17. Pearl, J., Verma, T.S.: A statistical semantics for causation. Stat. Comput. **2**(2), 91–95 (1992). https://doi.org/10.1007/BF01889587
18. Rau, A., Jaffrézic, F., Nuel, G.: Joint estimation of causal effects from observational and intervention gene expression data. BMC Syst. Biol. **7**(1), 111 (2013)
19. Sokolova, E., Groot, P., Claassen, T., von Rhein, D., Buitelaar, J., Heskes, T.: Causal discovery from medical data: dealing with missing values and a mixture of discrete and continuous data. In: Holmes, J.H., Bellazzi, R., Sacchi, L., Peek, N. (eds.) AIME 2015. LNCS (LNAI), vol. 9105, pp. 177–181. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19551-3_23
20. Spirtes, P., Glymour, C.N., Scheines, R., Heckerman, D.: Causation, Prediction, and Search. MIT Press, Cambridge (2000)
21. Stekhoven, D.J., Bühlmann, P.: Missforest-non-parametric missing value imputation for mixed-type data. Bioinformatics **28**(1), 112–118 (2012)
22. Thoemmes, F., Mohan, K.: Graphical representation of missing data problems. Struct. Eq. Model.: Multidiscip. J. **22**(4), 631–642 (2015)
23. Van Buuren, S.: Multiple imputation of discrete and continuous data by fully conditional specification. Stat. Methods Med. Res. **16**(3), 219–242 (2007)
24. Wintjens, D.S., et al.: Novel perceived stress and life events precede flares of inflammatory bowel disease: a prospective 12-month follow-up study. J. Crohn's Colitis **13**(4), 410–416 (2019)