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 → Observation → Observed data

$\mathrm{law}_\theta(X)$ $\quad$ $\mathrm{law}_\phi(Y|X)$ $\quad$ $\mathrm{law}_{\theta,\phi}(Y)$

$X = x \in E$

observe $Y = A \subseteq E$

$A \ni x$

$\#E < \infty$

Notation: $y \equiv A$

Examples (?)

- partition (fixed, or random but independent)
  CCAR

- 3 door problem X=door with car behind
  Y=two doors still closed
  = {your first choice, other door left closed}

forgetful quizplayer ↗

- 3 door problem X=door with car behind
  Y=(your first choice, other door left closed)

Coarsening AT RANDOM

$$\Longrightarrow$$

- can do statistical analysis of data at face value ie, as if we observe $\mathbb{1}\{X \in y\}$
- Likelihood is $\mathrm{P}_\theta(X \in y)$
- Can use naieve EM

Coarsening at Random

IS

$$\mathrm{P}_{\phi}(Y = y | X = x) \quad \text{is same for all} \quad x \in y$$

$$\Longrightarrow$$

$$\mathrm{P}_{\theta,\phi}(Y = y) \quad = \quad \mathrm{P}_{\phi}(Y = y | X = x) \cdot \mathrm{P}_{\theta}(X \in y)$$

any $x \in y$

How to simulate an arbitrary CAR mechanism?

WRONG ANSWER:

- generate x from $\mathrm{law}_\theta(X)$
- generate y from $\mathrm{law}_\phi(Y|X=x)$
- report Y=y

Gill, vdLaan, Robins

- "randomized monotone coarsening" ?
  but $\exists$ CAR models which are not RMC

- $\exists$ CAR models which cannot be honest
  (do not tell the truthth

- honest CAR $\Longleftrightarrow$ RMC ?

Grünwald and Halpern

- $\exists$ cute CAR algorithms but which are frail, ie, become non CAR under perturbation of parameters --- need delicate fine tuning

Manfred Jaeger (Ålborg, CS)

- robust CAR $\Longleftrightarrow$ CCAR
- honest CAR $\Longleftrightarrow$ CCAR
- RMC $\Longleftrightarrow$ CCAR

Gill and Grünwald

$$\pi_A \;=\; \pi_A^x \;=\; \mathrm{P}(Y = A | X = x)$$

$$\sum_{A \ni x} \pi_A = 1 \qquad \forall x \in E$$

linear equalities

$$\pi_A \geq 0 \quad \forall A$$

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 $\Longleftrightarrow$ random uniform multicover

- multicover $\mathcal{A}$: set of nonempty subsets of E, allowing multiplicity, covering E

- uniform multicover: each x in E is in the same number of elements of $\mathcal{A}$

- depth of uniform multicover: this number

- rational CAR $\Longleftrightarrow$ $\exists$ uniform multicover $\mathcal{A}$
given x in E, choose element of $\mathcal{A}$ covering x, uniformly at random, ie, prob = 1 / depth

PROOF:

- Intersection of rational hyperplanes is rational point
- take LCM, write rational $\pi_A = n_A / n$

multiplicity (n_A) depth (n)

$$\sum_{A \ni x} n_A = n \quad \forall x \in E$$

$\Longrightarrow$

- 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

- $M_{\mathcal{A}}$ = incidence matrix of support of $\mathcal{A}$
- $\mathcal{A}$ is extreme iff $M_{\mathcal{A}}\,\vec{x} = \vec{1}$
  has unique positive solution

Fibonacci CAR

- #E = 1 : only one CAR; it is Fibonnaci
- #E = 2m+1 :
  - take support of Fibonnaci 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:

- $M_A \vec{x} = \vec{1}$ has a unique positive solution for every odd n=#E
- $\pi_{A_i} = x_{A_i} = F_i / F_n, i = 0, \ldots, n-1$

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
- Fibonnaci (3) is the forgetful quizplayer :
  is there a natural mechanism for Fibonnaci (2m+1) ?
- statistical inference with non-CCAR CAR
- statistical inference with non-CAR
- relative CAR -> useful non CAR models?