

Sequential models for coarsening and missingness

Richard D. Gill
James M. Robins

ABSTRACT In a companion paper we described what intuitively would seem to be the most general possible way to generate Coarsening at Random mechanisms, a sequential procedure called *randomized monotone coarsening*. Counter-examples showed that CAR mechanisms exist which cannot be represented in this way. Here, we further develop these results in two directions. Firstly, we consider what happens when data is coarsened at random in two or more phases. We show that the resulting coarsening mechanism is not CAR anymore, but under suitable assumptions is identified and can provide interesting alternative analysis of data under a non-CAR model. Secondly, we look at sequential mechanisms for generating MAR data, missing components of a multivariate random vector. Randomised monotone missingness schemes, in which one variable at a time is observed and depending on its value, another variable is chosen or the procedure is terminated, supply in our opinion the broadest class of physically interpretable MAR mechanisms. We show that every randomised monotone missingness scheme can be represented by a Markov monotone missingness scheme, in which the choice of which variable to observe next only depends on the set of previously observed variables and their values, not on the sequence in which they were measured. We also show that MAR mechanisms exist which cannot be represented sequentially.

1. Sequential CAR is not CAR but it is everything

We use the notation of our companion paper Gill, van der Laan and Robins (1997). In particular \mathcal{X} is a coarsening of an underlying random variable X . We suppose that X takes values in a finite set E . Its power set (the set of all subsets of E) is denoted by \mathcal{E} . So \mathcal{X} takes values in $\mathcal{E} \setminus \{\emptyset\}$ and $X \in \mathcal{X}$ with probability one.

In this section, we study k -sequential coarsening, by which we mean that the observed random set \mathcal{X} is formed by first coarsening the underlying random variable X to a random set \mathcal{X}_1 , then further coarsening \mathcal{X}_1 to a new random set \mathcal{X}_2, \dots , and finally coarsening \mathcal{X}_k to \mathcal{X} . Only the final random set \mathcal{X} is observed. We say the coarsening is k -sequentially CAR if each of the $k+1$ sequential coarsenings are individually CAR. Robins (1996) introduced sequential CAR in the special case of missing components of a multivariate random vector which he called generalised permutation missingness and we call sequential MAR. We will show that $\mathcal{X}_1, \dots, \mathcal{X}_k, \mathcal{X}$ being k -sequentially CAR does not imply that the overall coarsening of X to \mathcal{X} is CAR. Thus,

modeling the observed data \mathcal{X} as CAR versus k -sequentially CAR implies different marginal distributions for the underlying variable X . Furthermore, we will show that, in the discrete case, k -sequential CAR is, like ordinary CAR, everything, so that no data evidence can distinguish CAR from k -sequential CAR. We will argue that, in many contexts, the assumption that the coarsening mechanism is k -sequential CAR rather than CAR has greater substantive plausibility, in which case the data should be analyzed under a k -sequential CAR model. We now formalize these ideas.

Definition. We say that the random sets $\mathcal{X}_1, \dots, \mathcal{X}_k, \mathcal{X}$ with each \mathcal{X}_m and $\mathcal{X} \in \mathcal{E} \setminus \{\emptyset\}$ form a k -sequential coarsening of a random variable X if for $m = 0, \dots, k+1$, $\mathcal{X}_m \subseteq \mathcal{X}_{m+1}$ with probability 1 where $\mathcal{X}_0 \equiv \{X\}$ and $\mathcal{X}_{k+1} \equiv \mathcal{X}$. Thus a 0-sequential coarsening is just the ordinary coarsening studied above.

Definition. A k -sequential coarsening is k -sequentially CAR if, for $m = 1, \dots, k$, the conditional distribution of \mathcal{X}_m given \mathcal{X}_{m-1} does not depend on the particular realization of \mathcal{X}_{m-1} except through the fact that it is compatible with \mathcal{X}_m . In the discrete case, this means $\Pr(\mathcal{X}_m = A \mid \mathcal{X}_{m-1} = B)$ is the same for all B in the support of \mathcal{X}_{m-1} such that $B \subseteq A$.

Throughout we suppose we only observe the final random set $\mathcal{X} = \mathcal{X}_{k+1}$ of a k -sequential coarsening. One might hope that \mathcal{X} is a CAR coarsening of X (in the previous sense) if $\mathcal{X}_1, \dots, \mathcal{X}_{k+1}$ is k -sequentially CAR. However, this is not the case as the following example indicates.

Example. Suppose $X = (Y, W)$ where Y and W take values in $\{0, 1\}$. Let $E_w = \{(0, w), (1, w)\}$ denote the event that $W = w$ but Y is not observed. Let $E_y = \{(y, 0), (y, 1)\}$ denote an event in which W is not observed. Finally, let the entire sample space $E = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ be the event that neither Y nor W is observed. Now suppose that $\Pr(\mathcal{X}_1 = \{(y, w)\} \mid X = (y, w)) = c_1(w)$, $\Pr(\mathcal{X}_1 = E_w \mid X = (y, w)) = 1 - c_1(w)$ so that the probability that Y is not observed in \mathcal{X}_1 depends on the observed value of W . Further, suppose $\Pr(\mathcal{X}_2 = \{(y, w)\} \mid \mathcal{X}_1 = \{(y, w)\}) = c_2(y)$, $\Pr(\mathcal{X}_2 = E_y \mid \mathcal{X}_1 = \{(y, w)\}) = 1 - c_2(y)$ so the probability w is not observed in \mathcal{X}_2 depends on the observed value of y in \mathcal{X}_1 . Suppose $\Pr(\mathcal{X}_2 = E_w \mid \mathcal{X}_1 = E_w) = c_3$ and $\Pr(\mathcal{X}_2 = E \mid \mathcal{X}_1 = E_w) = 1 - c_3$ so the probability that W is not observed in \mathcal{X}_2 when Y is not observed in \mathcal{X}_1 is constant. It is easy to check that $\mathcal{X}_1, \mathcal{X}_2$ is 1-sequentially CAR, more specifically, it is 1-sequentially MAR. However, $\mathcal{X} = \mathcal{X}_2$ is not a CAR coarsening of X since $\Pr(\mathcal{X}_2 = E_y \mid X = (y, w)) = c_1(w)(1 - c_2(y))$ which depends on w .

However, if X is discrete, k -sequential CAR is still everything as formally stated in the following theorem and its corollary. The theorem assumes that the possible values of the intermediate coarsening \mathcal{X}_m are specified in advance and moreover that each value arises in one particular way from its predecessor \mathcal{X}_{m-1} . This was in fact the case in the example just discussed, as the reader can easily check. These a priori restrictions on the

intermediate coarsenings ensure that a k -sequential CAR model reproducing a particular distribution of observed coarsened data \mathcal{X} not only exists, but is also essentially unique.

Theorem. Let E be a finite set and let \mathcal{X} be a random nonempty set with distribution f_A , $A \in \mathcal{E}_{k+1} = \mathcal{E} \setminus \{0\}$. Let $\mathcal{E}_0 = \{\{x\} : x \in E\}$ be the subset of \mathcal{E} consisting of all the singletons. Let $\mathcal{E}_1, \dots, \mathcal{E}_k$ be given subsets of \mathcal{E} such that, for each $m = 0, \dots, k$ and for each $A \in \mathcal{E}_{m+1}$, A is a disjoint union of specified elements B of \mathcal{E}_m . Write $B \in_m A$ if $B \in \mathcal{E}_m$ is one of the specified sets partitioning a given $A \in \mathcal{E}_{m+1}$. Then there exist ‘CAR probabilities’ π_{mA} for $A \in \mathcal{E}_{m+1}$ for $m = 0, \dots, k$, and distributions p_{mB} over $B \in \mathcal{E}_m$ for $m = 0, \dots, k+1$, with $p_{(k+1)A} = f_A$, satisfying (i) for each $B \in \mathcal{E}_m$, $\sum \pi_{mA} = 1$ with summation over $\{A \in \mathcal{E}_{m+1} : B \in_m A\}$; (ii) $\sum p_{mB} = 1$ with summation over all $B \in \mathcal{E}_m$; and (iii) $p_{(m+1)A} = p_{mA} \pi_{mA}$ for each $m \leq k$ and each $A \in \mathcal{E}_{m+1}$ and $p_{mA} = \sum p_{mB}$ with summation over $\{B \in \mathcal{E}_m : B \in_m A\}$. Furthermore, for each $A \in \mathcal{E}_{k+1}$ with $f_A > 0$, π_{kA} and p_{kA} are unique. Similarly, for each $A \in \mathcal{E}_{m+1}$ with $p_{(m+1)A} > 0$, π_{mA} and p_{mA} are unique. Thus if $f_A > 0$ for all singletons $A = \{x\}$, then $p_x = p_{0\{x\}}$, $x \in E$, is a uniquely defined distribution on E .

Proof. This follows by repeated application of the main theorem of Section 2 of our companion paper, starting from $m = k+1$ downwards.

Corollary. If we draw $\mathcal{X}_0 = \{X\}$ from $p_{0\{x\}}$, $x \in E$, and then recursively draw \mathcal{X}_{m+1} from \mathcal{E}_{m+1} under the conditional law π_{mA} given $\mathcal{X}_m = B$ (where B is the value of \mathcal{X}_m drawn in the previous step and A such that $B \in_m A$), then $\mathcal{X}_1, \dots, \mathcal{X}_{k+1}$ will be a k -sequential CAR coarsening of X and $\mathcal{X} \equiv \mathcal{X}_{k+1}$ will have marginal distribution f_A .

Thus, for discrete E , one cannot use the data to distinguish whether the underlying coarsening process is CAR or is k -sequentially CAR based on specified sets of intermediate coarsened values $\mathcal{E}_1, \dots, \mathcal{E}_k$ and relations between them. However, a CAR process compatible with the observed data \mathcal{X} will imply a different marginal distribution for the underlying variable X than will a k -sequential CAR process compatible with the data. Thus the choice of modeling the observed data \mathcal{X} as CAR versus k -sequentially CAR will depend solely on one’s prior subject matter beliefs. As the following example due to Robins (1996) shows, the assumption that the coarsening process is k -sequentially CAR can be much more substantively plausible than that it is CAR.

Example. Suppose in a prison, a voluntary HIV test is offered to prisoners. Thirty percent agree to the test, of which fifteen percent are positive. It is believed that the decision to be tested (i) is not independent of HIV status due to the differing views on testing held by HIV-positive versus negative inmates but (ii) is independent of HIV conditional on whether a prisoner has sought previous treatment for an STD in the prison hospital. The goal is to estimate the marginal probability of HIV infection among

the inmates. Thus, in order to identify this probability, it is decided to randomly sample hospital records and record previous STD treatment with a prisoner's probability of being included in the hospital record sample depending on whether the prisoner agreed to the HIV test and, if so, his test result. Letting Y and W be the dichotomous indicator variables for HIV infection status and previous STD treatment, the coarsening mechanism is exactly that discussed in the previous example, implying the data is 1-sequential CAR but not CAR.

We note that a k -sequential CAR process and an m -sequential CAR process would also imply different marginal distributions for the underlying variable X and, by our theorem, cannot be distinguished from one another by any data evidence. Indeed, two k -sequential CAR processes based on different sequences of selected sets will imply different marginal distributions for the underlying X and also cannot be distinguished from one another based on the data. Thus, multiple sequential CAR models can be used to conduct a sensitivity analysis to help determine the sensitivity of the distribution of the underlying variable X to assumptions about the coarsening mechanism.

2. MAR and randomized monotone missingness

We now restrict attention from coarsened to missing data, and investigate the possibility that all MAR mechanisms can be constructed by random monotone coarsening at the variables level. Recall that in section 3 of our companion paper we showed that, for CAR, this programme fails. However our counter-example there cannot be used in the new situation.

The sample space E for the underlying data X is now a Cartesian product of k finite sets E_i , $i \in \mathcal{K} = \{1, \dots, k\}$. Write the coordinates of X as $X = (X_1, \dots, X_k)$. The only sets A which can be observed, i.e., with $\pi_A > 0$, are Cartesian products either of the whole coordinate E_i or of a singleton $\{x_i\} \in E_i$, $i \in \mathcal{K}$. Knowing $X \in A$ is the same as knowing exactly the values of some of the X_i but knowing nothing about the others. We say that the first X_i are *observed*, the others are *missing*. In this situation, coarsening at random is called *missing at random* (MAR), Rubin (1976) or Little and Rubin (1987), and one can define *monotone missingness*, *randomized monotone missingness*, and *Markov randomized monotone missingness* analogously to the CAR case. In monotone schemes, rather than choosing at each stage (if we do not stop) a partition of the current set, we now choose a further variable to observe.

The counter-example which we gave in the CAR case to the conjecture that all CAR mechanisms can be represented by randomized monotone coarsening schemes does not give a counter-example to the analogous conjecture for MAR. It took us some effort to find a new counter-example for MAR, though at the end, as the reader will discover, the counter-example is quite simple. We will first show that any set of MAR probabilities which

admit a randomized monotone representation, also admit a Markov randomized monotone representation. So it is sufficient to investigate the possible ‘Markov monotone missingness’ or MMM property of a given MAR system. This can be done by solving a certain *linear programming problem*. So one can find out, in a finite number of steps, if a specific MAR system admits an MMM representation. Finding a representation comes down to solving a set of linear equalities in more unknowns than there are equations (so typically many solutions exist) subject to a collection of linear inequalities (which rule out some or even all solutions).

Recall we have $X = (X_1, \dots, X_k)$ where X_i takes values x_i in a finite set E_i . Let \mathcal{R} denote the set of labels of actually observed variables: so \mathcal{R} is a random subset of \mathcal{K} (the letter \mathcal{R} stands for ‘recorded’, the other variables are missing). Observation of \mathcal{X} is equivalent to observation of the pair $(\mathcal{R}, X_{\mathcal{R}})$ where $X_{\mathcal{R}}$ denotes the vector (a subvector of X) with components X_i for $i \in \mathcal{R}$. Possibly \mathcal{R} is empty. Under MAR, the probability to observe the variables \mathcal{R} , given $X = x$, only depends on the X_i with $i \in \mathcal{R}$, or:

$$\Pr\{\mathcal{R} = R | X = x\} = \Pr\{\mathcal{R} = R | X_R = x_R\}$$

We denote this probability by $\pi_R^{x_R}$. Thus a MAR mechanism is determined by a system of probabilities $\pi_R^{x_R}$ such that for all $x \in E = E_1 \times \dots \times E_k$,

$$\sum_{R \subseteq \mathcal{K}} \pi_R^{x_R} = 1 \tag{1}$$

Now we specify a randomized monotone missingness (RMM) mechanism. This consists of a set of probability rules for choosing a sequence of distinct indices $I_1, I_2, \dots, I_K \in \mathcal{K}$ where $0 \leq K \leq k$ is the random number of variables finally observed and $\mathcal{R} = \{I_1, \dots, I_K\}$. We shall summarize the rules in a big collection of probabilities ($\tilde{\psi}$); later we will write (ψ), without the tilde, to specify a (less complicated) Markov mechanism. To keep the formulas short, we write $i_{<n}$ for i_1, \dots, i_{n-1} . The *set* of elements $\{i_1, \dots, i_{n-1}\}$ is then denoted by $\{i_{<n}\}$. The *vector* or *sequence* (i_1, \dots, i_{n-1}) may be written $(i_{<n})$ for emphasis. We drop the curly or round brackets when the distinction between set and sequence is not crucial. When we write events like ‘ $I_n = i_n$ ’ it is implicitly understood that ‘ $K \geq n$ ’ also holds, and conversely, ‘ $K < n$ ’ implies ‘ I_n is undefined’.

Suppose at the n ’th stage we have already determined $I_{<n} = i_{<n}$ and have observed $X_{<n} = x_{<n}$. The order in which the variables were selected is included in the expression $i_{<n}$. Next, with conditional probability

$$\begin{aligned} \tilde{\psi}_{i_n | i_{<n}}^{x_{i_{<n}}} &= \Pr\{\text{At stage } n \text{ go on to observe variable } i_n \mid \text{past}\} \\ &= \Pr\{I_n = i_n | I_{<n} = i_{<n}, X_{i_{<n}} = x_{i_{<n}}\}, \end{aligned}$$

we choose and observe another variable X_{i_n} , $i_n \notin \{i_{<n}\}$. With the comple-

mentary probability

$$1 - \sum_{i_n} \tilde{\psi}_{i_n|i_{<n}}^{x_{i_{<n}}} = \Pr\{I_n \text{ is not defined} \mid \text{past}\}$$

we stop and set $K = n - 1$ and thus $\mathcal{R} = \{I_{<n}\}$.

Now define

$$\tilde{\phi}_{i_{<n}, i_n}^{x_{i_{<n}}} = \Pr\{I_{<n} = i_{<n}, I_n = i_n \mid X_{i_{<n}} = x_{i_{<n}}\}$$

and

$$\begin{aligned} \tilde{\xi}_{i_{<n}}^{x_{i_{<n}}} &= \tilde{\phi}_{i_{<n-1}, i_{n-1}}^{x_{i_{<n-1}}} - \sum_{i_n} \tilde{\phi}_{i_{<n}, i_n}^{x_{i_{<n}}} \\ &= \Pr\{K = n - 1, I_{<n} = i_{<n} \mid X_{i_{<n}} = x_{i_{<n}}\}. \end{aligned} \tag{2}$$

Then

$$\tilde{\psi}_{i_n|i_{<n}}^{x_{i_{<n}}} = \tilde{\phi}_{i_{<n}, i_n}^{x_{i_{<n}}} / \tilde{\phi}_{i_{<n-1}, i_{n-1}}^{x_{i_{<n-1}}} \tag{3}$$

showing that the RMM system can be described in terms of the system of probabilities $(\tilde{\phi})$ instead of the system $(\tilde{\psi})$. An advantage of this is that the MAR probabilities (π) can be expressed *linearly* in $(\tilde{\phi})$: for any given $R \subseteq \mathcal{K}$ of size $n - 1$,

$$\pi_R^{x_R} = \sum_{i_{<n}: \{i_{<n}\} = R} \tilde{\xi}_{i_{<n}}^{x_{i_{<n}}} \tag{4}$$

where the $(\tilde{\xi})$ are given by (2). The $(\tilde{\phi})$ are also convenient because they can be chosen freely up to a collection of *linear* restrictions:

$$\left. \begin{aligned} \tilde{\phi}_{i_{<n}, i_n}^{x_{i_{<n}}} &\geq 0 \\ \sum_{i_n} \tilde{\phi}_{i_{<n}, i_n}^{x_{i_{<n}}} &\leq \tilde{\phi}_{i_{<n-1}, i_{n-1}}^{x_{i_{<n-1}}} \end{aligned} \right\} \tag{5}$$

where for $n = 1$ the right hand side of the last inequality is replaced by the number 1.

It is easy to check that any system $(\tilde{\phi})$ satisfying (5) leads via (2) and (4) to a system (π) satisfying (1) and $\pi_R^{x_R} \geq 0$ for all R and x . Equation (3) shows how to calculate the actual decision rule probabilities $(\tilde{\psi})$ from the $(\tilde{\phi})$. Conversely, $(\tilde{\phi})$ can be computed recursively from $(\tilde{\psi})$ by rewriting (3) as

$$\tilde{\phi}_{i_{<n}, i_n}^{x_{i_{<n}}} = \tilde{\phi}_{i_{<n-1}, i_{n-1}}^{x_{i_{<n-1}}} \tilde{\psi}_{i_n|i_{<n}}^{x_{i_{<n}}}$$

where for $n = 1$ this is interpreted as $\tilde{\phi}_{\cdot, i_1} = \tilde{\psi}_{i_1|\cdot}$.

Our question: ‘does any MAR mechanism admit representation by a RMM scheme?’ boils down therefore to the question: given a system of nonnegative numbers (π) satisfying (1), do the linear equations (2) and

(4) admit a solution ($\tilde{\phi}$) satisfying the linear constraint (5)? We first show that if a solution exists, then one can also find a Markov monotone missing (MMM) solution by which we mean that the decision probabilities $\tilde{\psi}_{i_n|i_{<n}}^{x_{i_{<n}}}$ (to observe at stage n variable i_n , given the previous sequence of choices and observed values) are invariant under permutations of $i_{<n}$.

3. Markov monotone missingness

Under what we call Markov monotone missingness (MMM), the decision probabilities $\tilde{\psi}$ are invariant under permutations. We denote then, for a subset R of $n - 1$ elements of \mathcal{K} ,

$$\psi_{i_n|R}^{x_R} = \Pr\{I_n = i_n \mid \{I_{<n}\} = R, X_R = x_R\},$$

$$\phi_{R,i_n}^{x_R} = \Pr\{I_n = i_n, \{I_{<n}\} = R \mid X_R = x_R\}.$$

So $\phi_{R,i_n}^{x_R}$ is the *sum*, over all permutations $i_{<n}$ of the elements of R , of $\tilde{\phi}_{i_{<n},i_n}^{x_R}$; while $\psi_{i_n|R}^{x_R}$ is *equal* to each of the formerly introduced $\tilde{\psi}_{i_n|i_{<n}}^{x_R}$ for all permutations of R .

Now suppose a given MAR system (π) has a, not necessarily Markov, randomized monotone missingness (RMM) representation characterized by decision probabilities ($\tilde{\psi}$) or equivalently by a system ($\tilde{\phi}$). *Define* for each set R of $n - 1$ elements the quantities

$$\phi_{R,i_n}^{x_R} = \sum_{i_{<n}:\{i_{<n}\}=R} \tilde{\phi}_{i_{<n},i_n}^{x_{i_{<n}}}. \quad (6)$$

Combining (2) and (4) gives

$$\pi_R^{x_R} + \sum_{j \notin R} \phi_{R,j}^{x_R} = \sum_{j \in R} \phi_{R \setminus \{j\},j}^{x_{R \setminus \{j\}}}. \quad (7)$$

Moreover, *define*

$$\psi_{i_n|R}^{x_R} = \phi_{R,i_n}^{x_R} / \sum_{j \in R} \phi_{R \setminus \{j\},j}^{x_{R \setminus \{j\}}}. \quad (8)$$

The new (ψ) form a system of MMM decision rules. The new (ϕ) can be recovered from (ψ) by recursive use of (8), and since (7) holds we have hereby constructed a Markov representation for the original RMM-generated MAR probabilities (π). Here we use the fact that for a Markov system, described equivalently by a system (ϕ) or a system (ψ) because of the equivalence (8), the MAR probabilities can be calculated from (7); this fact follows from (2) and (4) again.

In conclusion, if a given MAR system can be represented by a randomized monotone missingness scheme, then it can also be represented by a Markov monotone missingness scheme.

4. Is every MAR system MMM?

In view of the results so far, the question to be answered is now: given any system of numbers $(\pi_R^{x_R})$ such that

$$\pi_R^{x_R} \geq 0 \quad \forall R, x_R \quad (9)$$

$$\sum_{R \subseteq \mathcal{K}} \pi_R^{x_R} = 1 \quad \forall x \quad (10)$$

do there exist $(\phi_{R,j}^{x_R})$ for $j \notin R \subseteq \mathcal{K}$ such that

$$\phi_{R,j}^{x_R} \geq 0 \quad \forall R, x_R, j \notin R \quad (11)$$

$$\pi_R^{x_R} + \sum_{j \notin R} \phi_{R,j}^{x_R} = \sum_{j \in R} \phi_{R \setminus \{j\}, j}^{x_R \setminus \{j\}} \quad \forall R, x_R \quad (12)$$

If so we can define

$$\psi_{j|R}^{x_R} = \phi_{R,j}^{x_R} / \sum_{j \in R} \phi_{R \setminus \{j\}, j}^{x_R \setminus \{j\}} \quad (13)$$

and the system $(\psi_{j|R}^{x_R})$ is a set of Markov monotone missingness decision probabilities which generates $(\pi_R^{x_R})$. Note that when R is the empty set \emptyset , (12) is rewritten as

$$\pi_\emptyset + \sum_{j \in \mathcal{K}} \phi_{\emptyset, j} = 1.$$

Because of this, on adding (12) over all $R \subseteq \mathcal{K}$ for given x , we obtain (10). Note, when $R = \emptyset$, equation (13) is rewritten as $\psi_{j|\emptyset} = \phi_{\emptyset, j}$.

Collect all $\pi_R^{x_R}$ and $\phi_{R,j}^{x_R}$ into column vectors $\boldsymbol{\pi}$, $\boldsymbol{\phi}$. Equation (12) can be rewritten in matrix form as $A\boldsymbol{\phi} = \mathbf{b}$ where \mathbf{b} coincides with $\boldsymbol{\pi}$ except in the ' \emptyset ' component, where π_\emptyset is replaced by $1 - \pi_\emptyset$. A is a matrix of 0's, 1's and -1 's. The question is now: for given A and \mathbf{b} , does there exist $\boldsymbol{\phi} \geq \mathbf{0}$ (componentwise inequality) such that $A\boldsymbol{\phi} = \mathbf{b}$? This problem can be solved in a specific instance by transforming it into a linear program, solvable in a finite number of steps by the simplex algorithm. Note that we also have $\mathbf{b} \geq \mathbf{0}$ and let $\mathbf{1}$ denote a vector of 1's of appropriate length. Consider the problem:

$$\text{maximize } -\mathbf{1}^\top \mathbf{z}$$

over $\mathbf{z} \geq \mathbf{0}$ subject to $A\mathbf{x} + \mathbf{z} = \mathbf{b}$, for some $\mathbf{x} \geq \mathbf{0}$; equivalently, the standard LP problem

$$\text{maximize } (-\mathbf{1}^\top; \mathbf{z}^\top) \begin{pmatrix} \mathbf{z} \\ \mathbf{x} \end{pmatrix}$$

$$\text{over } \begin{pmatrix} \mathbf{z} \\ \mathbf{x} \end{pmatrix} \geq \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \text{ subject to } (I; A) \begin{pmatrix} \mathbf{z} \\ \mathbf{x} \end{pmatrix} = \mathbf{b}.$$

Since $\mathbf{b} \geq \mathbf{0}$,

$$\begin{pmatrix} \mathbf{z} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix}$$

is a feasible solution, so the simplex algorithm can be initialized. Since $-\mathbf{1}^\top \mathbf{z} \leq 0$ for all $\mathbf{z} \geq \mathbf{0}$, the solution is finite. The simplex algorithm finds it in a finite number of steps. If at the solution the object function $-\mathbf{1}^\top \mathbf{z} = 0$, so also $\mathbf{z} = \mathbf{0}$, then we have found a solution $\phi = \mathbf{x} \geq \mathbf{0}$ to the equations $A\phi = \mathbf{b}$. If conversely there exists a solution $\phi \geq \mathbf{0}$ to $A\phi = \mathbf{b}$ then the maximization problem is solved by $\mathbf{x} = \phi$, $\mathbf{z} = \mathbf{0}$, with maximum 0. So: a maximizer of the linear program yields a maximum value 0 if and only if $A\phi = \mathbf{b}$ admits a nonnegative solution.

MAR systems are represented by vectors $\boldsymbol{\pi}$ satisfying $\boldsymbol{\pi} \geq \mathbf{0}$, $\sum_R \pi_R^{x_R} = 1$ for all x . This means that (for given $E = E_1 \times \dots \times E_k$) the set of all MAR systems is a *convex, compact polytope*. The set of all MMM systems is the subset of such $\boldsymbol{\pi}$ which can be written as

$$A\phi = \mathbf{b}, \quad b_R^{x_R} = \pi_R^{x_R} \quad \forall R \neq \emptyset, \quad b_\emptyset = 1 - \pi_\emptyset.$$

for some $\phi \geq \mathbf{0}$. This is also a convex, compact polytope, contained in the former one.

The system $A\phi = \mathbf{b}$, \mathbf{b} given, ϕ unknown, always has solutions. We have a consistent set of equations with more unknowns than equations. This means that it is not unreasonable, on the basis of just counting parameters, to hope that all MAR systems are MMM. However, the inequality constraint $\phi \geq \mathbf{0}$ may rule out many or even all solutions of $A\phi = \mathbf{b}$.

Our geometric picture gives a procedure for determining whether or not, for given $E = E_1 \times \dots \times E_k$, all MAR systems are MMM: inspect the *extremal points*—the *vertices*—of the set of MAR probabilities. If each vertex has an MMM representation, then the whole set does. Conversely, if any set of MAR probabilities does not admit an MMM representation, then at least one vertex also yields such a counterexample.

The next question is: how can we enumerate the extremal points $\boldsymbol{\pi}$ in the convex, compact polytope of $\boldsymbol{\pi}$ satisfying (9) and (10)? One way to do this is to select a sufficient number of the inequalities (9), turn them into equalities, and try to solve the resulting augmentation of the equalities (10). The procedure is ‘finite’ but already in the smallest non-trivial case— $k = 3$ binary variables—computationally unfeasible, since millions of cases have to be examined. We need to take account of symmetries and other special features to reduce the number of cases to be investigated.

Let us first study a special subcollection of extremal points: $\boldsymbol{\pi}$ that are *vectors of 0’s and 1’s only*. If we can find a counter-example here, we are done already. These extremal points, as CAR models, are deterministic (or Coarsened Completely at Random): the coarsening is generated by a fixed partition of the sample space.

In the case of 3 binary variables we enumerated all deterministic extremal points in MAR space, up to symmetries. Taking any particular $\pi_R^{x_R}$ equal to

1 forces all other $\pi_R^{x_R}$ in the equations (10) in which the given one appears to equal zero. A $\pi_R^{x_R}$ with R ‘small’ appear in many equations—for instance, π_\emptyset appears in all—so the enumeration is quite easy, for instance, by ordering according to the minimum number of elements in R with $\pi_R^{x_R} = 1$. We found in this way just 17 different types (i.e., up to permutation of variables and/or values of variables).

Of these 17, all but one were monotone. The exception had exactly

$$\pi_{23}^{11} = \pi_{13}^{10} = \pi_{12}^{00} = \pi_{123}^{101} = \pi_{123}^{010} = 1.$$

In words, if variables 1 and 2 both take the value 0, variable 3 is not observed; if variables 1 and 3 take the values 1 and 0 respectively, variable 2 is not observed; and if variables 2 and 3 both take the value 1, variable 1 is not observed. Otherwise all three variables are observed. Note that each variable has a chance to be observed, but also each variable has a chance *not* to be observed. That makes it obvious that this MAR system does not have a monotone missingness representation since there is no way any monotone procedure (selecting variables to be observed one by one, without knowing their values in advance) can start.

The existence of this example produces uneasiness as to the interpretation of MAR. Consider a computer simulation model, in which first X is generated, and then \mathcal{R} (hence also $X_{\mathcal{R}}$) according to given MAR probabilities (π). Let us suppose that after X has been realised, the computer has to generate the random indices \mathcal{R} . The computer program is allowed to look at values of components of X one by one, and on the basis of these values decide which values to report. We have shown that MAR probabilities exist which, in order to be simulated, require the computer to look at values of components of X which are *omitted* from the collection of observed variables \mathcal{R} . Still, given $\mathcal{R} = R$ and $X_R = x_R$, $X_{\mathcal{K}\setminus R}$ has the same distribution as when one only conditions on $X_R = x_R$ so the information about $X_{\mathcal{K}\setminus R}$ which was used in the procedure is not in the least revealed ‘statistically’.

We believe that the only plausible natural MAR mechanisms have a randomised monotone representation. If this belief is correct then, in practical applications, an analyst should perform a test of the hypothesis that the missingness process has such a representation. Robins and Gill (1996) provide an appropriate test statistic. If the test rejects, likelihood analysis under the MAR assumption should be avoided altogether or understood to be at best an approximation to a more appropriate analysis under a non MAR model. We note that the set of RMM probabilities is a closed subset of the set of MAR probabilities, so there exist tests that are consistent under the alternative.

Research remains to be done to support these arguments. Development of special methods for use in RMM models and for testing the RMM assumption only makes sense if the non-RMM MAR models are not highly special rarities as the number of variables and the number of their values becomes larger. As the number of variables increases, are the non-RMM

MAR models more and more of an exception or more and more the rule? On the other side, how convincing is our claim that the notion of RMM captures exactly the class of honest MAR mechanisms (the MAR mechanisms which one can expect to meet with in practice)?

Bibliography

- R.D. Gill, M.J. van der Laan and J.M. Robins (1997), Coarsening at random: characterizations, conjectures, counter-examples, *Proc. First Seattle Symposium on Biostatistics: Survival Analysis*, ed. D.Y. Lin, Springer-Verlag.
- R.J.A. Little and D.B. Rubin (1987), *Statistical Analysis with Missing Data*, Wiley, New York.
- J.M. Robins (1996), Non-response models for the analysis of non-monotone non-ignorable missing data. *Statistics in Medicine*, Special Issue, to appear.
- J.M. Robins and R.D. Gill (1996), Non-response models for the analysis of non-monotone ignorable missing data, *Statistics in Medicine*.
- D.B. Rubin (1976), Inference and missing data, *Biometrika* **63**, 581–592.

Authors:

R.D.Gill, Mathematical Institute, University Utrecht, PO Box 80 010, 3508 TA Utrecht, Netherlands; email: gill@math.ruu.nl

J.M. Robins, Depts of Epidemiology and Biostatistics, Harvard School of Public Health, 677 Huntington Road, Boston Ma. 02115, USA; email: robins@hsph.harvard.edu