In H. Geffner, R. Dechter, and J.Y. Halpern (Eds.), Probabilistic and Causal Inference: The Works of Judea Pearl, Association for Computing Machinery, 2021.

34

Causal Graphs for Missing Data: A Gentle Introduction

Karthika Mohan (University of California)

In this chapter we describe how causal graphs can be used for processing missing data. In particular, we model the missingness process using causal graphs and present graph-based definitions of various missingness mechanisms. Given a graph and a target quantity to be estimated, we present various methods for determining if and how a consistent estimate of the quantity can be computed. We further present techniques for detecting misspecifications in the model. We demonstrate all of the above using toy examples and small graphs, thus making it easy to understand the various intricacies and nuances of graph-based missing data analysis.

34.1 Introduction

Consider the following problems: (i) estimating the average income of a population in which the wealthy are reluctant to reveal their income, (ii) estimating the causal effect of diet and stress on obesity, given a dataset in which teenagers left several questions unanswered, and (iii) making product recommendations using data in which customers rated items only when they loved it. The underlying common theme in (i), (ii), and (iii) above is the estimation of a desired target quantity given missing data, that is, data in which values of one or more variables are missing.

Problems caused due to missing data are notoriously complex, afflict all branches of empirical sciences, and could potentially bias the outcomes of studies. Much of the research on missing data has focused on identifying conditions (such as Missing at Random [MAR] and Missing Completely at Random [MCAR]

[Rubin 1976]) under which the causes of missingness can be ignored when estimating quantities of interest. A widely held belief is that when the underlying cause of missingness is not random (Missing Not at Random (MNAR) [Rubin 1976]), it is rarely possible to compute estimates with any degree of confidence (example 1.17 in Little and Rubin [2002]).

In this chapter we discuss the recent advances in missing data theory that facilitate processing of MNAR data (i.e., non-ignorable missingness); in particular, we focus on *recoverability* (i.e., computing consistent estimates of quantities of interest) and *testability* (i.e., developing tests to determine the compatibility of a model with the available data). The following section describes missingness graphs (m-graphs), which are causal graphs that encode the (causal and statistical) assumptions about the process that generated missing data.

34.2 Missingness Graphs

Let $G(\mathbf{V}, E)$ be the causal directed acyclic graph (DAG) where $\mathbf{V}$ is the set of nodes and E is the set of edges. Nodes in the graph correspond to variables in the dataset and are partitioned into five categories, namely, $\mathbf{V} = V_o \cup V_m \cup U \cup V^* \cup R$ as described in Table 34.1. For example, in Figure 34.1(a), $V_o = \{G\}, V_m = \{I\}, R = \{R_I\}, V^* = \{I^*\}$, and $U = \{\}$.

Table 34.1 **Notations**

V_O	Set of all fully observed variables
V_m	Set of variables with missing values
U	Set of unobserved (latent) variables
R	Set of missingness mechanisms
V^*	Set of all proxy variables

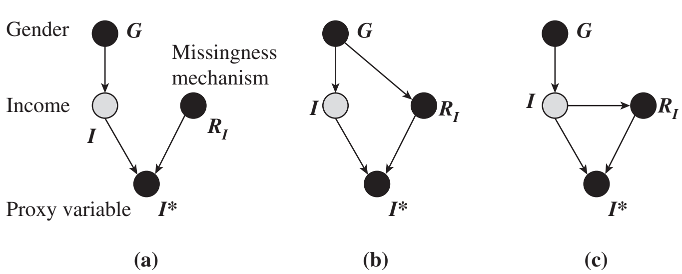

Figure 34.1 Causal graphs depicting various missingness mechanisms: (a) Missing Completely at Random (MCAR), (b) Missing at Random (MAR), and (c) Missing Not at Random (MNAR).

Every $X \in V_m$ is associated with two variables R_X and X^*, where X^* is the proxy variable that is actually observed and R_X represents the status of the mechanism responsible for the missing values in X^*; formally,

$$x^* = f(r_x, x) = \begin{cases} x & \text{if } r_x = 0 \\ m & \text{if } r_x = 1 \end{cases} \tag{34.1}$$

Unless stated otherwise, it is assumed that no variable in $V_o \cup V_m \cup U$ is a child of an R variable. U is the set of unobserved nodes, also called latent variables. Two nodes X and Y can be connected by a directed edge, that is, $X \rightarrow Y$, indicating that X is a cause of Y, or by a bi-directed edge, $X \leftrightarrow Y$, denoting the existence of a latent variable $U_i \in U$ that is a parent of both X and Y. This graphical representation is called a *missingness graph* (or *m*-graph) [Mohan et al. 2013]. $P^*(V^*, V_o, R)$ is called the observed data distribution, and $P(V_m, V_o, R)$ is called the true distribution. Any given true and observed data distribution are said to be compatible if the latter can be constructed from the former by repeatedly applying Equation (34.1). Conditional independencies are read off the graph using the d-separation criterion [Pearl 2009]. For any binary variable X, x' and x denote $X = 0$ and $X = 1$, respectively.

34.2.1 Graphical Representation of Missingness Categories

The graphical model–based definitions of the various missingness mechanisms [Rubin 1976, Little and Rubin 2002] that can be used to effortlessly decide the missingness categories are as follows:

1. Data are MCAR if $V_m \cup V_o \cup U \perp\!\!\!\perp R$ holds in the m-graph. Example: m-graph in Figure 34.1(a) in which $\{G, I\} \perp\!\!\!\perp R_I$ holds. Essentially, parents of R variables can only be other R variables.
2. Data are MAR if $V_m \cup U \perp\!\!\!\perp R|V_o$ holds in the m-graph. Example: m-graph in Figure 34.1(b) in which $\{I\} \perp\!\!\!\perp R_I|G$ holds. For MAR to hold, no parent of any R variable should belong to $V_m \cup U$; put differently, parents of R variables may only belong to $V_O \cup R$.
3. Data that are not MAR fall under the MNAR category. Example: m-graph in Figure 34.1(c) in which $\{I\} \not\perp\!\!\!\perp R_I|G$. In this case at least one R variable will have a parent that is either a latent variable or a variable with missing values, that is, belonging to $V_m \cup U$.

Remark 34.1 Observe that the graphical condition for MCAR immediately satisfies that for MAR; if parents of R variables may only be other R variables, then they clearly cannot be in $V_m \cup U$. Thus any model that is MCAR is MAR as well; this also follows from the weak union graphoid axiom [Pearl 2009].

34.3 Recoverability

Let Q denote a quantity of interest such as the joint/conditional distribution and causal effect. Given an m-graph G, Q is recoverable if there exists an algorithm that can (asymptotically) compute the true value of Q *as if* no data were missing. In the remainder of this section, we exemplify various recoverability techniques and their intricacies using small graphs as favored and taught by Judea Pearl, and as seen in many of his publications.

34.3.1 Recoverability in MAR and MCAR Problems

Consider the problem of recovering the joint distribution $P(G, I)$ given the m-graph in Figure 34.1(a) and the observed data distribution in Table 34.2.

$$\begin{aligned} P(G,I) &= P(G,I|r_I') \text{ (since } \{G,I\} \perp\!\!\!\perp R_I \text{ in the m-graph)} \\ &= P(G,I^*|r_I') \text{ (using Equation 34.1)} \end{aligned} \tag{34.2}$$

The preceding equations demonstrated how $P(G, I)$, which is a function of the partially observed variable I and fully observed variable G, is transformed into one over variables in the observed data distribution, I^* and G. The final expression derived in Equation (34.2), $P(I^*, G|r_I')$, is an estimand for $P(G, I)$, that is, it is an expression for $P(G, I)$ in terms of the available data that precisely defines what needs to be estimated. Recoverability is established once we derive an estimand. Note that the observed data distribution per se played no part in recoverability, which was established using assumptions in the m-graph ($\{G, I\} \perp\!\!\!\perp R_I$) and the missingness equation Equation (34.1). Thus, *recoverability is a property of the m-graph*.

34.3.1.1 Recoverability of Joint Distribution in MCAR and MAR Models

We shall now show that the joint distribution, $P(V_m, V_o)$, is recoverable in all MCAR and MAR m-graphs.

Table 34.2 Observed data distribution generated by the m-graph in Figure 34.1(a)

G	I^*	R_I	$P(G, I^*, R_I)$
M	H	0	p_1
M	L	0	p_2
F	H	0	p_3
F	L	0	p_4
M	M	1	p_5
F	M	1	p_6

G and I are binary variables that can take values Male (M) and Female (F), and High (H) and Low (L), respectively. P_is denote probabilities such that $\sum_{i=1}^{6} p_i = 1$.

Recoverability of joint distribution $P(V_o, V_m)$ in MCAR problems:

$$\begin{aligned} P(V_o, V_m) &= P(V_o, V_m|R=0) \text{ (since } (V_m, V_O) \perp\!\!\!\perp R \text{ when MCAR holds in an m-graph)} \\ &= P(V_o, V^*|R=0) \text{ (using Equation 34.1)} \end{aligned} \tag{34.3}$$

Recoverability of joint distribution $P(V_o, V_m)$ in MAR problems:

$$\begin{aligned} P(V_o, V_m) &= P(V_m|V_O)P(V_O) \\ &= P(V_m|V_O, R=0)P(V_O) \text{ (since } V_m \perp\!\!\!\perp R|V_O \text{ when MAR holds in an m-graph)} \\ &= P(V^*|V_O, R=0)P(V_O) \text{ (using Equation 34.1)} \end{aligned} \tag{34.4}$$

Equations (34.3) and (34.4) establish recoverability by presenting an estimand for the joint distribution.

34.3.1.2 Recoverability as a Guide for Estimation

Having established recoverability for all MAR and MCAR problems, we will now show how recoverability serves as a guide for estimation. We will exemplify estimation using deletion-based procedures.

The estimand in Equation 34.2 can be expressed as,

$$P(I^*, G|r_I') = \frac{P(I^*, G, r_I')}{P(r_I')}$$

It licenses the estimation of $P(G, I)$ exclusively from cases/samples in which $V_m = \{I\}$ is always observed, that is, $R_I = 0$. This procedure is known as *listwise deletion* or *complete case analysis*. In order to estimate using this method we may only use the first four rows in Table 34.2 in which $R_I = 0$. Table 34.3 shows the joint distribution estimated in this manner. However, notice that the information contained in the last two rows of Table 34.2 in which $R_I = 1$ has been left unused, thus resulting in wastage of samples [McKnight et al. 2007, Enders 2010]. Hence this procedure, while convenient and fast to implement, is not recommended in practice even if it guarantees consistent estimates. We describe below an alternate procedure that utilizes samples more efficiently.

As stated in Remark 34.1, any model that is MCAR is also MAR; hence, any estimation algorithm applicable to MAR is applicable to MCAR as well. Thus, to recover $P(G, I)$ given the MCAR graph in Figure 34.1(a), we could apply Equation (34.4) to obtain:

$$P(G, I) = P(I^*|G, r_I')P(G)$$

Table 34.3 **Complete case analysis–based estimation of joint distribution given the m-graph in Figure 34.1(a) and the data in Table 34.2**

G	I	$P(G,I)$
M	H	$\frac{p_1}{p_1+p_2+p_3+p_4}$
M	L	$\frac{p_2}{p_1+p_2+p_3+p_4}$
F	H	$\frac{p_3}{p_1+p_2+p_3+p_4}$
F	L	$\frac{p_4}{p_1+p_2+p_3+p_4}$

The estimand above dictates that we compute $P(I^*|G, r'_I)$ exclusively from samples in which I is observed and $P(G)$ from all samples, including those in which I is missing as shown in Table 34.4. Clearly, this utilizes data in a better manner compared to listwise deletion exemplified in Table 34.3. Efficient graph-based deletion procedures for MCAR and MAR that exploit available samples to a greater extent, thus yielding better quality estimates, are discussed in Van den Broeck et al. [2015].

34.3.2 Recoverability in MNAR Problems

In this subsection, we exemplify various recoverability techniques for MNAR using simple models.

34.3.2.1 Recovering $P(X, Y)$ Given the m-graph G in Figure 34.2(a)

G is one of the simplest examples of MNAR in which missingness in R_X is caused by Y, a variable with missing values. $V_m = \{X, Y\}$, $V_o = \{\}$ and due to the edge from Y to R_X, MAR does not hold, that is, $\{X, Y\} \not\perp\!\!\!\perp \{R_x, R_y\}$. Joint distribution $P(X, Y)$ is recoverable given G as shown below:

$$
\begin{aligned}
P(X,Y) &= P(X|Y)P(Y) \text{ (using chain rule)}\\
&= P(X|Y, r'_x, r'_y)P(Y|r'_y) \text{ (since } X \perp\!\!\!\perp R_x, R_y|Y \text{ and } Y \perp\!\!\!\perp R_y \text{ hold in } G)\\
&= P(X^*|Y^*, r'_x, r'_y)P(Y^*|r'_y) \text{ (using Equation 34.1)}
\end{aligned}
$$

We call the above technique sequential factorization [Mohan and Pearl 2018]. It is sensitive to the order of factorization. Had we factorized $P(X, Y)$ as $P(Y|X)P(X)$ in

Table 34.4 **A deletion-based method with less sample wastage for estimating joint distribution given the m-graph in Figure 34.1(a) and the data in Table 34.2**

G	I	$P(G,I)$
M	H	$\frac{p_1(p_1+p_2+p_5)}{p_1+p_2}$
M	L	$\frac{p_2(p_1+p_2+p_5)}{p_1+p_2}$
F	H	$\frac{p_3(p_3+p_4+p_6)}{p_3+p_4}$
F	L	$\frac{p_4(p_3+p_4+p_6)}{p_3+p_4}$

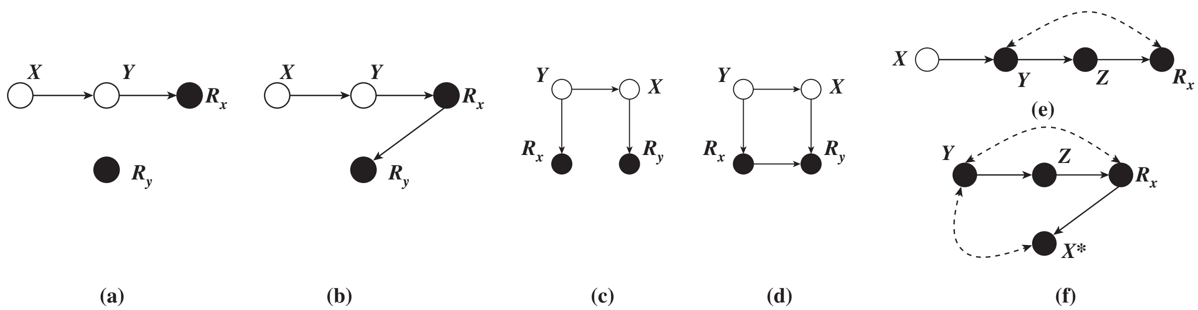

Figure 34.2 (a)–(e) m-graphs depicting MNAR missingness. Proxy variables have not been explicitly portrayed to keep the figures simple and clear. (f) Graph corresponding to m-graph (e) in which X is treated as a latent variable.

the first step, it would have been harder to establish recoverability. We further note that the estimand dictates that $P(X|Y)$ be estimated from samples in which both X and Y are observed and $P(Y)$ be estimated from samples in which Y is observed, regardless of the missingness status of X.

34.3.2.2 Recovering $P(X, Y)$ Given the m-graph in Figure 34.2(b)

For exactly the same reasons as those described in Section 34.3.2.1, this m-graph also depicts MNAR. However, notice that m-graphs in Figure 34.2(a) and (b) differ in the way the R variables are connected. An edge exists between the R variables in m-graph (b) whereas in (a) $R_x \perp\!\!\!\perp R_y$. We show below that this seemingly minor change results in a substantially different estimand (and estimation process).

$$\begin{aligned}
P(X, Y) &= P(X|Y)P(Y) \\
&= P(X|Y, r'_x, r'_y)P(Y) \text{ (since } X \perp\!\!\!\perp R_x, R_y|Y) \\
&= P(X|Y, r'_x, r'_y) \sum_{R_x} P(Y|R_x)P(R_x) \\
&= P(X|Y, r'_x, r'_y) \sum_{R_x} P(Y|R_x, r'_y)P(R_x) \text{ (since } Y \perp\!\!\!\perp R_y|R_x) \\
&= P(X^*|Y^*, r'_x, r'_y) \sum_{R_x} P(Y^*|R_x, r'_y)P(R_x) \text{ (using Equation 34.1)}
\end{aligned}$$

This example underscores the importance of modeling the causal relationship among R variables. For instance, had the m-graph been $X \rightarrow Y \rightarrow R_x \leftrightarrow R_y$, the estimand for $P(X, Y)$ would have been identical to the one derived in Section 34.3.2.1.

34.3.2.3 Recovering $P(X, Y)$ Given the m-graph in Figure 34.2(c)

The parents of both R variables in this m-graph are variables with missing values. Hence the m-graph depicts MNAR missingness. Recoverability of $P(X, Y)$ given this m-graph is discussed in Mohan et al. [2013] and the recoverability procedure presented therein forms the basis for most recoverability methods for MNAR. In this subsection we present an alternate method that requires inspecting all missingness patterns one by one.

$$
\begin{aligned}
P(X,Y) &= \sum_{R_x,R_y} P(X,Y,R_x,R_y) \\
&= P(X,Y,r'_x,r'_y) + P(X,Y,r'_x,r_y) \\
&\quad + P(X,Y,R_x=,r'_y) + P(X,Y,r_x,r_y)
\end{aligned}
$$

To prove recoverability of $P(X,Y)$, we will show that each term in the sum is recoverable. It follows from Equation (34.1) that $P(X,Y,r'_x,r'_y) = P(X^*,Y^*,r'_x,r'_y)$ and hence $P(X,Y,r'_x,r'_y)$ is recoverable. We will now show that $P(X,Y,r_x,r'_y)$ is recoverable.

$$
\begin{aligned}
P(X,Y,r_x,r'_y) &= P(X|Y,r_x,r'_y)P(Y|r_x,r'_y)P(r_x,r'_y) \\
&= P(X|Y,\mathbf{r}'_x,r'_y)P(Y|r_x,r'_y)P(r_x,r'_y) \text{ (since } X \perp\!\!\!\perp R_x|Y,R_y) \\
&= P(X^*|Y^*,r'_x,r'_y)P(Y^*|r_x,r'_y)P(r_x,r'_y) \text{ (using Equation 34.1)}
\end{aligned}
$$

In a similar manner we can show that $P(X,Y,r'_x,r_y) = P(Y^*|X^*,r'_x,r'_y)P(X^*|r'_x,r_y)$ $P(r'_x,r_y)$ and hence recoverable. What remains to be shown is that $P(X,Y,r_x,r_y)$ is recoverable.

$$
\begin{aligned}
P(X,Y,r_x,r_y) &= P(X|Y,r_x,r_y)P(r_x|Y,r_y)P(Y,r_y) \\
&= P(X|Y,\mathbf{r}'_x,r_y)P(r_x|Y,r_y)P(Y,r_y) && (34.5) \\
&= \frac{P(X,Y,r'_x,r_y)}{P(Y,r'_x,r_y)}P(r_x|Y,r_y)P(Y,r_y) \\
&= \frac{P(Y|X,r'_x,r_y)P(X,r'_x,r_y)}{P(r'_x|Y,r_y)P(Y,r_y)}P(r_x|Y,r_y)P(Y,r_y) \\
&= \frac{P(Y^*|X^*,r'_x,\mathbf{r}'_y)P(X^*,r'_x,r_y)}{P(r'_x|Y^*,\mathbf{r}'_y)}P(r_x|Y^*,\mathbf{r}'_y) && (34.6)
\end{aligned}
$$

In Equation (34.5), we replaced r_x with r'_x since $X \perp\!\!\!\perp R_x|Y,R_y$ holds in the graph. In Equation (34.6), we first cancelled out $P(Y,r_y)$ from the numerator and denominator, and then replaced r_y with r'_y in (i) $P(Y|X,r'_x,r_y)$ by applying $Y \perp\!\!\!\perp R_y|X,R_x$ and in (ii) $P(r'_x|Y,r_y)$ and $P(r_x|Y,r_y)$ by applying $R_x \perp\!\!\!\perp R_y|Y$. Finally, using Equation (34.1) we replaced Y with Y^* and X with X^*.

34.3.2.4 Recovering $P(X, Y)$ Given the m-graph in Figure 34.2(d)

The m-graph depicts MNAR for exactly the same reasons discussed in Section 34.3.2.4. Here we are recovering a conditional distribution as opposed to all previous examples of recoverability that discussed joint distributions.

$$P(X|Y) = P(X|Y, r'_x) \text{ (since } X \perp\!\!\!\perp R_x|Y)$$
$$= \frac{P(X, Y, r'_x)}{\sum_X P(X, Y, r'_x)} \quad (34.7)$$
$$P(X, Y, r'_x) = P(Y|X, r'_x)P(X, r'_x)$$
$$= P(Y|X, r'_x, r'_y)P(X, r'_x) \text{ (since } Y \perp\!\!\!\perp R_y|X, R_x)$$
$$= P(Y^*|X^*, r'_x, r'_y)P(X^*, r'_x) \quad (34.8)$$

Substituting the right-hand side (RHS) of Equation (34.8) in the place of $P(X, Y, r'_x)$ in Equation (34.7), we get

$$P(X|Y) = \frac{P(Y^*|X^*, r'_x, r'_y)P(X^*, r'_x)}{\sum_X P(Y^*|X^*, r'_x, r'_y)P(X^*, r'_x)}$$

34.3.2.5 Recovering $P(X)$ Given the m-graph in Figure 34.2(e)

The dotted bi-directed edge indicates that there exists a latent variable that is a parent of both Y and R_X, and this makes the model MNAR. This graph is different from all the other m-graphs that we have examined thus far. Notice that here, although X and R_X are not connected by an edge, there exists no separating set that can d-separate them. This is because there are two paths between X and R_X; on one path Y is a collider and Z, the descendant of a collider, and on the other path Y and Z are part of a chain. So, including Y or Z in the separating set will open the collider path, while excluding either one of them would leave the chain open. Interestingly, $P(X)$ is still recoverable as detailed below:

$$P(X) = P(X|do(z)) \text{ (using rule 3 of do-calculus [Pearl 2009])}$$
$$= P(X|do(z), r'_x) \text{ (using rule 1 of do-calculus [Pearl 2009])}$$
$$= P(X^*|do(z), r'_x) \text{ (using Equation 34.1)}$$

We have reduced the problem of recovering $P(X)$ to the problem of identifying the causal effect such that the causal query is defined over variables in the observed data distribution. Since the causal query is not a function of X, it can be identified using methods available in Shpitser and Pearl [2006] and the graph shown in Figure 34.2(f) in which X is treated as a latent variable.

Table 34.5 **Observed data distribution $P(X^*, R_x)$ corresponding to the m-graph $X \longrightarrow R_x$**

X	R_X	$P(X^*, R_X)$
0	0	0.3
1	0	0.5
m	1	0.2

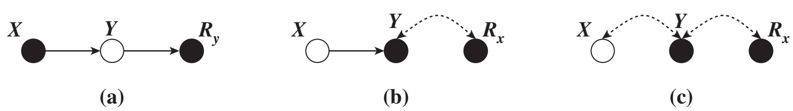

Figure 34.3 m-graphs in which $P(X, Y)$ is not recoverable.

Finally, we note that although in this chapter we focus on discrete variables, recoverability techniques exist for continuous variables and have been discussed in Pearl [2013] and Mohan et al. [2018].

34.3.3 Non-recoverability

Consider the problem of recovering $P(X)$ given the m-graph $G : X \longrightarrow R_x$. R_x is dependent on X and we have no additional information regarding this dependence. Table 34.5 presents a dataset generated by G. It could be that X is missing only when its value is 1 or it could be that X is missing only when its value is 0. In the former case $P(x') = 0.3$, whereas in the latter case $P(x') = 0.5$. Using the available information in G, it is not possible to find the (true) value of $P(X)$ even if we are given infinitely many samples, that is, $P(X)$ is non-recoverable. In fact, non-recoverability of $P(X)$ would persist even if G had more variables in it (formally proved in Mohan et al. [2013], Mohan and Pearl [2014a, 2014b]). In general, joint distribution is non-recoverable whenever there exists a variable X with missing values (i.e., $X \in V_m$) such that either:

1. X and R_X are neighbors or
2. X and R_X are connected by a path in which all intermediate nodes are colliders.

Thus, $P(X, Y)$ is non-recoverable in all the three m-graphs in Figure 34.3. However, in Figure 34.3(a) $P(X|Y)$ is recoverable, and in Figure 34.3(b) and (c) $P(X)$ is recoverable.

34.4 Testability

Testability when there is no missingness: When X and Y are fully observed variables, the independence statement $X \perp\!\!\!\perp Y$ is testable, that is, there exist distributions over X and Y in which $X \perp\!\!\!\perp Y$ does not hold. Therefore, given a graph G and a distribution $P(X, Y)$, if the graph portrays $X \perp\!\!\!\perp Y$ and the claim does not hold in the distribution, then we can conclude that the graph and distribution are not compatible. Thus, under no missingness, d-separations serve as testable implications of a graphical model [Pearl 2009].

Non-testability under missingness: The simplest missing data distribution is $P(X^*, R_X)$, which is obtained when the substantive variable of interest is a single variable X. Let the query to be recovered be $P(X)$. As shown in the previous sections, recoverability of $P(X)$ hinges on $X \perp\!\!\!\perp R_x$; if it holds then $P(X)$ is recoverable, otherwise not. Given the decisive nature of this independence, can we test it?

$X \perp\!\!\!\perp R_x$ is testable only if it is refutable in all true distributions that are compatible with the observed data distribution. However, for any observed data distribution $P(X^*, R_X)$, there exists a true distribution $P'(X, R_x)$ in which $X \perp\!\!\!\perp R_x$ holds. It can be constructed as $P'(X, Rx) = P(X^*|Rx = 0)P(R_x)$. Hence the claim is not refutable. Put differently, independence claims between a variable and its mechanism are not testable [Mohan and Pearl 2014b].

Testable implications of m-graphs: d-separations that abide by the following syntactic rules are testable under missingness (X and Y are singletons) [Mohan and Pearl 2014a].

$$X \perp\!\!\!\perp Y|Z, R_x, R_y, R_z \tag{34.9}$$

$$X \perp\!\!\!\perp R_y|Z, R_x, R_z \tag{34.10}$$

$$R_x \perp\!\!\!\perp R_y|Z, R_z \tag{34.11}$$

Example of testability: Figure 34.2(a) encodes the conditional independence $X \perp\!\!\!\perp R_y|R_x$, which matches the syntactic rule 34.10 above when $Z = \{\}$. It follows from $X \perp\!\!\!\perp R_y|R_x$ that:

$$P(X|r_y, r'_x) = P(X|r'_y, r'_x)$$

Using Equation (34.1) we can rewrite the above as,

$$P(X^*|r_y, r'_x) = P(X^*|r'_y, r'_x)$$

The preceding claim, which is defined over X^*, R_x and R_y, is testable given the observed data distribution. If the claim is violated, we conclude that the model and data are not compatible. Note that this test not only detects incompatibility but also helps in locating the source of incompatibility.

On the indispensability of causal assumptions: Let $G_1 : X \quad R_X$ and $G_2 : X \longrightarrow R_X$. G_1 encodes the assumption $X \perp\!\!\!\perp R_x$, whereas G_2 does not. Since $X \perp\!\!\!\perp R_X$ is not testable, G_1 and G_2 are statistically indistinguishable, that is, any given observed data distribution $P(X^*, R_X)$ compatible with G_1 is also compatible with G_2. However, they encode different causal assumptions. In G_1 where X does not cause its own missingness $P(X)$ is recoverable, whereas in G_2 where X causes its own missingness $P(X)$ is not recoverable. Thus, there exists no universal algorithm that can determine recoverability without examining the model and taking into account the embedded causal assumptions.

In conclusion, *missing data is a causal inference problem!*

References

C. Enders. 2010. *Applied Missing Data Analysis*. Guilford Press, London, New York.

R. Little and D. Rubin. 2002. *Statistical Analysis with Missing Data*. Wiley, New York.

P. McKnight, K. McKnight, S. Sidani, and A. Figueredo. 2007. *Missing Data: A Gentle Introduction*. Guilford Press, London, New York.

K. Mohan and J. Pearl. 2014a. On the testability of models with missing data. *Proceedings of AISTAT*. 33, 643–650.

K. Mohan and J. Pearl. 2014b. Graphical models for recovering probabilistic and causal queries from missing data. In *Advances in Neural Information Processing Systems 27*. Curran Associates, Inc., Red Hook, NY, 1520–1528.

K. Mohan and J. Pearl. 2018. Graphical models for processing missing data. *arXiv preprint arXiv:1801.03583*.

K. Mohan, J. Pearl, and J. Tian. 2013. Graphical models for inference with missing data. In *Advances in Neural Information Processing Systems 26*. Curran Associates, Inc., Red Hook, New York, 1277–1285.

K. Mohan, F. Thoemmes, and J. Pearl. 2018. Estimation with incomplete data: The linear case. In *IJCAI*. 5082–5088. DOI: https://doi.org/10.24963/ijcai.2018/705.

J. Pearl. 2009. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York. DOI: https://doi.org/10.1017/CBO9780511803161.

J. Pearl. 2013. Linear models: A useful "microscope" for causal analysis. *J. Causal Inference* 1, 1, 155–170. DOI: https://doi.org/10.1515/jci-2013-0003.

D. Rubin. 1976. Inference and missing data. *Biometrika* 63, 581–592. DOI: https://doi.org/10.2307/2335739.

I. Shpitser and J. Pearl. 2006. Identification of conditional interventional distributions. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*. 437–444.

G. Van den Broeck, K. Mohan, A. Choi, A. Darwiche, and J. Pearl. 2015. An efficient method for Bayesian network parameter learning from incomplete data. In *Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence*. 161–170.