Algorithmic + Geometric characterization of CAR
(Coarsening at Random)

Richard Gill - Utrecht Leiden
Peter Grünwald - CWI Amsterdam
both EURANDOM, Eindhoven

arXiv.org: math.ST/0510276
Ann. Statist. to appear
Coarsening

Underlying data \( \text{law}_\theta(X) \)

Observation \( \text{law}_\phi(Y|X) \)

Observed data \( \text{law}_{\theta,\phi}(Y) \)

\( X = x \in E \)

\( \#E < \infty \)

observe \( Y = A \subseteq E \)

\( A \ni x \)

Notation: \( y \equiv A \)
Examples (?)

- partition (fixed, or random but independent) CCAR
- 3 door problem $X=\text{door with car behind}$
  $Y=\text{two doors still closed}$
  $= \{\text{your first choice, other door left closed}\}$
- forgetful quizplayer
- 3 door problem $X=\text{door with car behind}$
  $Y=\text{(your first choice, other door left closed)}$
Coarsening AT RANDOM

- can do statistical analysis of data at face value
  ie, as if we observe $1\{X \in y\}$
- Likelihood is $P_\theta(X \in y)$
- Can use naive EM
Coarsening at Random

\[ P_\phi(Y = y | X = x) \text{ is same for all } x \in y \]

\[ P_{\theta,\phi}(Y = y) = P_\phi(Y = y | X = x) \cdot P_\theta(X \in y) \]

any \( x \in y \)
How to simulate an arbitrary CAR mechanism?

**WRONG ANSWER:**

- generate $x$ from $\text{law}_\theta(X)$
- generate $y$ from $\text{law}_\phi(Y | X = x)$
- report $Y = y$
“randomized monotone coarsening” ?
but ∃ CAR models which are not RMC
∃ CAR models which cannot be honest
     (do not tell the truth)
honest CAR ⇔ RMC ?
∃ cute CAR algorithms but which are frail, ie, become non CAR under perturbation of parameters --- need delicate fine tuning
Manfred Jaeger (Ålborg, CS)

- $\diamond$ robust CAR $\iff$ CCAR
- $\diamond$ honest CAR $\iff$ CCAR
- $\diamond$ RMC $\iff$ CCAR
\[ \pi_A = \pi_A^x = P(Y = A|X = x) \]

\[ \sum_{A \ni x} \pi_A = 1 \quad \forall x \in E \]

\[ \pi_A \geq 0 \quad \forall A \]

Gill and Grünwald

linear equalities

linear inequalities
Gill and Grünwald (Jaeger almost):

\[ \vec{\pi} = (\pi_A : A \subseteq E) \]

\{CAR $\vec{\pi}$ \} is a convex polytope

every CAR $\vec{\pi}$ is a mixture of extreme CAR models

each extreme CAR has rational probabilities

rational CAR $\iff$ random uniform multicover
\textit{multicover} \( \mathcal{A} \): set of nonempty subsets of \( E \), allowing multiplicity, covering \( E \)

\textit{uniform} multicover: each \( x \) in \( E \) is in the same number of elements of \( \mathcal{A} \)

\textit{depth} of uniform multicover: this number

\textit{rational CAR} \iff \exists \text{ uniform multicover } \mathcal{A} \text{ given } x \text{ in } E, \text{ choose element of } \mathcal{A} \text{ covering } x, \text{ uniformly at random, ie, prob } = 1 / \text{ depth}
PROOF:

- Intersection of rational hyperplanes is rational point
- take LCM, write rational \[ \pi_A = \frac{n_A}{n} \]

\[ \sum_{A \ni x} n_A = n \quad \forall x \in E \]
rational CAR are dense in all CAR

every CAR is mixture of extreme (rational) CAR

every CAR has nice, robust (?) algorithmic description

but obviously, dishonest when depth > 1
But unattractive when depth is very large

So, HOW LARGE COULD IT BE ??
VERY LARGE! : Fibonacci CAR
FURTHER CHARACTERIZATION:

Extreme CAR $\equiv$ multicover unique for its support

1. $M_A = \text{incidence matrix of support of } A$

2. $A$ is extreme iff $M_A \vec{x} = \vec{1}$ has unique positive solution
Fibonacci CAR

- \#E = 1: only one CAR; it is Fibonacci

- \#E = 2m+1:
  - take support of Fibonacci CAR for \#E = 2m-1
  - add two points to E
  - add ONE of the new points to each coarsening in OLD support
  - add to support: \{OTHER new point, all old points\}
  - add to support: \{two new points\}
DEFINE: $F_0 = F_1 = 1; F_n = F_{n-1} + F_{n-2}$

CLAIMS:

1. $M_A \vec{x} = \vec{1}$ has a unique positive solution for every odd $n = \#E$

2. $\pi_{A_i} = x_{A_i} = \frac{F_i}{F_n}, i = 0, \ldots, n - 1$

PROOF: Induction

CONCLUSION: The maximal depth of extremal CAR grows at least exponentially with $n$
FINAL REMARKS

- robustness is what you make of it

- a cute algorithm is not necessarily a NATURE-al mechanism

- Fibonacci (3) is the forgetful quizplayer:
  is there a natural mechanism for Fibonacci (2m+1)?

- statistical inference with non-CCAR CAR

- statistical inference with non-CAR

- relative CAR $\rightarrow$ useful non CAR models?