Optimal nonparametric testing of Missing Completely At Random, and its connections to compatibility

Tom Berrett (joint work with Richard Samworth)
University of Warwick

Structural Breaks and Shape Constraints ICMS Workshop

May 20th, 2022

Missing data

One of the most common gaps between theory and practice is missing data.

With incomplete data traditional methodology often becomes inapplicable, uninterpretable or unreliable.

The best solution to handle missing data is to have none.

– R.A. Fisher

Nevertheless, recent work has introduced methodology and theory for modern statistical problems (Loh & Wainwright, 2012; Loh & Tan, 2018; Zhu, Wang & Samworth, 2019; Elsener & van de Geer, 2019; Cai & Zhang, 2019; Follain, Wang & Samworth, 2022).

Missingness mechanisms

When data is missing, some quantities are fundamentally out of reach. We must make assumptions about the mechanisms causing the missingness.

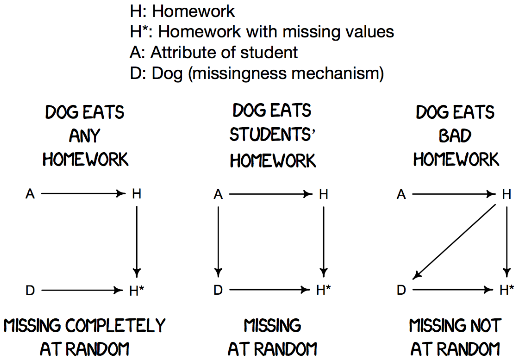


Image: Richard McElreath

The simplest setting is where data are MCAR, and when this holds the analysis is much easier and more interpretable.

Formal setting

Variables of interest X taking values in $\mathcal{X} = \prod_{j=1}^{d} \mathcal{X}_j$ and Ω taking values in $\{0,1\}^d$ indicating which variables are observed. Writing

$$(x \circ \omega)_j = \begin{cases} x_j & \text{if } \omega_j = 1 \\ \star & \text{if } \omega_j = 0 \end{cases},$$

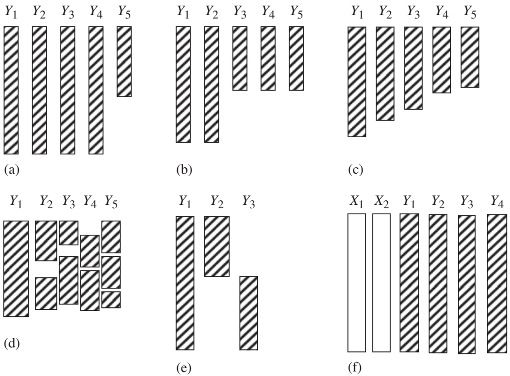
we observe i.i.d. copies of the random vector $X \circ \Omega$. MCAR says that

$$X \perp\!\!\!\perp \Omega.$$

Here we aim to test $H_0 : X \perp\!\!\!\perp \Omega$.

Observables

Write $\mathbb{S} = \{S \subseteq [d] : \mathbb{P}(\Omega = \mathbb{1}_S) > 0\}$ for the set of possible observation patterns.



Little & Rubin (2019)

Write P_S for the distribution of $X_S | \{\Omega = \mathbb{1}_S\}$ for $S \in \mathbb{S}$ and write $P_{\mathbb{S}} = (P_S : S \in \mathbb{S})$. These are the distributions we have access to.

Little's test

If data is *Gaussian* and *all pairs* of variables are observed together, the EM algorithm can be used to find MLEs for the population mean and covariance matrix.

Little (1988) estimates means and covariances within each observation pattern and compares to null MLEs with LR test.

$$d^2 = \sum_{j=1}^{J} m_j (\bar{\mathbf{y}}_{\text{obs},j} - \hat{\boldsymbol{\mu}}_{\text{obs},j}) \tilde{\boldsymbol{\Sigma}}_{\text{obs},j}^{-1} (\bar{\mathbf{y}}_{\text{obs},j} - \hat{\boldsymbol{\mu}}_{\text{obs},j})^T.$$

Fuchs's test

When $\mathcal{X}$ is discrete and complete cases are available ($[d] \in \mathbb{S}$), the EM algorithm can be used to find the MLE for the population distribution.

Fuchs (1982) derives the LR test statistic comparing this to observed distributions. With a large number of *complete cases* this has an approximate χ^2 distribution.

$$G^2 = \sum_i \sum_j \sum_k {y_{ijk}}^{ABC} \ln \frac{{y_{ijk}}^{ABC}}{n_0 \hat{p}_{ijk}}$$

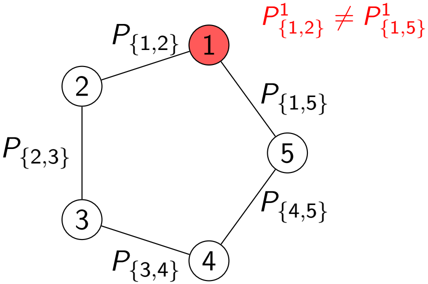
$$+ 2 \sum_t \sum_{t(ijk)} {y_{t(ijk)}}^{t(ABC)} \ln \frac{{y_{t(ijk)}}^{t(ABC)}}{n_t \sum_{\sim t(ijk)} \hat{p}_{ijk}}$$

Nonparametric tests of MCAR

In nonparametric settings, many works have proposed tests based on two sample testing methodology (Li & Yu, 2015; Michel, Spohn & Meinshausen, 2021).

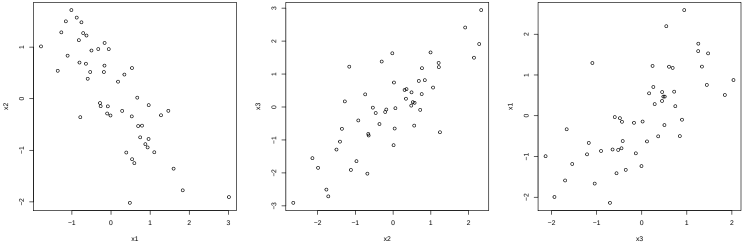
We can rule out H_0 if there are two observation patterns $S_1, S_2 \in \mathbb{S}$ for which P_{S_1} and P_{S_2} have different marginal distributions on $\mathcal{X}_{S_1 \cap S_2}$.

We say that $P_{\mathbb{S}}$ is *consistent* if $P_{S_1}^{S_1 \cap S_2} = P_{S_2}^{S_1 \cap S_2}$ whenever $S_1, S_2 \in \mathbb{S}$ have $S_1 \cap S_2 \neq \emptyset$.



Consistency is not sufficient

It is straightforward to see that there exist non-MCAR settings where all such tests would have trivial power.



$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}\right), \quad \begin{pmatrix} X_2 \\ X_3 \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \quad \begin{pmatrix} X_3 \\ X_1 \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

However, we can rule out MCAR if $\rho > 1/2$.

Our aims

We would like to introduce methods that:

- Do not rely on parametric assumptions;
- Can be used for any $\mathbb{S}$, without a need for complete cases (or, e.g. data on each pair of variables);
- Have power against all detectable alternatives.

Outline of the rest of the talk

1. Introduction
2. Fréchet classes and compatibility
3. Testing compatibility
    - Simple universal discrete test
    - More powerful tests
4. Examples
    - $d = 3$
    - Reductions
    - $d = 4$
    - Continuous data
5. Monte Carlo method and numerical results

Fréchet classes

Define the *Fréchet class*

$$\mathcal{F}(P_{\mathbb{S}}) := \{P \text{ on } \mathcal{X} : P \text{ has marginal distribution } P_S \text{ on } \mathcal{X}_S\},$$

where $\mathcal{X}_S = \prod_{j \in S} \mathcal{X}_j$. Say $P_{\mathbb{S}}$ is compatible if $\mathcal{F}(P_{\mathbb{S}}) \neq \emptyset$.

[1] More generally, compatibility is equivalent to consistency if $\mathbb{S}$ is *decomposable* (Lauritzen & Spiegelhalter, 1988)

Fréchet classes

Define the *Fréchet class*

$$\mathcal{F}(P_{\mathbb{S}}) := \big\{P \text{ on } \mathcal{X} : P \text{ has marginal distribution } P_S \text{ on } \mathcal{X}_S\big\},$$

where $\mathcal{X}_S = \prod_{j \in S} \mathcal{X}_j$. Say $P_{\mathbb{S}}$ is compatible if $\mathcal{F}(P_{\mathbb{S}}) \neq \emptyset$.

- If $\mathbb{S} = \big\{\{1\}, \ldots, \{d\}\big\}$ then $\mathcal{F}(P_{\mathbb{S}}) \neq \emptyset$ for any $P_{\mathbb{S}} = (P_{\{1\}}, \ldots, P_{\{d\}})$.
- If $[d] = \{1, \ldots, d\} \in \mathbb{S}$ then either $\mathcal{F}(P_{\mathbb{S}}) = \{P_{[d]}\}$ or $\mathcal{F}(P_{\mathbb{S}}) = \emptyset$. Here compatibility is equivalent to consistency[1].
- If $\mathbb{S} = \big\{\{1,2\}, \{2,3\}, \{1,3\}\big\}$ consistency is not sufficient for compatibility, as per earlier Gaussian example.

[1] More generally, compatibility is equivalent to consistency if $\mathbb{S}$ is *decomposable* (Lauritzen & Spiegelhalter, 1988)

Compatibility

If H_0 holds then $X_S | \{\Omega = \mathbb{1}_S\} \stackrel{\mathrm{d}}{=} X_S$, so $\mathcal{L}(X) \in \mathcal{F}(P_{\mathbb{S}})$ and $\mathcal{F}(P_{\mathbb{S}}) \neq \emptyset$.

Compatibility

If H_0 holds then $X_S | \{\Omega = \mathbb{1}_S\} \stackrel{\mathrm{d}}{=} X_S$, so $\mathcal{L}(X) \in \mathcal{F}(P_{\mathbb{S}})$ and $\mathcal{F}(P_{\mathbb{S}}) \neq \emptyset$.

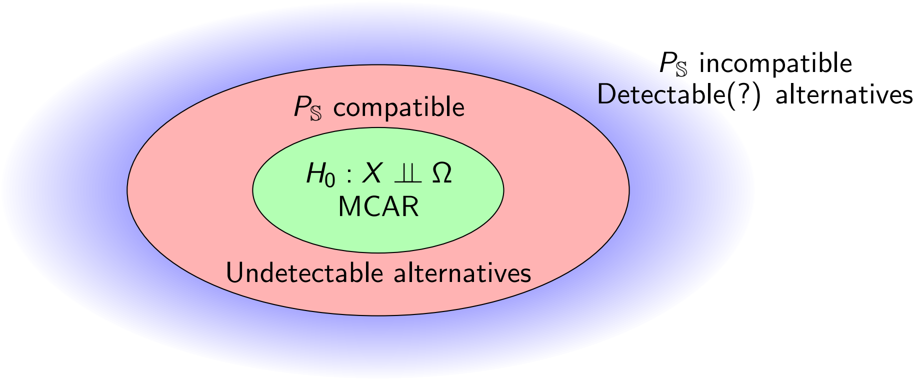
On the other hand, if $\mathcal{F}(P_{\mathbb{S}}) \neq \emptyset$, then there exists P such that, if $\tilde{X} \sim P$ is independent of (X, Ω), then

$$\tilde{X} \circ \Omega \stackrel{\mathrm{d}}{=} X \circ \Omega.$$

But the distribution of $(\tilde{X}, \Omega)$ satisfies H_0.

Compatibility

$$H_0 \implies \mathcal{F}(P_{\mathbb{S}}) \neq \emptyset \quad \text{and} \quad \mathcal{F}(P_{\mathbb{S}}) \neq \emptyset \implies \text{cannot rule out } H_0.$$



The best we can do is test the compatibility of $P_{\mathbb{S}}$.

Testing compatibility

We slightly change our model. For fixed $\mathbb{S} \subseteq 2^{[d]}$, distributions $(P_S : S \in \mathbb{S})$ with P_S on $\mathcal{X}_S$, and deterministic sample sizes $(n_S : S \in \mathbb{S})$ we observe

$$X_{S,1}, \ldots, X_{S,n_S} \overset{\text{i.i.d.}}{\sim} P_S \quad \forall S \in \mathbb{S}, \text{ independently.}$$

With this data we aim to test

$$H_0' : \mathcal{F}(P_{\mathbb{S}}) \neq \emptyset.$$

Testing compatibility

We slightly change our model. For fixed $\mathbb{S} \subseteq 2^{[d]}$, distributions $(P_S : S \in \mathbb{S})$ with P_S on $\mathcal{X}_S$, and deterministic sample sizes $(n_S : S \in \mathbb{S})$ we observe

$$X_{S,1}, \ldots, X_{S,n_S} \overset{\text{i.i.d.}}{\sim} P_S \quad \forall S \in \mathbb{S}, \text{ independently.}$$

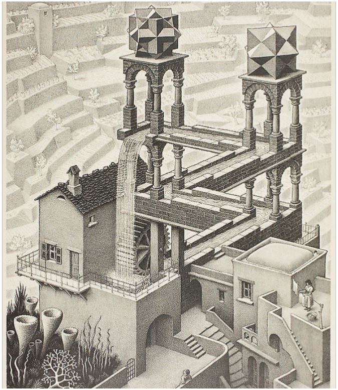
With this data we aim to test

$$H_0' : \mathcal{F}(P_{\mathbb{S}}) \neq \emptyset.$$

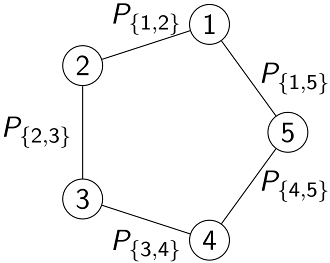
In fact, tests of compatibility are needed in other areas beyond missing data.

Quantum contextuality

'...measurements of quantum observables cannot simply be thought of as revealing pre-existing values' (Wikipedia). See Bell (1966).



M. C. Escher (Cunha, 2019)



Local measurements $P_{\mathbb{S}}$ may not glue together for a sensible global picture.

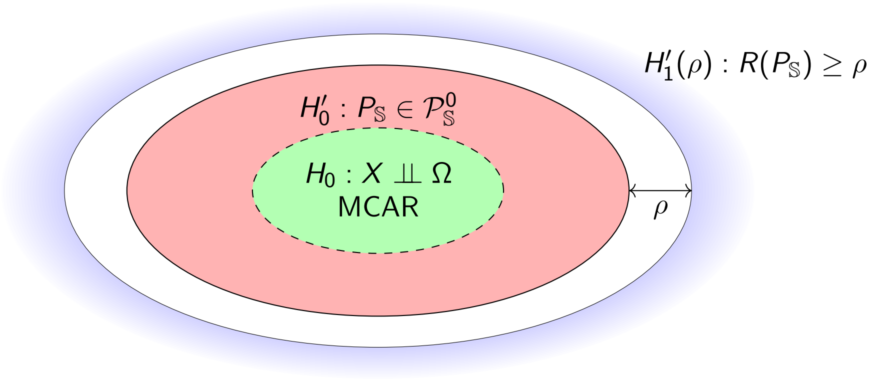
Other relevant areas

- Expert systems (Lauritzen & Spiegelhalter, 1988).
- Meta analysis (Massa & Lauritzen, 2010).
- Relational database theory (Abramsky, 2013).
- Quantitative risk management (Puccetti & Rüschendorf, 2012).

Incompatibility index

To formalise the problem we define an incompatibility index $R(\cdot)$ so that we test

$$H_0' : R(P_{\mathbb{S}}) = 0 \quad \text{vs.} \quad H_1'(\rho) : R(P_{\mathbb{S}}) \geq \rho$$



Incompatibility index

Let $\mathcal{G}_{\mathbb{S}}$ be the set of sequences $(f_S : S \in \mathbb{S})$, where $f_S : \mathcal{X}_S \to [-1, \infty)$ is upper semi-continuous. Take

$$\mathcal{G}_{\mathbb{S}}^{+} = \left\{ f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}} : \inf_{x \in \mathcal{X}} \sum_{S \in \mathbb{S}} f_S(x_S) \geq 0 \right\}.$$

We may then define

$$R(P_{\mathbb{S}}) := \sup_{f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^{+}} R(P_{\mathbb{S}}, f_{\mathbb{S}}), \quad \text{where } R(P_{\mathbb{S}}, f_{\mathbb{S}}) = -\frac{1}{|\mathbb{S}|} \sum_{S \in \mathbb{S}} \int_{\mathcal{X}_S} f_S(x_S)\, dP_S(x_S).$$

Using Kellerer (1984) we have[2] $R(P_{\mathbb{S}}) = 0$ if and only if $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{0}$.

[2] Farkas: $\exists p \geq 0$ s.t. $\mathbb{A}p = p_{\mathbb{S}} \iff p_{\mathbb{S}}^{T} f_{\mathbb{S}} \geq 0 \ \forall f_{\mathbb{S}}$ s.t. $\mathbb{A}^{T} f_{\mathbb{S}} \geq 0$

Dual form of $R(\cdot)$

When each $\mathcal{X}_j$ is a locally compact Hausdorff space such that every open set in $\mathcal{X}$ is σ-compact, we have

$$R(P_{\mathbb{S}}) = \inf\big\{\epsilon \in [0,1] : P_{\mathbb{S}} \in (1-\epsilon)\mathcal{P}_{\mathbb{S}}^0 + \epsilon\mathcal{P}_{\mathbb{S}}\big\}.$$

Clearly $R(P_{\mathbb{S}}) \in [0,1]$. When $\mathcal{X}$ is discrete $R(P_{\mathbb{S}}) < 1$ if and only if there exists $x \in \mathcal{X}$ with $P_S(\{x_S\}) > 0$ for all $S \in \mathbb{S}$.

A simple test for discrete $\mathcal{X}$

$R(P_{\mathbb{S}})$ can be computed by linear programming and, writing $\hat{P}_{\mathbb{S}}$ for the empirical distributions, we can take the test statistic $\hat{R} := R(\hat{P}_{\mathbb{S}})$.

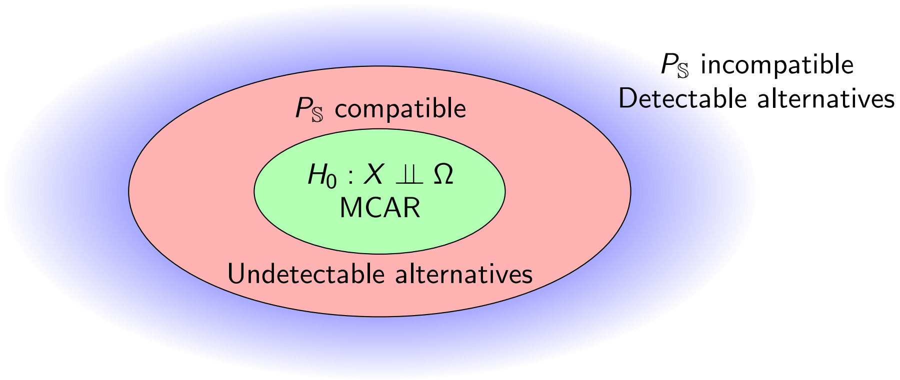
The challenge is to find a critical value: general bounds give $\mathbb{P}_{H_0'}(\hat{R} \geq C_\alpha) \leq \alpha$ when

$$C_\alpha := \frac{1}{2} \sum_{S \in \mathbb{S}} \Big(\frac{|\mathcal{X}_S| - 1}{n_S} \Big)^{1/2} + \bigg\{ \frac{1}{2} \log(1/\alpha) \sum_{S \in \mathbb{S}} \frac{1}{n_S} \bigg\}^{1/2}.$$

When $\mathcal{X}$ is fixed and $\min_S n_S \to \infty$, we have asymptotic power 1 against all fixed alternatives. In fact, whenever $R(P_{\mathbb{S}}) \geq C_\alpha + C_\beta$ we have

$$\mathbb{P}(\hat{R} \geq C_\alpha) \geq 1 - \beta.$$

Compatibility and detectability



Improved tests

In choosing the critical value C_α we used the bound

$$\sup_{f_{\mathbb{S}} \in \mathcal{G}_{\mathbb{S}}^+} R(\hat{P}_{\mathbb{S}}, f_{\mathbb{S}}) \leq \sup_{-1 \leq f_{\mathbb{S}} \leq |\mathbb{S}|-1} R(\hat{P}_{\mathbb{S}}, f_{\mathbb{S}}),$$

which is generally loose as it ignore the constraints

$$\min_{x \in \mathcal{X}} (\mathbb{A}^T f_{\mathbb{S}})_x = \min_{x \in \mathcal{X}} \sum_{S \in \mathbb{S}} f_S(x_S) \geq 0.$$

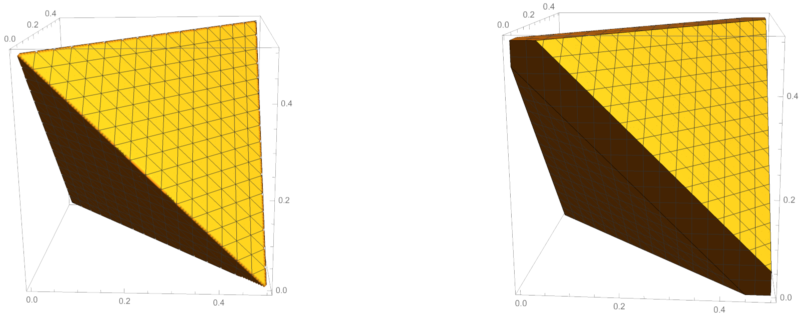
Although generally very complicated, we can do better with more understanding of $R(\cdot)$.

Testing membership of a convex polytope

The null space $\mathcal{P}^0_{\mathbb{S}}$ is the convex hull of the columns of $\mathbb{A}$, an $\mathcal{X}_{\mathbb{S}} \times \mathcal{X}$ matrix with

$$\mathbb{A}_{(S,x_S),y} = \mathbb{1}_{\{x_S=y_S\}}.$$

In fact, $\mathcal{P}^0_{\mathbb{S}}$ is a full-dimensional subset of $\mathcal{P}^{\mathrm{cons}}_{\mathbb{S}}$, the set of consistent sequences.



Optimal testing over convex polyhedra depends on the specific geometry

(Blanchard, Carpentier & Gutzeit, 2018; Wei, Wainwright & Guntuboyina, 2019).

A decomposition

Let $F \in \mathbb{N}_0$ and $f_{\mathbb{S}}^{(1)}, \ldots, f_{\mathbb{S}}^{(F)} \in \mathcal{G}_{\mathbb{S}}^{+}$ be such that

$$R(P_{\mathbb{S}}) = \max_{\ell \in [F]} R(\hat{P}_{\mathbb{S}}, f_{\mathbb{S}}^{(\ell)}) \qquad \text{for } P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}.$$

Proposition

For $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}$ we have

$$R(P_{\mathbb{S}}) \asymp_{\mathbb{S}} \max_{\ell \in [F]} R(P_{\mathbb{S}}, f_{\mathbb{S}}^{(\ell)}) + \max_{S_1, S_2 \in \mathbb{S}} d_{\mathrm{TV}}\left(P_{S_1}^{S_1 \cap S_2}, P_{S_2}^{S_1 \cap S_2}\right).$$

A decomposition

Let $F \in \mathbb{N}_0$ and $f_{\mathbb{S}}^{(1)}, \ldots, f_{\mathbb{S}}^{(F)} \in \mathcal{G}_{\mathbb{S}}^{+}$ be such that

$$R(P_{\mathbb{S}}) = \max_{\ell \in [F]} R(\hat{P}_{\mathbb{S}}, f_{\mathbb{S}}^{(\ell)}) \qquad \text{for } P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}.$$

Proposition

For $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}$ we have

$$R(P_{\mathbb{S}}) \asymp_{\mathbb{S}} \max_{\ell \in [F]} R(P_{\mathbb{S}}, f_{\mathbb{S}}^{(\ell)}) + \max_{S_1, S_2 \in \mathbb{S}} d_{\mathrm{TV}}\left(P_{S_1}^{S_1 \cap S_2}, P_{S_2}^{S_1 \cap S_2}\right).$$

If F is known we can choose a critical value

$$C'_\alpha \asymp_{\mathbb{S}} \frac{\log(F/\alpha)}{\min_S n_S} + \max_{S_1, S_2 \in \mathbb{S} : S_1 \cap S_2 \neq \emptyset} \frac{|\mathcal{X}_{S_1 \cap S_2}|}{n_{S_1} \wedge n_{S_2}}.$$

$d = 3$ example

Theorem

Let $\mathbb{S} = \big\{\{1,2\},\{2,3\},\{1,3\}\big\}$ *and* $\mathcal{X} = [r] \times [s] \times [2]$ *for* $r, s \geq 2$. *Then for* $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$ *we have*

$$R(P_{\mathbb{S}}) = 2 \max_{A \subseteq [r], B \subseteq [s]} \Big\{ P_{\{1,2\}}(A \times B) + P_{\{1,3\}}(A \times \{1\}) - P_{\{2,3\}}(B^c \times \{1\}) \Big\}_+$$

In particular, we may take $F = (2^r - 2)(2^s - 2)$ and design a test with separation rate

$$\left(\frac{r+s}{n_{\{1,2\}}}\right)^{1/2} + \left(\frac{r}{n_{\{1,3\}}}\right)^{1/2} + \left(\frac{s}{n_{\{2,3\}}}\right)^{1/2}.$$

Idea of proof

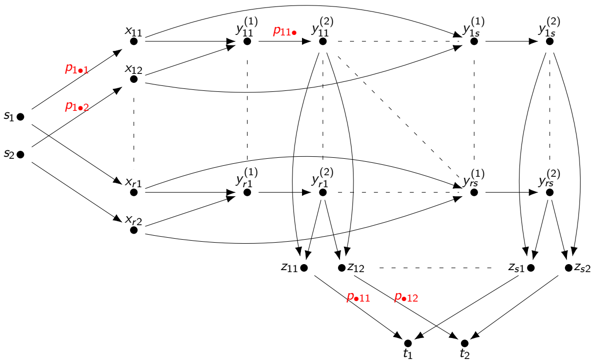
Lower bound via primal problem $R(P_{\mathbb{S}}) \geq \max_{A \subseteq [r], B \subseteq [s]} R(P_{\mathbb{S}}, f_{\mathbb{S}}^{A,B})$.

Idea of proof

Lower bound via primal problem $R(P_{\mathbb{S}}) \geq \max_{A\subseteq[r],B\subseteq[s]} R(P_{\mathbb{S}}, f_{\mathbb{S}}^{A,B})$.

Upper bound via dual problem

$$R(P_{\mathbb{S}}) = 1 - \max\left\{\sum_{i,j,k} p_{ijk} : \sum_{i=1}^{r} p_{ijk} \leq p_{\bullet jk}, \sum_{j=1}^{s} p_{ijk} \leq p_{i\bullet k}, p_{ij1}+p_{ij2} \leq p_{ij\bullet}\right\}.$$



Lower bound for this example

Theorem

In the previous setting, when $n_{\{1,2\}} \geq (r+s)\log(r+s)$, $n_{\{1,3\}} \geq r \log r$ *and* $n_{\{2,3\}} \geq s \log s$ *we have a minimax lower bound of the order*

$$\Big(\frac{r+s}{n_{\{1,2\}}\log(r+s)}\Big)^{1/2} + \Big(\frac{r}{n_{\{1,3\}}\log r}\Big)^{1/2} + \Big(\frac{s}{n_{\{2,3\}}\log s}\Big)^{1/2}$$

Up to log factors, our test is rate optimal.

Lower bound for this example

Theorem

In the previous setting, when $n_{\{1,2\}} \geq (r+s)\log(r+s)$, $n_{\{1,3\}} \geq r\log r$ and $n_{\{2,3\}} \geq s\log s$ we have a minimax lower bound of the order

$$\left(\frac{r+s}{n_{\{1,2\}}\log(r+s)}\right)^{1/2} + \left(\frac{r}{n_{\{1,3\}}\log r}\right)^{1/2} + \left(\frac{s}{n_{\{2,3\}}\log s}\right)^{1/2}$$

Up to log factors, our test is rate optimal.

When $s=2$ and $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$ with $p_{\bullet 1 \bullet} = p_{\bullet\bullet 1} = 1/2, p_{\bullet 21} \geq 1/4$ and with $p_{i\bullet\bullet} = 1/r, p_{i\bullet 1} = 1/(2r)$ for $i \in [r]$, we have

$$R(P_{\mathbb{S}}) = \left(\sum_{i=1}^{r}\left|p_{i1\bullet} - \frac{1}{2r}\right| + 2p_{\bullet 21} - 1\right)_+ .$$

This testing problem is then at least as hard as estimating L_1 distances, and we can adapt Cai & Low (2011); Jiao, Han & Weissman (2018).

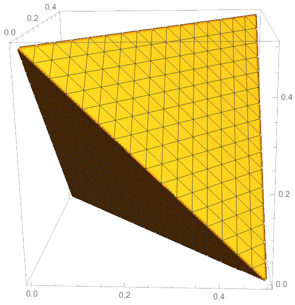
Testing against consistent alternatives

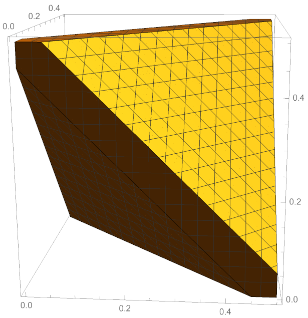
The constructions in previous lower bound satisfied $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\text{cons}}$.

Testing against

$$H_1''(\rho) : R(P_{\mathbb{S}}) \geq \rho, P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\text{cons}}$$

is as hard as the original problem $(H_1'(\rho) : R(P_{\mathbb{S}}) \geq \rho)$.





Reductions

For other $(\mathbb{S}, \mathcal{X})$, analytic expressions for $R(P_{\mathbb{S}})$ can be difficult, but we can sometimes reduce to simpler problems.

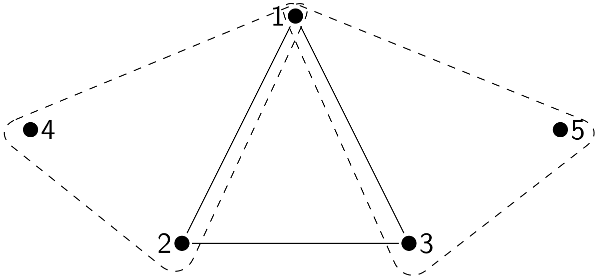
Reductions

For other $(\mathbb{S}, \mathcal{X})$, analytic expressions for $R(P_{\mathbb{S}})$ can be difficult, but we can sometimes reduce to simpler problems.

If there exists $J \subseteq [d]$ and $S_0 \in \mathbb{S}$ with $J \subseteq S_0$ and $J \cap S = \emptyset$ for all $S \in \mathbb{S} \setminus \{S_0\}$, then

$$R(P_{\mathbb{S}}) = R(P_{\mathbb{S}}^{-J}).$$

$\mathbb{S} = \{\{1,2,4\},\{2,3\},\{1,3,5\}\}$ reduces to $\mathbb{S} = \{\{1,2\},\{2,3\},\{1,3\}\}$.



Reductions

If there exists $J \subseteq [d]$ such that $J \subseteq S$ and $P_S^J = P^J$ for all $S \in \mathbb{S}$, then

$$R(P_{\mathbb{S}}) = \sum_{x_J \in \mathcal{X}_J} R(P_{\mathbb{S}|X_J = x_j}) p^J(x_j).$$

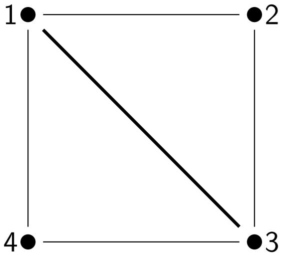
When $\mathbb{S} = \big\{\{1,2,3\},\{1,3,4\},\{1,2,4\}\big\}$ with $\mathcal{X} = [r] \times [s] \times [t] \times [2]$ then when $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$ we have

$$R(P_{\mathbb{S}}) = 2 \sum_{i=1}^{r} \max_{A \subseteq [s], B \subseteq [t]} (-p_{iAB\bullet} + p_{iA\bullet 1} + p_{i\bullet B1} - p_{i\bullet\bullet 1})_+.$$

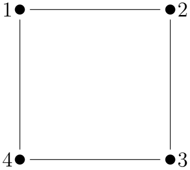
Reductions

If $\mathbb{S}_1, \mathbb{S}_2 \subseteq \mathbb{S}$ are such that there exists $J \in \mathbb{S}$ with $\mathbb{S}_1 \cap \mathbb{S}_2 = \{J\}$ and $(\cup_{S \in \mathbb{S}_1} S) \cap (\cup_{S \in \mathbb{S}_2} S) = J$, then

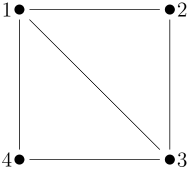
$$\max\{R(P_{\mathbb{S}_1}), R(P_{\mathbb{S}_2})\} \leq R(P_{\mathbb{S}}) \leq R(P_{\mathbb{S}_1}) + R(P_{\mathbb{S}_2}).$$



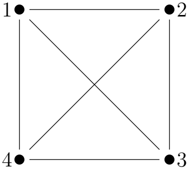
Irreducible $d = 4$ examples



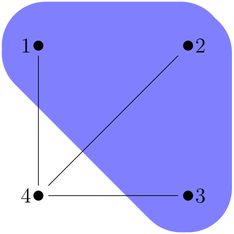
(a) Chain pairs



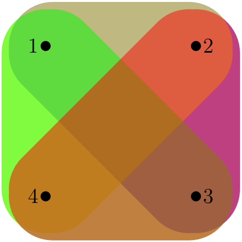
(b) All pairs except one



(c) All pairs



(d) Single triple



(e) All triples

Mixed discrete and continuous cases

By binning continuous variables we can apply our tests designed for the discrete setting.

In particular, when $\mathcal{X} = [0,1]^2 \times \{1,2\}$ and the densities on $\mathcal{X}_j$ are (r_j, L)-Hölder smooth ($j = 1, 2$), we have a test with power whenever

$$R(P_{\mathbb{S}}) \geq C_{\mathbb{S},L} n^{-\frac{r_1 \wedge r_2}{1+2(r_1 \wedge r_2)}}.$$

Numerical results

Tests so far have had uniform, finite-sample Type I error control, but can be conservative. We propose a simple Monte Carlo test that can perform well in practice.

When $\mathcal{X}$ is discrete we can solve the dual linear program for $R(\hat{P}_{\mathbb{S}})$ to find a decomposition

$$\hat{P}_{\mathbb{S}} = \{1 - R(\hat{P}_{\mathbb{S}})\}\hat{Q}_{\mathbb{S}} + R(\hat{P}_{\mathbb{S}})\hat{T}_{\mathbb{S}} \in \{1 - R(\hat{P}_{\mathbb{S}})\}\mathcal{P}^0_{\mathbb{S}} + R(\hat{P}_{\mathbb{S}})\mathcal{P}_{\mathbb{S}}.$$

Here $\hat{Q}_{\mathbb{S}}$ can be thought of as a closest compatible sequence of marginal distributions to $\hat{P}_{\mathbb{S}}$. We can generate bootstrap samples $\hat{Q}^{(1)}_{\mathbb{S}}, \ldots, \hat{Q}^{(B)}_{\mathbb{S}}$ and reject H_0' if and only if

$$1 + \sum_{b=1}^{B} \mathbb{1}_{\{R(\hat{Q}^{(b)}_{\mathbb{S}}) \leq R(\hat{Q}_{\mathbb{S}})\}} \leq \alpha(B + 1).$$

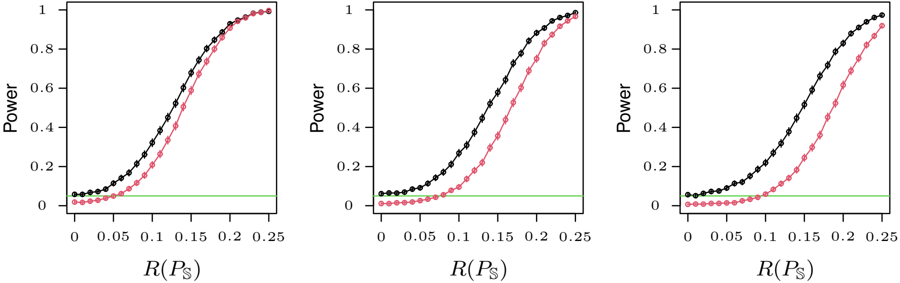
Numerical results

We compare with Fuchs's LR test. With $\mathbb{S} = \big\{\{1,2\},\{2,3\},\{1,3\}\big\}$, with $\mathcal{X} = [r] \times [2]^2$ for $r \in \{2,4,6\}$ and with $P_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{\mathrm{cons}}$ defined by

$$p_{i\bullet\bullet} = \frac{1}{r}, p_{\bullet 1 \bullet} = p_{\bullet\bullet 1} = \frac{1}{2}, p_{i\bullet 1} = \frac{1}{2r}, p_{i\bullet 1} = \frac{1+(-1)^i}{2r}$$

and $p_{\bullet 21} \in [0.25, 0.375]$, we take $n_{\mathbb{S}} = (200, 200, 200)$, $B = 99$, $\alpha = 0.05$.

Fuchs's test requires complete cases, so we allow it access to 200 observations from a distribution on $\mathcal{X}$.



Numerical results

Now take $d = 5$, $\mathcal{X} = [2]^5$ and

$$\mathbb{S} = \big\{\{1,2,3,4\},\{1,2,3,5\},\{1,2,4,5\},\{1,3,4,5\},\{2,3,4,5\}\big\}.$$

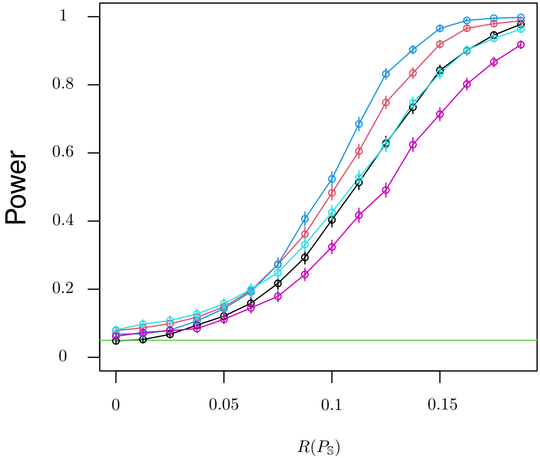
For $\epsilon \in [0.2, 0.35]$ and $i, j, k, \ell, m \in [2]$, we set

$$\begin{aligned} p_{ijk\ell\bullet} = p_{ijk\bullet\ell} = p_{ij\bullet k\ell} = p_{i\bullet jk\ell} &= \frac{1 + \epsilon(-1)^{i+j+k+\ell}}{16}, \\ p_{\bullet ijk\ell} &= \frac{1 - \epsilon(-1)^{i+j+k+\ell}}{16}, \end{aligned}$$

for which $R(P_{\mathbb{S}}) = (5\epsilon - 1)_+/4$.

Numerical results

Allowing Fuchs's test $\{25, 50, 100, 200\}$ complete cases



Conclusion

- Shown testing MCAR is equivalent to testing compatibility;
- General test with asymptotic power against fixed alternatives for discrete/discretisable data;
- Improved tests given knowledge of underlying geometry (rate-optimal in cases);
- Monte Carlo method with good empirical power.

THANK YOU!

Berrett, T. B. & Samworth, R. J. (2022) Optimal nonparametric testing of Missing Completely At Random, and its connections to compatibility arXiv:2205.08627.

References

Cai, T. T. & Low, M. G. (2011) Testing composite hypotheses, Hermite polynomials & optimal estimation of a nonsmooth functional. *Ann. Statist.*, **39**, 1012–1041.

Fuchs, C. (1982) Maximum likelihood estimation & model selection in contingency tables with missing data. *J. Amer. Statist. Assoc.*, **77**, 270–278.

Jiao, J., Han, Y. & Weissman, T. (2018) Minimax estimation of the L_1 distance. *IEEE Trans. Inf. Theory*, **64**, 6672–6706.

Kellerer, H. G. (1984) Duality theorems for marginal problems. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **67**, 399–432.

Lauritzen, S. L. & Spiegelhalter, D. J. (1988) Local computations with probabilities on graphical structures & their application to expert systems. *J. Roy. Statist. Soc., Ser. B*, **50**, 157–194.

Little, R. J. (1988) A test of missing completely at random for multivariate data with missing values. *J. Amer. Statist. Assoc.*, **83**, 1198–1202.

References

Abramsky, S. (2013) Relational databases & Bell's theorem. *In Search of Elegance in the Theory & Practice of Computation*, 13–15.

Bell, J. S. (1966) On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.*, **38**, 447.

Blanchard, G., Carpentier, A. & Gutzeit, M. (2018) Minimax Euclidean separation rates for testing convex hypotheses in $\mathbb{R}^d$. *Electr. J. Statist.*, **12**, 3713–3735.

Cai, T. T. & Zhang, L. (2019) High dimensional linear discriminant analysis: Optimality, adaptive algorithm & missing data. *J. Roy. Statist. Soc., Ser. B*, **81**, 675–705.

Cunha, M. T. (2019) On measures & measurements: a fibre bundle approach to contextuality. *Philos. Trans. R. Soc.* **377.2157**, 20190146.

Dwork, C. & Lei, J. (2009) Differential privacy & robust statistics. *Annual ACM Symposium on Theory of Computing*, 371–380.

Elsener, A. & van de Geer, S. (2019) Sparse spectral estimation with missing & corrupted measurements. *Stat*, **8**, e229.

Follain, B., Wang, T. & Samworth, R. J. (2022) High-dimensional changepoint estimation with heterogeneous missingness. *J. Roy. Statist. Soc., Ser. B, to appear.*

Li, J. & Yu, Y. (2015) A nonparametric test of missing completely at random for incomplete multivariate data. *Psychometrika*, **80**, 707–726.

Little, R. J. & Rubin, D. B. (2019) *Statistical Analysis with Missing Data*. John Wiley & Sons.

References

Loh, P.-L. & Tan, X. L. (2018) High-dimensional robust precision matrix estimation: Cellwise corruption under ϵ-contamination. *Electr. J. Statist.*, **12**, 1429–1467.

Loh, P.-L. & Wainwright, M. J. (2012) High-dimensional regression with noisy & missing data: Provable guarantees with nonconvexity. *Ann. Statist.*, **40**, 1637–1664.

Massa, M. S. & Lauritzen, S. L. (2010) Combining statistical models. *Contemporary Mathematics: Algebraic Methods in Statistics & Probability II*, 239–260.

Michel, L., Näf, J., Spohn, M.-L. & Meinshausen, N. (2021) PKLM: A flexible MCAR test using Classification. *arXiv preprint arXiv:2109.10150*.

Puccetti, G. & Rüschendorf, L. (2012) Bounds for joint portfolios of dependent risks. *Statistics & Risk Modeling*, **29**, 107–132.

Wei, Y., Wainwright, M. J. & Guntuboyina, A. (2019) The geometry of hypothesis testing over convex cones: Generalized likelihood ratio tests and minimax radii. *Ann. Statist.*, **47**, 994–1024.

Zhu, Z., Wang, T. & Samworth, R. J. (2019) High-dimensional principal component analysis with heterogeneous missingness. *arXiv preprint arXiv:1906.12125*.